APPLICATIONS OF SEQUENCING TECHNOLOGIES IN MONITORING THE EVOLUTION AND TRANSMISSION OF VIRAL PATHOGENS

by

Kimia Kamelian

B.Sc., University of British Columbia, 2016

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Experimental Medicine)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2019

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, a thesis/dissertation entitled:

APPLICATIONS OF SEQUENCING TECHNOLOGIES IN MONITORING THE EVOLUTION AND TRANSMISSION OF VIRAL PATHOGENS

submitted by Kimia Kamelian in partial fulfillment of the requirements for

the degree of Master of Science

in Experimental Medicine

Examining Committee:

Dr. P. Richard Harrigan (Faculty of Medicine)
Supervisor

Dr. Jeffrey B. Joy (Faculty of Medicine)
Supervisory Committee Member

Dr. Chanson J. Brumme (BC Centre for Excellence in HIV/AIDS)
Supervisory Committee Member

Dr. Soren Gantt (Faculty of Medicine)
Additional Examiner

Additional Supervisory Committee Members:

Dr. M-J Milloy (Faculty of Medicine)
Supervisory Committee Member

Abstract

Viruses are obligate intracellular parasites, a diverse set of rapidly-evolving pathogens capable of causing severe disease in humans. New approaches to modern medicine have provided viral treatments that improve the clinical status of patients and prevent transmission. Genetic sequencing is a valuable tool in the field of infectious diseases, being used to investigate various properties of infections including host immunity, pathogen characteristics, and evolutionary trends over time.

Viral infections are responsible for millions of deaths per year, and persistent as well as emerging viral infections are associated with high morbidity and mortality around the world. Human immunodeficiency virus (HIV), the virus which causes acquired immune deficiency syndrome (AIDS), was responsible for nearly one million AIDS-related deaths in 2018. Although the current paradigm of HIV treatment involves the use of antiretroviral medications, drug resistance affects the potency and efficacy of antiretroviral therapy. Emerging viruses, such as the Zika virus (ZIKV), present additional threats to the global managements of infectious diseases. Prior to ZIKV outbreak in Yap Island in 2007 that resulted in 74 confirmed cases, the ZIKV was associated with small sporadic outbreaks in Africa and Asia. In 2016, there were approximately 400,000-1.3 million confirmed cases of ZIKV infections in Brazil alone, indicating increased geographical range of ZIKV away from previous known regions of infections, with detrimental neurological outcomes.

In this thesis, the primary objective is to investigate the utility of current sequencing technologies in monitoring and surveillance of circulating viral pathogens. I discuss the extent to which viral sequencing can be used to examine the evolution and transmission of current and emerging viral pathogens, HIV and ZIKV. The aims are to 1) identify the longitudinal annual prevalence of HIV drug resistance in British Columbia, Canada using Sanger sequencing, 2) evaluate the prevalence and the impact of pretreatment HIV drug resistance on treatment outcome in Mbarara, Uganda using “next-generation” sequencing, and 3)

assess the potential utility of “next-generation” sequencing in identifying origins of travel-related ZIKV infections in a proof-of-concept study and examine the changing evolutionary trends of circulating strains of ZIKV.

Lay Summary

Viruses are microscopic agents that can cause illness in humans. Genetic sequencing is useful in the detection and monitoring of viruses. Although human immunodeficiency virus (HIV) can be treated using medications, drug resistance can occur, weakening treatment. Sequencing of the HIV genome can detect drug resistance and is used by physicians to better personalize treatments. Additionally, sequencing is useful in understanding characteristics of newly relevant viruses such as the Zika virus (ZIKV), and may reveal information about their spread in communities.

The objective of this thesis is to investigate the use of genetic sequencing in understanding currently relevant diseases, HIV and ZIKV. There are three aims: 1) identify the annual trend of HIV drug resistance in British Columbia, Canada, 2) discuss the level and impact of pretreatment HIV drug resistance in Uganda, and 3) explore the use of sequencing in determining the origins of travel-related ZIKV infections.

Preface

All of the experimental works presented in the main body of this thesis (Chapter Two, Chapter Three, and Chapter Four) occurred at the British Columbia Centre for Excellence in HIV/AIDS (BC-CfE), an affiliated laboratory of the University of British Columbia in Vancouver BC, Canada.

Chapter Two has previously been published by Oxford Press in Open Forums Infectious Diseases, an open-access journal and has been reprinted here under the terms of the Creative Commons Attribution Licence (CC BY-NC-ND):

Kamelian K, Lepik KJ, Chau W, et al. Prevalence of Human Immunodeficiency Virus-1 Integrase Strand Transfer Inhibitor Resistance in British Columbia, Canada Between 2009 and 2016: A Longitudinal Analysis. Open Forum Infect Dis. **2019**; 6(3):ofz060

I (Kimia Kamelian) am the primary author and was responsible for data analysis and manuscript composition. Katherine J. Lepik and P. Richard Harrigan conceived of the study concept presented. Co-authors P. Richard Harrigan and Rolando Barrios gained funding for the study. Rolando Barrios provided access to the Drug Treatment Program. Wendy W. Zhang carried out the experimental procedures. Conan Woods, William Chau, Marjorie A. Robbins, and Benita Yip were involved in the acquisition of the data. Conan Woods, Benita Yip, and Viviane Dias Lima contributed to the analysis of the data. P. Richard Harrigan, Katherine J. Lepik, Benita Yip, Andrea Olmstead, and Jeffrey B. Joy provided critical feedback on the manuscript. P. Richard Harrigan and Jeffrey B. Joy supervised the project.

The University of British Columbia Providence Health Care Research Ethics Board granted ethical approval for the BC Centre for Excellence in HIV/AIDS Drug Treatment Program (H05-50123), the population of study in Chapter Two.

At current time, Chapter Three has been submitted for publication. I (Kimia Kamelian) am the primary author and was responsible for performing laboratory experimental procedures, data analysis, and

manuscript composition. In detail, P. Richard Harrigan and I conceived of the study idea. Ryan H. Lapointe and I planned and executed the experimental procedures. Suzanne M. McCluskey and Jessica E. Haberer provided critical data. I contributed to genomic analysis and interpretation of the data. David Bangsberg was the original investigator of the UARTO cohort. I took lead in writing the manuscript. P. Richard Harrigan, Mark J. Siedner, and Jessica E. Haberer supervised the project. Jessica E. Haberer, Chanson J. Brumme, Yap Boum, Mwebesa Bosco Bwana, and Conrad Muzoora provided feedback on the manuscript. Institutional Review Boards at Partners Healthcare, University of California San Francisco, Mbarara University of Science and Technology, and Uganda National Council for Science and Technology granted ethical approval for the UARTO cohort, the population of study in Chapter Three. The University of British Columbia Providence Health Care Research Ethics Board granted ethical approval for the study presented in Chapter Three (H11-01642).

At this time, Chapter Four has been submitted for publication. I (Kimia Kamelian) am the primary author and was responsible for laboratory experimental procedures, data analysis, and composed the manuscript. In detail, co-authors Jeffrey B. Joy, P. Richard Harrigan and Muhammad Morshed conceived of the study idea. Andrea Olmstead gained ethical approval for the study. Andrea Olmstead, Winnie Dong, and I planned the experimental procedures. Mohammad Morshed provided the sample set. I carried out the experimental procedures. Vincent Montoya contributed to genomic analysis and interpretation of the data. Jeffrey B. Joy, Andrea Olmstead, and I contributed to the phylogenetic analysis of the data. I took lead in writing the manuscript. P. Richard Harrigan and Jeffrey B. Joy supervised the project. Vincent Montoya and Jeffrey B. Joy provided critical feedback on the manuscript.

The University of British Columbia Providence Health Care Research Ethics Board granted ethical approval for the study presented in Chapter Four (H16-02865).

Table of Contents

List of Tables

List of Figures

List of Abbreviations

3'-O-dNTP – 3'-O-azidomethyl 2'-deoxy-nucleoside triphosphate
3TC – lamivudine
ABC – abacavir
AIDS – acquired immune deficiency syndrome
APS – adenosine 5'-phosphosulfate
ART – antiretroviral therapy
ARV – antiretroviral
AZT – zidovudine
BC – British Columbia
BC-CfE – BC Centre for Excellence in HIV/AIDS
BCCDC – BC Centre for Disease Control
bp – base-pair
CD4 - Cluster of Differentiation 4
CDC – Centres for Disease Control and Prevention
CRF – circulating recombinant form
Ct – cycle threshold
d4T – stavudine
dATP – deoxyadenylate
dCTP – deoxycytidylate
ddATP – dideoxyadenylate
ddCTP – dideoxycytidylate
ddGTP – dideoxyguanylate
ddNTP – dideoxynucleoside triphosphates
ddTTP – dideoxythymidylate
ddI – didanosine
dGTP – deoxyguanylate
DNA – deoxyribonucleic acid
dNTP – deoxyribonucleoside triphosphates
DRM – drug resistance-associated mutation
DTG – dolutegravir
DTP – Drug Treatment Program
dTTP – deoxythymidylate
EFV – efavirenz
EVG – elvitegravir
FTC – emtricitabine
HCV – Hepatitis C Virus
HIV or HIV-1 – Human Immunodeficiency Virus type 1
HIVdb –HIV Drug Resistance Database
INSTI – integrase strand transfer inhibitor
kbp – kilo-base-pair
Lat – latitude
Long – longitude
NNRTI – nonnucleoside reverse transcriptase inhibitor
NRTI – nucleoside reverse transcriptase inhibitor
NVP – nevirapine
PCR – polymerase chain reaction
PI – protease inhibitor
PI-RT – protease inhibitor and reverse transcriptase inhibitor

PPi – pyrophosphate
RAL – raltegravir
RNA – ribonucleic acid
RT-PCR – reverse transcription-polymerase chain reaction
SMRT –Single-Molecule Real-Time sequencing
TDF – tenofovir
UARTO – Uganda AIDS Rural Treatment Outcomes cohort
VIDUS – Vancouver Injection Drug Users Study
WGS – whole-genome sequencing
WHO – World Health Organization
ZIKV – Zika Virus
ZMW – zero-mode waveguide

Acknowledgements

To my supervisor Richard Harrigan, I am incredibly grateful for your encouragement, support, and patience over the years. Thank you for allowing this thesis to be my own work, steering me in the right direction whenever I needed help, and for keeping a sense of humour, when I had lost mine. I am forever thankful for all of the valuable opportunities and experiences you have helped me gain.

To other members of my graduate supervisory committee, Jeffrey B. Joy, Chanson J. Brumme, and M-J Milloy, I am truly thankful for your valuable support and input. I would also like to acknowledge the co-authors and collaborators of these studies, some of whom were countries and continents away, for their guidance and their willingness to help.

I must express my sincere gratitude to the laboratory staff at the BC Centre for Excellence in HIV/AIDS for supporting me since undergrad and providing me with the knowledge and tools I needed to complete this work. I'd like to thank fellow students for their kindness over the years and specifically Hope Lapointe for, essentially, teaching me how to survive in a laboratory. I must also thank the research participants of these studies, without whom, this work would not have been possible.

And finally, I would like to thank my family, abroad and at home for their love and counsel. Most importantly, my Mom and Dad for their humility and sacrifices over the years. I am forever in your debt.

Dedication

To my Mom and Dad.

Thank you for your never-failing sympathies and unwavering support during my repeated episodes of uncertainty.

1 Chapter One: Introduction

1.1 Relevant Background

Massive-scale and high-throughput viral sequencing present an alternative method to conventional detection and surveillance of viral pathogens. The identification of nucleotide changes, resistance patterns, and transmission routes through sequencing methods can be less time-consuming and less labour-intensive than conventional methods, and provide highly accurate results [1].

1.1.1 History of Sequencing Technologies

Sanger sequencing

The quest to understand the underlying genetic make-up and mechanisms of any organism's genome is best exemplified with the initiation of the Human Genome Project in 1985, an international collaboration with the initiative to sequence the entirety of the human genome [2]. Utilizing derivatives of Frederick Sanger's sequencing methodology known as Sanger sequencing, the first human genome containing three billion nucleotides was sequenced in 2001 [3].

Sanger sequencing is a simple yet profoundly consequential first-generation sequencing methodology which has revolutionized the field of genetics. Sanger sequencing incorporates chain-terminating dideoxyribonucleotides called dideoxyribonucleoside triphosphates (ddNTPs) into extending nucleic acid molecules [4,5]. Dideoxyribonucleotides are chemical analogs of deoxyribonucleotides, deoxyribonucleoside triphosphates (dNTPs), the building blocks of deoxyribonucleic acid (DNA). Deoxyribonucleotides are composed of a pentose derivative called 2-deoxy-D-ribose, a nitrogenous base, and one or more phosphates (mono-, di-, or triphosphates) (Figure 1.1) [6]. The four major types of dNTPs include deoxyadenosine triphosphate (dATP), deoxyguanosine triphosphate (dGTP), deoxythymidine triphosphate (dTTP), and deoxycytidine triphosphate (dCTP) (Figure 1.2). During DNA replication, DNA

polymerase synthesizes DNA in the 5' to 3' direction, adding one of the four major dNTPs to the 3'-hydroxyl group of the growing, extending DNA fragment [6]. Unlike dNTPs, ddNTPs lack the 3'-hydroxyl group required for chain elongation, inhibiting a phosphodiester bond formation with the 5'-phosphate group of an incoming dNTP or ddNTP (Figure 1.3) [4,7]. During sequencing of DNA templates, ddNTPs are infrequently and randomly incorporated into growing DNA strands, resulting in fragments of different lengths each terminated by a single ddNTP. The concentration of ddNTPs required is typically a fraction of the nucleotides within a sequencing reaction [5]. Chain-terminating ddNTPs may contain external moieties such as radioisotopes or fluorophores which can be detected and the sequence of the DNA template can be inferred.

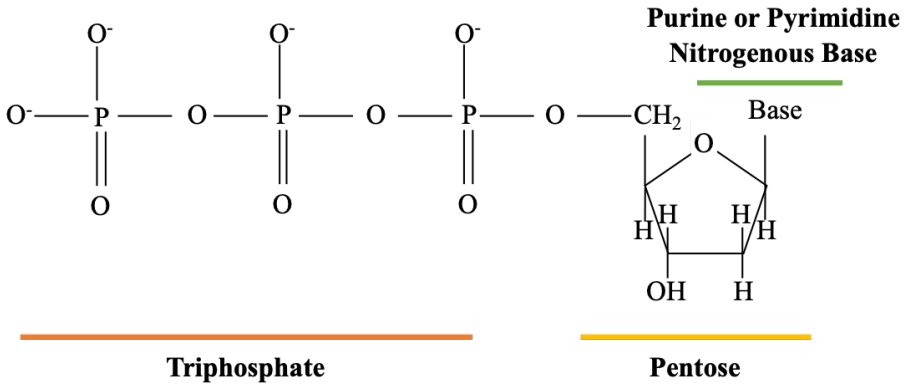


Figure 1.1 Deoxyribonucleoside triphosphate.
General structure of a deoxyribonucleoside triphosphate (dNTP) containing the three main components of any nucleotide: 1) deoxyribonucleotides are composed of a 2-deoxy-D-ribose, 2) nitrogenous base, and 3) phosphate group. (Image source: original creation by Kimia Kamelian)

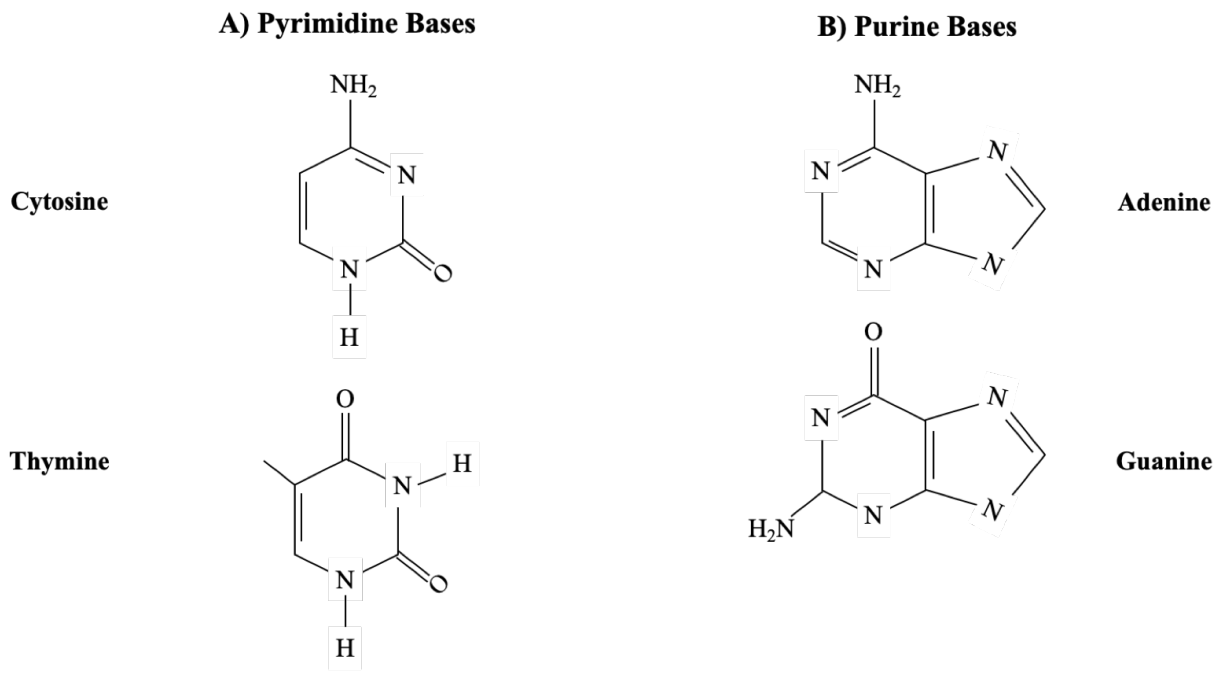


Figure 1.2 Nitrogenous bases.
The nitrogenous bases within DNA are derivatives of **A)** pyrimidines and **B)** purines. DNA contains two major pyrimidines, cytosine and thymine, and two major purines, adenine and guanine.
(Image source: original creation by Kimia Kamelian)

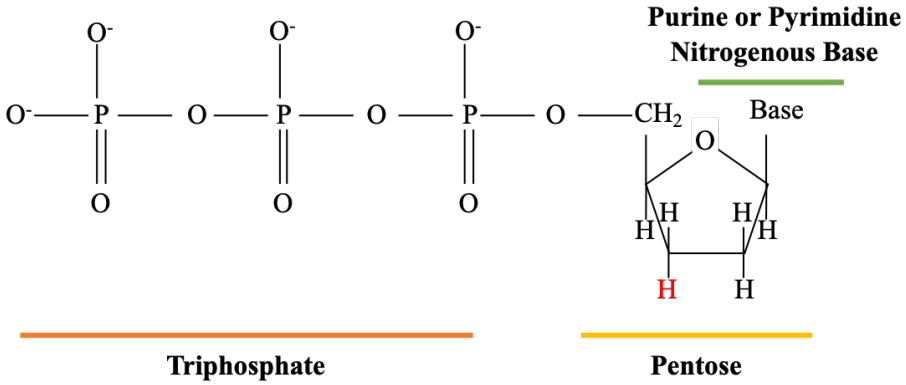


Figure 1.3 Dideoxyribonucleoside triphosphate.
In dideoxyribonucleoside triphosphates (ddNTPs) used in Sanger sequencing, the hydroxyl group on the 3' carbon of the pentose is replaced with hydrogen, inhibiting chain elongation through formation of phosphodiester bonds. The concentration of ddNTPs in a reaction are small relative to dNTPs. (Image source: original creation by Kimia Kamelian)

The original Sanger-sequencing method developed by Fredrick Sanger and colleagues [5] includes four parallel reactions, each reaction corresponding to the use of one of the four major types of ddNTPs: dideoxyadenosine triphosphate (ddATP), dideoxyguanosine triphosphate (ddGTP), dideoxythymidine triphosphate (ddTTP), and dideoxycytidine triphosphate (ddCTP). Each reaction comprises of a DNA template, DNA polymerase, buffer, primers, dNTPs, and radioisotope-labeled ddNTPs [8]. The resulting reactions are run through separate columns on a polyacrylamide gel and the sequencing products containing DNA segments of different lengths are separated (Figure 1.4). Autoradiography can be used to identify the radioactive bands in each ddNTP column, and the nucleotide sequence can be inferred [4].

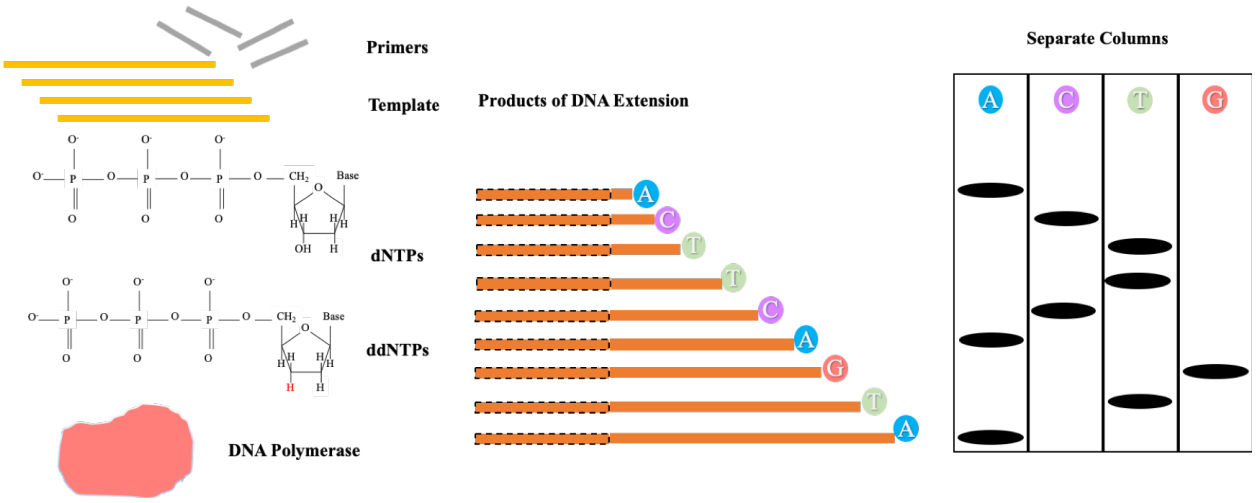


Figure 1.4 Sanger sequencing.
Template, primers, buffers, dNTPs, and DNA polymerase are mixed in four individual reactions corresponding to each of the four major ddNTPs (ddATP, ddTTP, ddGTP, ddCTP) added. DNA fragments of different lengths are generated. The reaction products are subject to gel electrophoresis and separated by length in the separate columns. (Image source: original creation by Kimia Kamelian)

A number of alterations and modifications have occurred to modernize Sanger sequencing. This modernization includes the utilization of non-radioactive fluorophore-labeled ddNTPs, permitting sequencing in a single reaction as opposed to four parallel reactions [4,8]. Advances in fluorescence technology have enabled ddNTP labeling with fluorophores of different dyes instead of a single radioisotope, allowing for mixing of ddNTPs within a single reaction. The use of automated

polyacrylamide-filled capillary electrophoresis has also increased the speed at which sequencing can occur by coupling efficient separation of DNA fragments with laser-induced fluorescence detection (Figure 1.5), making initiatives like the Human Genome Project possible [9,10].

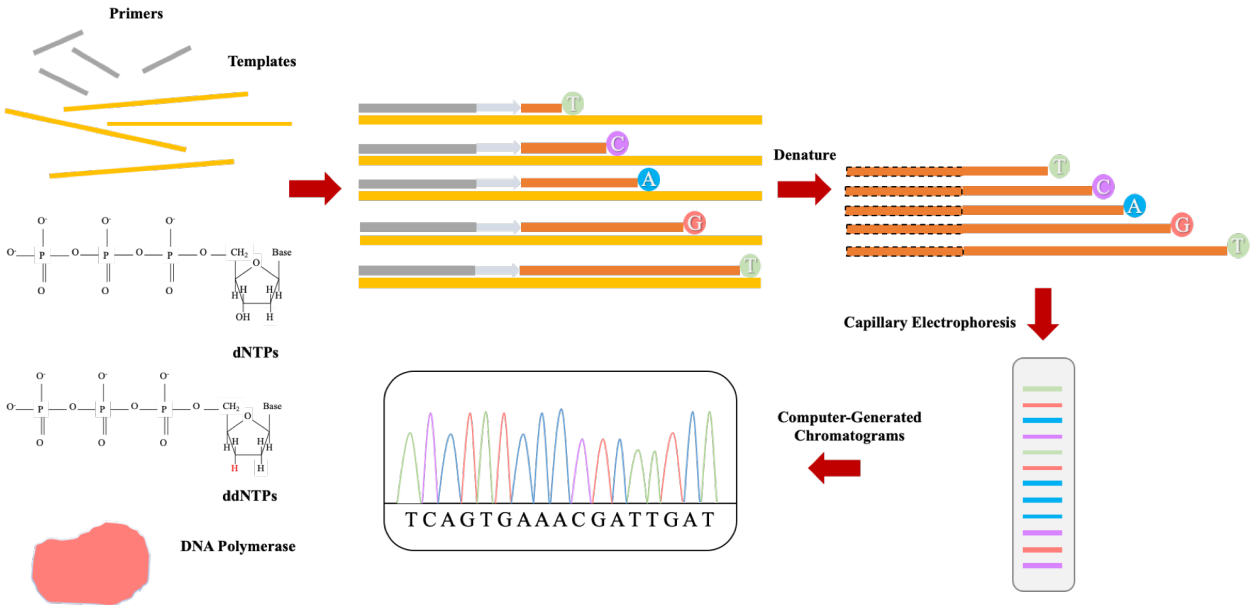


Figure 1.5 Automated Sanger sequencing.
Templates, primers, dNTPs, ddNTPs, and DNA polymerase are mixed causing extension of single-stranded DNA fragments of varying lengths in a single reaction. Double-stranded DNA molecules are denatured, strands are separated through capillary electrophoresis, and nucleotide bases are called through fluorophore detection by a laser. Computer-generated chromatograms indicate the sequence of nucleotides of the template DNA (Image amended from Nelson et al., 2013, p303-4 [6]).

Second-generation sequencing technologies

"Next-generation" sequencing refers to second-generation sequencing technologies that rely on an initial step involving the clonal amplification of isolated DNA molecules followed by massively-parallel sequencing (Figure 1.6) [8,11]. Although similar in concept to Sanger sequencing, next-generation sequencing allows for simultaneous clonal amplification of multiple DNA templates and subsequent sequencing of millions of template fragments within a single reaction. Sequencing occurs when thousands to millions of amplified single-stranded DNA molecules are attached to solid surfaces and often, a DNA polymerase replicates and extends the template DNA through the production of a complementary strand [11]. The resulting sequences from the clonally amplified template fragments range in base-pair (bp) length

(<700 bp) and are referred to as reads, which are overlapped to produce contiguous DNA segments called contigs. Contigs are assembled sequentially to form longer DNA segments called scaffolds and either through reference-based mapping or de novo contig assembly, produce read-based consensus sequences [8]. Although next-generation sequencing platforms encompass similar methodological concepts, they may differ in their template amplification method, sequencing chemistry, sequence length, cost, accuracy, and overall utility. For simplicity, three previous and current commonly-used next-generation sequencing technologies to date will be reviewed. These technologies include: Roche 454 pyrosequencing introduced in 2005 and discontinued by 2016, Ion Torrent sequencing through non-optical sensing from Thermo Fisher Scientific introduced in 2010, and Illumina/Solexa sequencing by Illumina introduced in 2006 [8,12].

The pyrosequencing technology [13] offered by Roche relied on oil-aqueous emulsion polymerase chain reaction (PCR) for its initial clonal template amplification to generate single-clone amplified beads (Figure 1.6A) [14]. The beads were then distributed and fixated onto separate single wells on a fiber optic plate called PicoTiterPlate^{TM} [8,14]. In addition to DNA polymerase, enzymes including adenosine triphosphate (ATP) sulphurylase, luciferase, apyrase, substrates adenosine 5'-phosphosulfate (APS) and luciferin were required for the sequencing reaction to take place. Sequencing occurred during repetitive cycles involving the addition of one dNTP at a time [15]. Each dNTP cycle was supplied by fluidic assembly and was incorporated into growing DNA strands with the use of DNA polymerase. During the incorporation of each dNTP a pyrophosphate (PPi) was released, and initiated a series of reactions to produce light [8,15]. The release of PPi was in a quantity equimolar to the number of dNTPs incorporated in each cycle [15]. Sulphurylase converted PPi to ATP in the presence of APS. The ATP facilitated a luciferase-mediated conversion of luciferin to oxyluciferin producing light detectable by a charge coupled device [8,15]. Lastly, apyrase degraded remaining dNTPs and ATP before the start of the next cycle [8,15].

The Ion Torrent platform from Thermo Fisher Scientific utilizes non-optical sensing sequencing methodology and detects a change in pH during sequencing [16]. The Ion Torrent system depends on bead-

based emulsion PCR for clonal amplification of DNA templates, similar to pyrosequencing technology offered by Roche and bead-bound DNA templates are located in single reaction sensor wells on specialized Ion Torrent chips [1]. Subsequently, sequencing occurs and released hydrogen ions are detected as dNTPs incorporate into growing DNA strands [1]. All four nucleotides are released in stages such that a single nucleotide is added at a time. After a single dNTP is incorporated, a hydrogen ion is released, detected through a sensor, and unused nucleotides are washed away [16]. The change in pH of the solution is proportional to the number of incorporated dNTPs and is determined by a sensor involving a complementary metal-oxide-semiconductor (CMOS), used for constructing integrated circuits, and an ion-sensitive field-effect transistor (ISFET), used for electrochemical detection of hydrogen molecules [16].

Over recent years, Illumina has become the dominant supplier of sequencing technologies. Relative to pyrosequencing and non-optical sensing sequencing, Illumina sequencing is less susceptible to homopolymer errors which occur during the incorporation of nucleotides within regions of repeated nucleotide bases [1]. Sample library preparation for Illumina sequencing generally starts with the fragmentation of large genomic DNA into smaller pieces or targeted amplification of amplicons [17,18]. Adapters are added to the ends of all DNA fragments and through reduced cycle amplification, additional motifs are introduced which include sequencing primer binding sites, indices, and regions complementary to oligos fixed to the surface of a glass slide called the flow cell. Used for isothermal bridge PCR (a technique involving amplification of DNA templates embedded on solid surface) and DNA cluster formation on Illumina platforms, the solid flow cell contains two types of covalently-attached oligos that are complementary to single strands of template DNA (Figure 1.6B) [17]. After initial hybridization to the flow cell, bridge amplification and cluster formation ensue [8,18]. The DNA polymerase creates a complimentary strand of the single template strand to produce covalently-attached DNA molecules on the flow cell [18]. The strands are then denatured and the original strand is washed away. The remaining strands are simultaneously amplified through bridge amplification where the 3’ end of the bounded single-stranded fragment folds and hybridizes with nearby complementary oligos, creating a bridge [8]. The DNA

polymerase then creates a complementary strand forming a double-stranded bridge which is then denatured and this PCR cycle is repeated, creating millions of clusters of clonally amplified DNA molecules [8]. Half of the templates are cleaved from the flow cell, and sequencing begins with the annealing of sequencing primers and a single nucleotide base extension involving the incorporation of fluorophore-labeled, reversible terminating dNTPs called 3'-O-azidomethyl 2'-deoxynucleoside triphosphates (3'-O-dNTPs). The reversibility of 3'-O-dNTPs is obtained by capping the 3'-hydroxyl group of dNTPs with a small reversible molecule, making 3'-O-dNTPs capable of still being recognized as a substrate by DNA polymerase [17,19]. After a single 3'-O-dNTP is incorporated into a growing, extending strand, the identity of the fluorophore-labeled 3'-O-dNTP is obtained through laser-induced excitation of the fluorophores and imaged. Next, the fluorophore is removed and the 3'-hydroxyl group on the incorporated nucleotide is regenerated. This extension, termination, cleavage process is repeated and, the fluorescent signals and intensity corresponding to nucleotide bases incorporated in each cluster are identified by fluorescence imaging to determine the sequence of the DNA fragments [17].

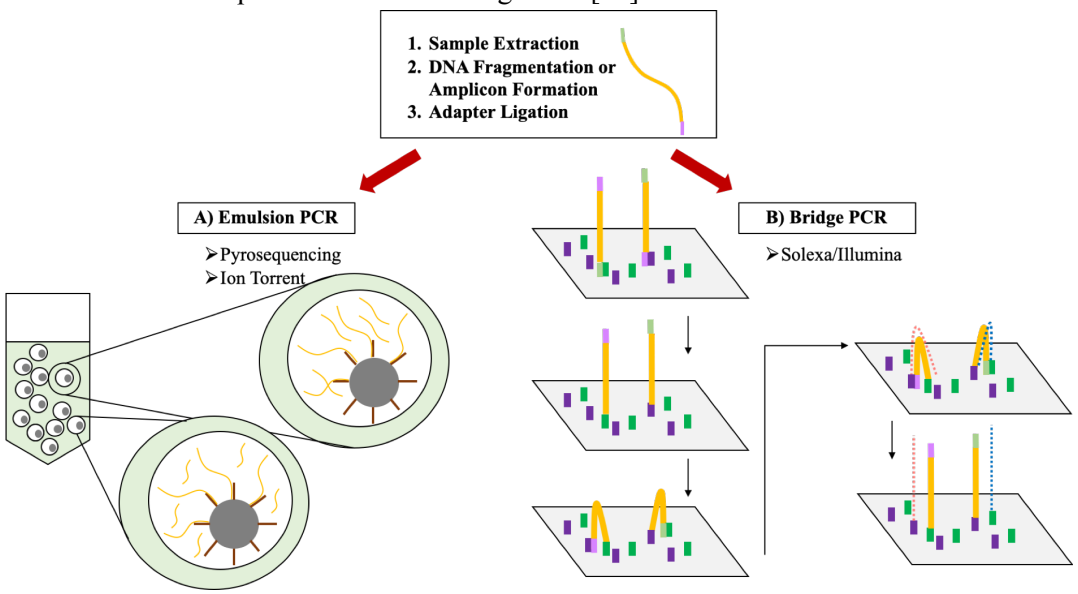


Figure 1.6 Clonal amplification involved in second-generation sequencing technologies.
A) Emulsion oil and PCR solution with beads and DNA library are mixed. Emulsions are broken and solution is vortexed, centrifuged, and magnetically separated. **B)** DNA strands are initially attached to flow cell surface, amplified, and the original strands washed away. Remaining complementary strands create a bridge to complementary oligos and are extended, creating a double strand. Denaturation creates two separate strands of DNA molecules. Repetition creates clusters of identical strands with identical identifiers, corresponding to the reads of a template. (Emulsion PCR figure amended from Vierstraete, 2018 [20] with permission from the author; Source of additional figures: original creation by Kimia Kamelian)

Third-generation sequencing technologies

Third-generation sequencing is characterized as single-molecule sequencing. Unlike second-generation sequencing, single-molecule sequencing does not rely on clonal amplification of template fragments prior to massively-parallel sequencing. Instead, it is able to obtain a sequence from a single template fragment, and is not hindered by PCR-based artifacts [1,21]. Third-generation sequencing platforms are also capable of sequencing longer fragments, potentially reading several continuous kilo-base-pairs (kbp) of template in real time [1,22,23]. Long-read sequences produced by single-molecule sequencing have several useful advantages including the ability to decipher the complexity of genomes that have large repetitive sequences, which is currently a barrier for short-read sequencers. Two current well-known and commonly used single-molecule sequencing technologies include Single-Molecule Real-Time sequencing by Pacific Biosciences introduced in 2011 [22], and nanopore sequencing by Oxford Nanopore Technologies introduced in 2014 [23].

Single-Molecule Real-Time (SMRT) sequencing by Pacific Biosciences occurs in specialized nanophotonic visualization chambers called zero-mode waveguide (ZMW) located in wells at the bottom of specialized SMRT flow cells (Figure 1.7) [21,22]. Each well consists of a template molecule, sequencing primer, and a DNA polymerase bound to the bottom of the well. The small confinement of each well permits single-fluorophore detection of incorporated dNTPs in extending DNA fragments [21]. Unlike other sequencing technologies which use fluorophores linked to the nitrogenous base of dNTPs, SMRT utilizes phospholinked dNTPs with fluorophores linked to the terminal phosphate molecule [22]. The extension of DNA fragments with the incorporation of phospholinked dNTPs occurs through phosphodiester bond formation by DNA polymerase, which cleaves the fluorophores from the dNTP during incorporation. Using a high-multiplex confocal fluorescence detection system, the fluorophores emitted and the duration of emitted light are able to be detected simultaneously from each well on a flow cell [1,22]. Through exploitation of the high catalytic rate of the stationary DNA polymerase and photo-protected nucleotide analogs which shield the polymerase from damage, sequencing of long-read fragments is enabled [21].

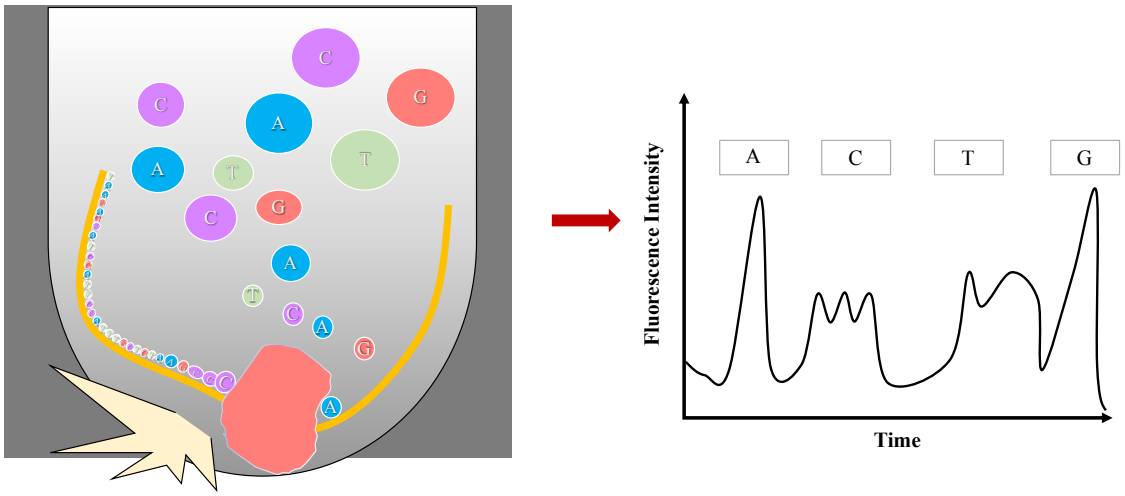


Figure 1.7 Single-Molecule Real-Time sequencing.
Single-Molecule Real-Time (SMRT) sequencing occurs in zero-mode waveguide (ZMW) chambers, nanophotonic structures capable of detecting single incorporations of fluorophore-tagged dNTPs in extending DNA segments by an immobilized DNA polymerase. The fluorophore emitted and the duration of the emitted light are identified by a fluorescence detection system. (Image amended from Eid et al., 2009, p134 [22] with permission from The American Association for the Advancement of Science, Licence: 4611510673553)

Nanopore sequencing by Oxford Nanopore Technologies is dependent on the use of electrically resistant membrane-embedded nanopores: biological, synthetic, or hybrid channels of nanometer length [8]. Unlike many other sequencing platforms, nanopore sequencing does not monitor the incorporation of fluorophore-labeled nucleotides into extending DNA fragments. Instead, it evaluates the nucleotide sequence of templates by detecting the change in voltage generated as a single strand of template DNA is passed through a nanopore while an ionic current flows through the aperture (Figure 1.8) [1]. During initiation of sequencing, double-stranded templates are directed to nanopores by template-ligated leader adapter sequences and motor proteins, and subsequently denatured into single-stranded template [24]. Single-stranded template is translocated through the nanopore as the current is passed through the pore causing disruptions in the current. The observed shift in voltage and its temporal magnitude are observed and recorded as a particular combination of nucleotides called a "k-mer" [1]. Rather than having four fluorescently distinct signals pertaining to the incorporation of dNTPs, nanopore sequencing has more than 1000 distinct signals – for each possible resulting k-mer. Base-calling occurs on 5-mers (five nucleotides:

4^5=1024 possible combinations) and 6-mers (six nucleotides: 4^6=4096 combinations) [24]. A template-ligated hairpin adapter also allows for the sequencing of the complementary strand, producing the sequence of both strands of template DNA.

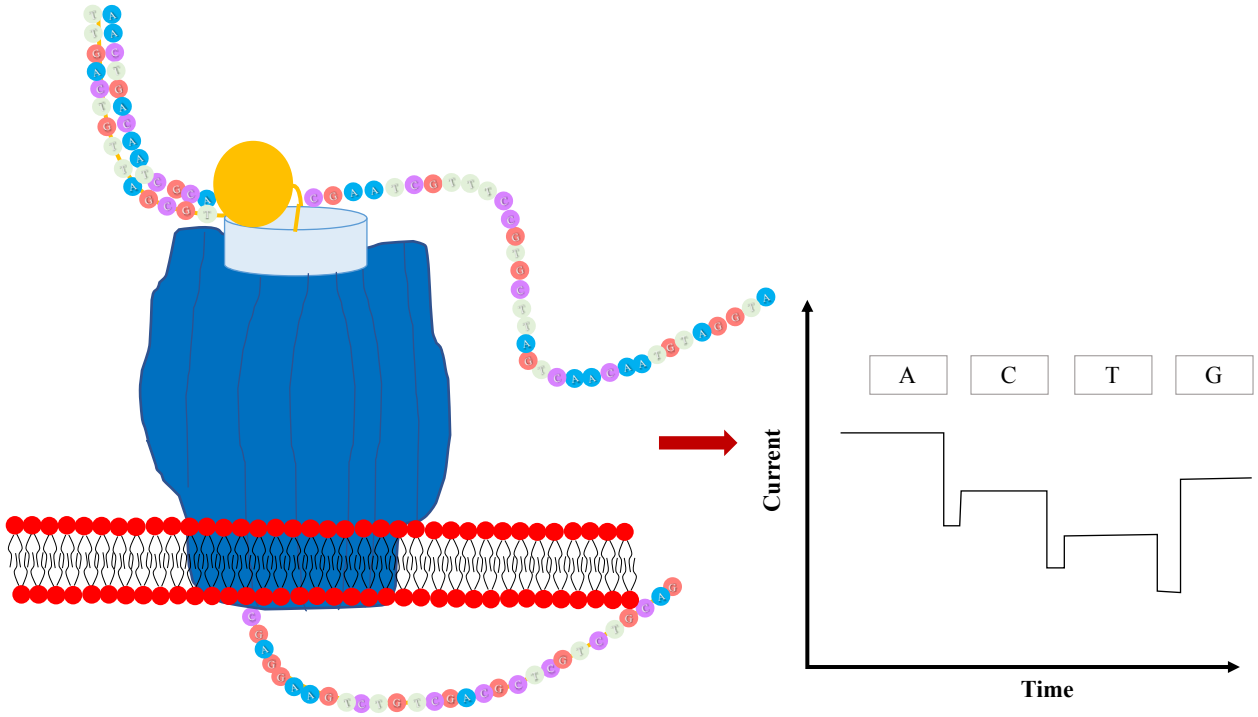


Figure 1.8 Nanopore sequencing.
Sequencing initiates as motor proteins unwind a double-stranded DNA molecule and a single strand passes through a nanopore. As nucleotides travel through the pore, a change in magnitude of the current passing through the pore is detected by a sensor and recorded in real time. Base-calling is performed on 5-mers or 6-mers. (Image amended from Lu et al., 2016, p266 [24] with permission under the terms of the Creative Commons Attribution License (CC BY))

Comparison

First- and second-generation sequencing technologies have enabled large-scale identification of novel and evolving pathogens [25–27], drug resistance testing [28–30], and discovery of nucleotide variants associated with various diseases [31,32]. However, the resulting short fragments of these sequencing platforms remains a limitation in the field of genetics considering the large sizes of bacterial and human genomes [33]. Sanger sequencing results in sequences <1000 bp while second-generation sequencing technologies result in sequences <700 bp (Table 1). Although third-generation sequencing technologies aim

to tackle this issue, their high error rate and low accuracy present additional complications [24]. A potential solution, that addresses the low accuracy of third-generation sequencing platforms and is gaining support, is the use of circularized DNA fragments to continuously sequence a single template to essentially produce long-reads which can then be assembled into a uniform, highly accurate, consensus sequence [1]. In addition, paired-end sequencing enables the sequencing of both ends of a DNA template and is often used during second- and third-generation sequencing for better reference-based alignment and read-based consensus sequence assembly. Compared to first-generation Sanger sequencing, second- and third-generation sequencing technologies can make genetic sequencing less time-consuming and more cost-effective (Table 1.1). However, the amount of raw data generated is vast and the technical assistance required to produce the finalized sequences is commonly an obstacle (Table 1) [1]. While sequencing technologies have evolved tremendously to be used in clinical and research settings, their utility varies based on their applications. For example, third-generation sequencing technologies may be best suited for transcriptome and large-genome sequencing due to their long-read capabilities, while targeted or amplicon sequencing can occur in an accurate and cost-effective manner on first- or second-generation sequencing technologies. Furthermore, second- and third-generation sequencing have lower limits of detections and can detect minority species nucleotide substitutions, potentially significant in identifying drug resistance-associated mutations. Choosing among different methods depends on the purpose of sequencing and the trade-offs in accuracy, speed, and cost [34].

Table 1.1 Overview of first-, second-, and third-generation sequencing technologies.

Platform	Method Chemistry	Run Time	Read Length (base-pair)	Output per Run	# Reads per Run	Reagent Cost	Error Rate (%)	Advantages	Disadvantages	References
ABI 3730xI	Sanger dideoxy-chain termination	2-3 hours	650	0.03Gb	N/A	$1500 per megabase	0.001-1.0	Low error rate; Long read length	Low throughput; High reagent cost	Glenn 2011; Harrison and Kidner, 2011
454 Roche	Pyrosequencing	23 hours	300-700	700 Mb	1 million	$7.00-$12.00 per megabase	1.0	Fast run time; Long read length	High reagent cost; High error rate (homopolymer errors)	Buermans et al. 2014; Metzker et al. 2009; Glenn 2011
Thermo Fisher Scientific/ Ion Torrent	Hydrogen ion detection	4 hours	200-400	1.5-2 Gb	4 million	$500-$1000 per run	~1.0	Fast run time; Long read length	High reagent cost; High error rate (homopolymer errors)	Buermans et al. 2014; Glenn 2011
Illumina	Reversible terminators	9 hours-4 days	300-600	1.2-600 Gb	4 million-20 billion	$0.04-$0.74 per megabase	0.01-3.8	High-throughput; Short run time; Various read lengths	Specialized analyses required for massive amount of data	Buermans et al. 2014; Illumina, Inc. 2019; Glenn 2011
Pacific BioSciences	Single-Molecule Real-Time (SMRT) sequencing	0.5-2 hours	10-15 kilo-base-pairs	1-2 Gb per SMRT flow cell	0.01 million	$11-$180 per megabase	15	Long read length	High cost; Limited throughput	Glenn 2011; Goodwin et al.2016; Lu et al.2016
Oxford Nanopore Technologies/ MinION	Sequencing by current detection	1 minute -72 hours	Hundreds of kilo-base-pairs->2 mega-base-pairs	10-30 Gb per flow cell	0.80 million	$99.00 per run	15-30	Small device and portable; Sequencing in real-time	Large error rate; Large RAM and hard desk space needed; High cost	Goodwin et al.2016; Lu et al.2016; Oxford Nanopore Technologies 2018; Tyler et al. 2018

1.1.2 Impact of Sequencing Technologies on Our Understanding of Viral Pathogens

Detection and diagnostics

In the field of infectious diseases, investigation of viral genomes facilitates efficient clinical diagnosis. Early prognosis and clinical management are beneficial to patients and rely on early detection of viral infections. Correct diagnosis of viral infections is clinically significant because viral infections can increase susceptibility to secondary viral and bacterial infections, resulting in an array of severe health-related complications [35]. For example, the influenza virus can result in secondary meningitis infection, while the herpes simplex virus can result in encephalitis [36,37].

Conventional diagnostic methods for viral infections include culture-based methods, targeted molecular testing, and immunoassay-based tests. However, these methods are limited by their hypothesis-driven circumstances as well as vary in their scope and sensitivity, and can be expensive, labor-intensive and time-consuming [37,38]. Culture-based methods can be non-specific and are restricted by their time requirements, targeted-molecular testing such as PCR may use primers that mismatch the strain of virus, and immunoassay-based testing may be cross-reactive among a virus family or genus and are prone to interferences [38]. Therefore, the aetiology of suspected infections can often be left undiagnosed [39].

Sequencing of viral pathogens provides an alternative methodology in confirming a viral diagnosis [38]. Because of the multiplex capabilities of current sequencing technologies, a large number of samples may be tested at a single time and in a cost-effective manner, appropriate for outbreak scenarios [40,41]. Although prior knowledge about potential unique viral exposures (hypothesis-driven) is an advantage in detecting viral pathogens through sequencing methodologies, often allowing for more targeted sequencing, it is not a necessity. This is because metagenomic sequencing, a near hypothesis-free diagnostic approach, sequences all nucleic acid content within a patient's biological sample [37]. Its unbiased approach can also be coupled with conventional diagnostic methods by narrowing the suspected viral pathogen through

identification of the viral family or species, which can then be precisely tested using molecular or immunoassay approaches.

Transmission

Sequencing of viral pathogens has enabled increased resolution of disease surveillance and enhanced the identification of transmission pathways. Surveillance of viral transmission is important in the development of public health guidelines to prevent viral outbreaks, and to ensure rapid diagnosis and adequate treatment are provided. Assessing the genomic diversity through transmission links can estimate the speed at which a virus is spreading and its future course [42].

Identifying new viral lineages, distinguishing between vaccine and wildtype strains of suspected viral infections, and linking imported cases of viral infections to international endemic areas are a few examples of how viral sequencing has contributed to our understanding of viral transmissions [43,44]. In 2015 during the Ebola virus epidemic in Guinea, a resource-limited setting, nanopore sequencing of the Ebola virus genome by scientists provided a quick, in-depth view of the virus's evolution and identified transmission links important in the viral spread [40]. Furthermore, in 2010 during the XXI Olympic Winter Games in Vancouver BC, Canada, a measles virus outbreak involving more than 80 confirmed cases, ensued. Although not done in real time, whole-genome sequencing of the measles virus and subsequent phylogenetic analysis revealed transmission routes, and the outbreak was determined to be the result of two independent importations of the measles virus from Olympic visitors travelling from measles endemic regions [45]. In addition, phylogenetic analysis of HIV transmissions in a Zambian and Rwandan cohort found distinct HIV sub-viral populations in the gastrointestinal tracts of infected individuals and further revealed their preferential transmission to sexual partners [46,47].

Drug resistance

The development of sequencing technologies has enabled applications that impact our fundamental understanding of the genetics of various pathogens, including their ability to evade the immune system and respective pharmaceutical treatments. Most viral pathogens have large population sizes, high replication capabilities, and high nucleotide substitution or mutation rates, resulting in high degrees of genetic diversity [48]. These factors are fundamentally important in concepts such as natural selection where, under different selective pressures certain virus particles are more likely to evade treatment and replicate [49]. In the areas of precision and personalized medicine, identification of the infecting viral strain and its nucleotide composition becomes significant in determining the course of treatment. Specifically in cases of rapidly evolving pathogens, such as human immunodeficiency virus (HIV) and hepatitis C virus (HCV), substitutions in the genome may confer resistance to treatments [50–52]. Drug resistance profiles and genotyping can occur through sequencing of viral particles present in patient biological samples and can be used to determine treatment susceptibility using genotypic drug resistance interpretation algorithms [51,53].

Surveillance of transmission of key viral nucleotides associated with drug resistance and immune evasion provide personalized clinical context as to why individuals are not responsive to treatments and whether a change in treatment regimen is required. The identification of sporadic resistance cases can warn clinicians of precautions that may need to be taken before the administration of specific pharmaceutical drugs. For example, in 2015, emergent drug resistance to antiretrovirals (ARVs) in the class of integrase strand transfer inhibitors (INSTIs), used to treat HIV caused the BC Centre for Excellence in HIV/AIDS (BC-CfE) in Vancouver BC, Canada to recommend additional clinical guidelines to be followed during the prescription of INSTIs in BC [54].

Moreover, with the introduction of newer therapies, surveillance of drug resistance is important because of the lack of knowledge associated with their genetic barrier to resistance. This is further evident in the case of INSTIs, the most novel class of HIV ARVs to date. After its introduction in 2007, there was

an explosive increase in INSTI prescriptions and INSTIs are now recommended as components within first-line HIV regimens [55]. As a result, potentially due to suboptimal (<80%) adherence and lack of resistance testing, the prevalence of INSTI resistance-associated mutations is continuously increasing in various populations [56,57].

1.1.3 Interpretation of Sequencing Data in the Epidemiology of Viral Pathogens

The mechanisms involved in the interpretation of sequencing data are crucial to epidemiological inferences of viral pathogens. The appropriate computational systems and tools required to analyze viral sequences is dependent on the pathogen sequenced and the initial purpose of sequencing. Moreover, interpretation of sequencing data often involves complex computational systems, and reproducibility and accessibility to relevant frameworks is a concern in resource-limited settings [58].

Drug resistance interpretation

Identification of substitutions which may confer drug resistance can reveal patient treatment history and, when performed in large populations, can expose the magnitude of drug resistance present in a community giving way to appropriate public health responses. This is best exemplified in the case of HIV. A series of commercially and freely accessible tools are available for the analysis of Sanger and next-generation HIV sequencing data allowing rapid incorporation of sequencing results into clinical practice. Automated Sanger sequence analysis tools such as RECall can assemble and align template reads into consensus sequences, and identify positions with two or more nucleotides (mixture nucleotides) within a user-friendly interface [59]. Raw chromatogram files can be processed using RECall. Analysis of next-generation sequencing data and production of consensus sequences involves an initial alignment of individual template reads to reference genomes using programs such as BWA-MEM [60], followed by sorting, merging, indexing and subsequent variant calling using programs such as Samtools [61]. Variants and substitutions can be identified using different detectable thresholds. Consensus sequences can be interpreted using genotypic drug resistance interpretation algorithms. Publicly-accessible resistance

interpretation programs include the geno2pheno program [62] and the Stanford University HIV Drug Resistance Database (HIVdb) Genotypic Resistance Interpretation Algorithm [53,63]. The HIVdb reports drug penalty scores for individual and combinations of mutations while the geno2pheno system uses two machine-learning algorithms: decision trees and support vector machines [62]. Interpretation of drug resistance-associated mutations and determining levels of resistance rely on three main types of data: 1) genotype-treatment, 2) genotype-phenotype, and 3) genotype-outcome [53]. Correlations between mutations identified in viral genomes extracted from patients and antiretroviral treatment produce genotype-treatment data and are based on Darwinian selection. Genotype-phenotype data is dependent on correlations which quantify the effect of mutations through in vitro drug susceptibility experiments. Correlations between mutations identified in viral genomes extracted from patients prior to initiation of treatment and virological response produce genotype-outcome data [53]. It is important to note that while there are publicly-accessible automated pipelines available for sequencing analysis of next-generation sequencing data, including MiCall [64] and HyDRA [65], they are independently developed with no specific consensus guidance and their results may differ [66].

Phylogenetic analysis

Phylogenetic analysis of sequencing data can reveal transmission events and characterize genetic variations. Transmission patterns can be inferred from phylogenetic tree topology as well as nucleotide evolutionary analysis [67]. For example, the emergence of the Zika virus (ZIKV) in Central America and Mexico in 2014 was determined, through temporal and spatial phylogenetic analysis, to be the result of one introduction of ZIKV from Brazil [68]. While there are different types of phylogenetic trees including radial trees, cladogram trees, and phylogram trees, the most informative tree type is a phylogram because it displays both evolutionary direction as well as genetic distance [69]. During phylogenetic tree construction, different methods are used to infer evolutionary relationship between viral sequences. Two commonly used methods include maximum parsimony, which states that the best phylogenetic tree is the tree with the minimum number of evolutionary changes, and maximum likelihood, which considers the evolutionary

relationship between all of the sequences and identifies the most likely tree topology [69]. There are commercial and freely publicly-accessible software packages available for phylogenetic and evolutionary analysis of viral sequences. Viral sequences must be aligned before a phylogenetic method can be applied. Popular sequence alignment programs are MUSCLE [70] and ClustalW/ClustalX [71]. To construct phylogenetic trees, Randomized Axelerated Maximum Likelihood (RAxML) can be used to constructs maximum likelihood phylogenies based on large numbers of sequences [72]. The Molecular Evolutionary Genetics Analysis (MEGA) software provides a maximum parsimony method of phylogeny construction and also provides comprehensive visualizing tools [73].

1.1.4 Current Examples of Viral Pathogens often Sequenced in Clinical Settings

Human immunodeficiency virus

In 1981, the Centres for Disease Control and Prevention (CDC) reported several opportunistic infections and a rare, unusually aggressive form of cancer in patients in California and New York, documenting for the first time, acquired immune deficiency syndrome (AIDS). By years end, nearly 40% of reportedly infected individuals (130/337) had died. The ribonucleic acid (RNA) retrovirus that became known as human immunodeficiency virus type 1 (HIV-1) was identified in 1983 as the agent that causes AIDS [74]. Soon afterwards, the main modes of transmission were determined and diagnostic tests confirming infection were created. The first ARV (zidovudine, also known as azidothymidine) targeting HIV's reverse transcriptase enzyme was approved in the United States of America in 1987, and the efficacy of combination ARV therapy (ART) was demonstrated in 1995 [75,76].

Even with the availability of potent ARVs, HIV drug resistance has become a significant concern in the treatment of HIV regardless of ARV class [77,78]. The lack of proofreading capability of the HIV reverse transcriptase enzyme leads to a high error rate leading to nucleotide substitutions [79,80]. Under ARV selective pressure and incomplete viral suppression, certain nucleotide substitutions capable of conferring resistance can be selected, causing ARV resistance [77,81]. Drug resistance threatens the long-

term efficacy of ART in terms of clinical and treatment outcomes [82,83]. The current gold standard of HIV genotypic resistance testing involves Sanger sequencing followed by analysis of HIV viral sequences by a genotypic drug resistance interpretation algorithm [53,84].

Hepatitis C virus

Hepatitis C virus (HCV) is the leading cause of chronic liver disease worldwide [85]. The RNA virus is a bloodborne pathogen, capable of causing acute and chronic infections. Acute HCV infection is typically asymptomatic, with many infected individuals clearing the virus within six months without treatment [85]. However, 60-80% of acute HCV infections progress to chronic HCV infections, often leading to cirrhosis and hepatocellular carcinoma [85,86]. Currently, there are more than 70 million individuals globally living with chronic HCV infections [85].

HCV is incredibly diverse due to its error-prone RNA-dependent RNA polymerase, a result of its lack of proofreading activity [80]. In the case of HCV, the World Health Organization (WHO) recognizes six genotypes (1, 2, 3, 4, 5, 6) and more than 50 subtypes [85]. Mainstream treatment for HCV has shifted from pegylated interferon and ribavirin combinations to the use of direct acting antivirals that target key components of the HCV replication cycle [87]. Identification of HCV genotype is important in the duration and cost of HCV treatment. Treatments can be genotype-specific and individuals can be infected with more than one strain [85]. Nevertheless, recent introductions of combination pangenotypic direct-acting antivirals are effective across all genotypes [88]. Besides genotypic classifications of HCV subtypes, identification of resistance-associated mutations to direct-acting antivirals is also a concern to healthcare providers [89]. HCV genotyping, referring to the identification of the HCV strain or subtype, and identification of resistance-associated mutations may occur through sequencing methods but more commonly occur through kit-based genotyping assays which do not require sequencing expertise [87,89].

Zika virus

The Zika virus (ZIKV) is an arthropod vector-borne RNA virus mainly transmitted through *Aedes* mosquitos [90]. Most commonly, ZIKV infections are asymptomatic or associated with non-specific symptoms such as fever, headache, and conjunctivitis. However, recent sporadic outbreaks around the world have been associated with increased cases of microcephaly and neurodegenerative disease in infants, and Guillain-Barré syndrome in adults, demanding increased public health attention [90]. Similar to HCV, the ZIKV genome has an error-prone RNA-dependent RNA polymerase but unlike HCV, treatment options are currently limited and the spread of ZIKV is mainly reliant on the use of vector control methods [91].

The emergence of ZIKV in the 21st century, capable of causing severe disease with greater geographical range than prior ZIKV infections, has been hypothesized and confirmed through sequencing analysis to partially be a result of nucleotide substitutions in structural and non-structural domains important for vector- and host-infectivity as well as neuropathogenicity [90,92]. While ZIKV outbreaks prior to 2010 have been endemic to Africa and certain countries in Asia, the risk of imported cases of ZIKV into regions with competent vectors, such as South America, and the risk of sexual transmission in regions without competent vectors, such as North America, is a concern [91]. Current serological techniques for ZIKV diagnosis may be impaired by cross-reactivity with similar viruses within ZIKV's genus *Flavivirus* and secondary confirmation by ZIKV-specific PCR is often required [91]. Sequencing of ZIKV may provide an alternative diagnostic tool for the confirmation of ZIKV infection and identification of transmission pathways and origins of infection [93].

1.2 Importance

1.2.1 Variables Significant in the Increasing Prevalence of Viral Pathogens

Although modern medicine has created antibiotics, antivirals, and vaccines to treat known infectious diseases, it must also be prepared for sporadic, resurging, and novel viral pathogens and their

increase in geographical spread [94]. There are several key parameters that are significant in the emergence of diseases in modern society. The epidemiological triad is the classic epidemiological approach to describe the disease process [95]. This model addresses the aetiology of diseases and involves three main entities: an agent capable of causing disease, a susceptible host, and an environment which permits interaction between agent and host. Factors such as viral strain influence an agent's ability to cause disease [96] while factors such as immune-status impact susceptibility of hosts [97].

Furthermore, environmental factors such as poverty, conflict, and environmental changes can facilitate the transmission of infectious diseases. For example, diseases have the propensity to flourish in poverty settings due to poor quality housing and close living quarters [98]. Poverty-stricken areas also lack adequate sanitation systems and can contain large standing bodies of water, prime breeding grounds for disease-causing vectors such as mosquitos (Figure 1.9). The largest Ebola virus outbreak occurred in 2014 in Guinea, Liberia, and Sierra Leona [98,99] and poverty was determined to be a significant driver in its transmission and spread, associated with population-dense, low sanitation regions [100]. Political destabilizations, leading to civil and international unrest can cause massive displacement of people and cause the dismantlement of healthcare infrastructures [98]. The second-largest Ebola virus outbreak is currently occurring in an active war zone in the Democratic Republic of Congo, and ongoing conflicts are interfering with healthcare response and surveillance attempts (Figure 1.9) [101]. The number of emerging diseases are analytically shown to be significantly increasing over time [102]. Mosquito abundance, geographical distributions, and transmission ability is dependent on weather patterns and has thrived in tropical and subtropical climates, such as Brazil. Warmer climates can maintain mosquito larvae for longer periods of time, and warmer global climates can expand the geographical habitat range of *Aedes* mosquitos [103].

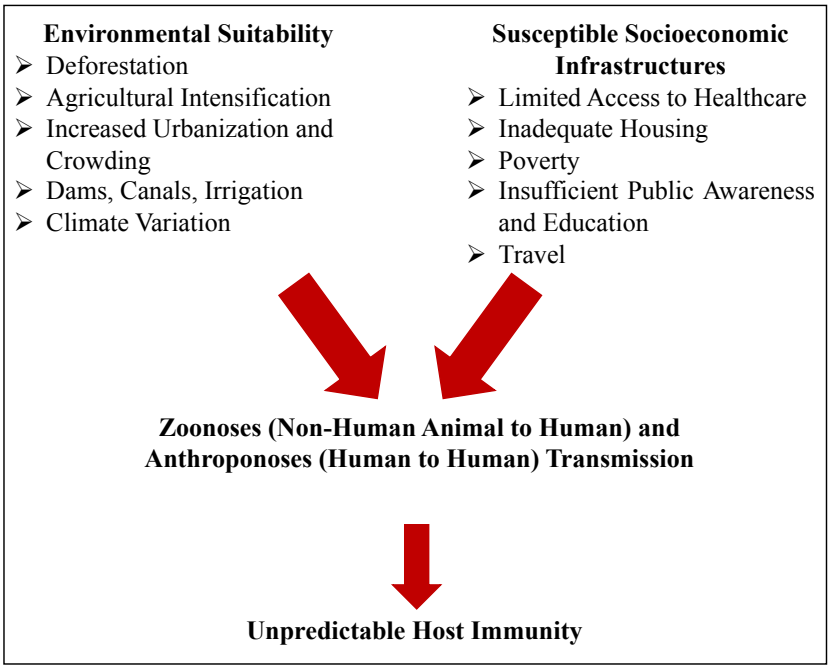


Figure 1.9 Environmental and socioeconomic factors that influence viral transmission.
Key environmental and socioeconomic factors important in the emergence and maintenance of viral pathogens are indicated. Environmental vulnerability may increase risk of zoonoses while susceptible socioeconomic factors can increase risk of anthroponoses. (Image amended in part from Ali et al., 2017 [103] with permission under the terms of the Creative Commons Attribution License (CC BY))

1.2.2 Limitations of Current Disease Surveillance Methods

Suitable interventions are required to mitigate the risk of viral emergence and their expansions. The control of a highly diverse set of pathogens such as viruses, requires timely, easy to navigate, and cost-effective surveillance systems that are readily available to scientists and healthcare providers [42]. Currently, routine surveillance of viral infections depends on healthcare providers and their detection of viral symptoms in patients, confirmed by previously mentioned culture-based methods, targeted molecular testing, immunoassay-based testing, and viral sequencing of suspected infections [38]. Although these are common routines in the monitoring of viral pathogens, they are associated with limitations.

By observing the genetic code of a virus, misdiagnosis of infections will occur to a lesser extent. Metagenomic sequencing may soon not require preliminary healthcare diagnosis [37]. During outbreaks, clinical samples can also be sequenced in real-time by portable sequencing machines, rather than samples

being sent to reference laboratories for processing, decreasing the turnover time from sample collection to data acquisition [42]. The extent to which genetic sequencing can aid in understanding rapidly evolving viruses is still being established.

1.3 Research Objectives and Aims, and Thesis Organization

1.3.1 Research Objectives and Aims

The primary objective of this thesis is to investigate the utility of current sequencing technologies in the detection and monitoring of circulating viral pathogens, focusing on three specific examples.

There are three aims:

1) Identify the longitudinal yearly prevalence of INSTI resistance in the HIV-1 population in British Columbia, Canada from 2009 to 2016,
2) Examine the transmission of minority species drug resistance and determine their impact on virological outcome in an HIV-1 population in Mbarara, Uganda,
3) Evaluate the use of whole-genome sequencing of ZIKV to correctly identify origins of travel-related ZIKV infections and provide insight into the genetic diversity of ZIKV.

1.3.2 Chapter Descriptions

This thesis is divided into five chapters. Tables and figures are placed after their first mention in this thesis, and a single bibliography is included after the final chapter. Additional supplemental data is provided in the appendices. Chapter One provides a general overview of current sequencing methodologies, and their current utility and limitations in the monitoring of viral pathogens. Chapters Two, Three, and Four each corresponds to one of the aims of this study and together, they address the primary objective of this study. The three aims investigate the transmission, detection, and surveillance of currently circulating viral pathogens, HIV and ZIKV.

Specifically, Chapter Two investigates the yearly prevalence of INSTI resistance in the HIV-1 population in British Columbia, Canada from 2009 to 2016. The use of INSTIs raltegravir, elvitegravir, and dolutegravir has increased dramatically over recent years. However, there is limited information about the

evolution and prevalence of INSTI resistance mutations in clinical HIV populations. HIV-positive Individuals within the Drug Treatment Program ≥19 years were included in this study if they received ≥1 dispensed prescription of ART in British Columbia between 2009 and 2016.

Although INSTI dolutegravir is now a recommended component of first-line regimens around the world, this option excludes women of reproductive age due to the potential risk of neural tube defects during pregnancies. Chapter Three examines the prevalence of pretreatment minority species resistance-associated mutations to a widely-used class of ARVs, nonnucleoside reverse transcriptase inhibitors (NNRTIs), and assesses their implication on virological response in Mbarara, Uganda, a region where NNRTI use is still common. Participants were treatment-naive individuals ≥18 years of age, initiating NNRTI-based regimens as part of the Uganda AIDS Rural Treatment Outcomes cohort.

Chapter Four examines the surveillance capability of whole-genome sequencing of travel-related cases of ZIKV infection in determining the suspected origins of ZIKV infection through phylogenetic analysis. At the same time, the genetic diversity of travel-related cases and currently available online whole-genome ZIKV sequences were analyzed.

Chapter Five summarizes the research findings of each chapter and presents the summary, limitations, and future applications of the research discoveries.

2 Chapter Two: Prevalence of Human Immunodeficiency Virus-1 Integrase Strand Transfer Inhibitor Resistance in British Columbia, Canada Between 2009 and 2016: A Longitudinal Analysis

2.1 Introduction

Integrase strand transfer inhibitors (INSTIs) are effective and well-tolerated antiretrovirals (ARVs) [104,105]. With increasing prevalence of pretreatment resistance to nonnucleoside reverse-transcriptase inhibitors (NNRTIs) [78], INSTIs are increasingly recommended in first-line and alternative first-line regimens [50,106–108]. Integrase strand transfer inhibitor-containing therapy also provides an alternative treatment option for individuals experiencing multidrug resistance or adverse reactions to other human immunodeficiency virus (HIV) ARVs [50,104]. However, INSTI resistance remains a barrier to the ongoing success of HIV treatment [109]. Although ARVs effectively suppress HIV replication enabling immune restitution, selection of resistance-associated mutations in the presence of ARVs is associated with lack of viral suppression, treatment failure, and increased likelihood of HIV transmission [109–111]. Furthermore, long-term persistence of drug resistance mutations has previously been reported and can limit regimen options [112]. Although pretreatment protease inhibitor (PI) and reverse transcriptase (RT) inhibitor resistance testing is the current standard of care in British Columbia (BC) [113,114], standard clinical guidelines currently do not recommend testing for INSTI resistance at initiation of therapy [55]. Instead, INSTI resistance testing is only recommended in patients who experience virologic failure while on INSTI-containing therapy [50,114].

There have been previous reports of transmitted and treatment-emergent INSTI resistance in clinical settings [115–118]. However, to better monitor and evaluate the current modalities of INSTI resistance and assess adequacy of INSTI resistance testing, more information is required about the change

in yearly prevalence of INSTI resistance and the specific integrase mutations associated with INSTI resistance selected, in large clinical HIV populations.

In a previous study, we observed low rates of treatment-emergent INSTI resistance (approximately 1 case per 100 person-years INSTI exposure) among individuals registered in the BC Centre for Excellence in HIV/AIDS (BC-CfE) Drug Treatment Program (DTP) who initiated INSTI-containing therapies between 2012 and 2014 [57]. This longitudinal, observational study examines the prevalence of INSTI resistance and the specific INSTI resistance mutations selected in antiretroviral therapy (ART)-treated DTP individuals in each calendar year between 2009 and 2016. For context, the yearly prevalence of PI and RT (PI-RT) resistance as well as the frequency of INSTI and PI-RT resistance testing were also investigated in the same time period.

2.2 Methods

2.2.1 Study Population

Participants within this study were HIV-1-positive individuals living in BC, Canada who received ART through the BC-CfE DTP. The BC-CfE monitors and maintains clinical patient profiles (ART, plasma HIV-1 ribonucleic acid [RNA] viral load, CD4 cell count, etc.) and collects sociodemographic data of DTP participants. A patient information sheet describing the potential uses of data for health research is provided at DTP enrollment, and consent is not required for use of anonymized data. The DTP is an open cohort and members can enter and leave at any time, for any duration of time, or leave indefinitely. These programs have been described in detail elsewhere [119,120]. The University of British Columbia Providence Health Care Research Ethics Board granted ethical approval for the DTP (H05-50123).

For each calendar year from January 1, 2009 to December 31, 2016, participants were counted in the denominator of ART-treated individuals if they were ≥19 years of age and were treated with at least 1

day of ART within that year. Treatment with ART was determined as ≥1 dispensed ART prescription in a calendar year through the BC-CfE's administrative records ART prescription database.

2.2.2 Drug Resistance Testing

Drug resistance tests were ordered by each individual's physician and processed at the BC-CfE research laboratories. However, if an initial HIV-1 RNA viral load level was requested by an individual's physician and the HIV-1 RNA viral load was >250 copies/mL, a drug resistance test was automatically ordered by the BC-CfE and the results were sent to the physician. Each individual in the study contributed cumulative drug resistance data from time of DTP enrollment until December 31, 2016. An individual contributed to the count of "tested for resistance" in the first year that the test was performed, and this "tested" status was automatically carried forward to each subsequent year the individual was treated with ART in the DTP. Viral nucleic acids were extracted from 500 μL plasma using the NucliSENS easyMAG from bioMérieux (Montreal, Canada). The protease, RT, and integrase genes were amplified by "nested" reverse transcription-polymerase chain reaction (PCR) using the Expand High Fidelity PCR system from Roche Diagnostics (Laval, Canada) as described elsewhere [121,122]. Sanger sequencing was performed bidirectionally using the BigDye Terminator version 3.1 Cycle Sequencing Kit from Applied Biosystems (ABI, Foster City, CA) on an ABI 3730xl DNA Sequencer. A consensus sequence was produced, and chromatograms were analyzed by in-house custom software called RECall [59].

2.2.3 Prevalence of Integrase Strand Transfer Inhibitor and Protease Inhibitor and Reverse-Transcriptase Inhibitor Resistance

In each year, an individual contributed to the total count of ART-treated individuals if they were treated with ART through the DTP within that year. Because annual prevalence is calculated separately for each year, deceased and/or lost to follow-up participants are accounted for accordingly. Samples were defined as drug resistant to either INSTI or PI-RT if they contained cumulative mutations that result in intermediate- or high-level resistance to at least 1 ARV in the INSTI or PI-RT drug classes, respectively,

as defined by a score ≥30 based on the Stanford University HIV Drug Resistance Database Genotypic Resistance Interpretation Algorithm version 7.0.1 [63]. To account for the potential long-term persistence of drug resistance mutations [112], each ART-treated individual tested for drug resistance was classified as having drug resistance (INSTI or PI-RT resistance) in the first year they had a sample meeting the criteria for resistance, and this resistance status was carried forward to each subsequent year the individual was treated with ART in the DTP. The annual prevalence of INSTI and PI-RT resistance per 1000 ART-treated individuals was calculated at the end of each calendar year.

2.2.4 Integrase Strand Transfer Inhibitor Resistance by Year of First Detection

To estimate the contribution of treatment-emergent INSTI resistance to the prevalence of INSTI resistance over time, new cases of INSTI resistance among ART-treated individuals were identified in each year and counted in the first year they were detected. It is important to note that newly identified cases of INSTI resistance do not correspond to incidence of INSTI resistance during the study period, but rather to the first year an individual who contributed to the count of prevalence had study-defined INSTI resistance during the study period. The drug associated with newly detected INSTI resistance was categorized as the last prescribed INSTI (either raltegravir, elvitegravir, or dolutegravir) before detection of resistance. If the first identified INSTI resistance occurred before the first known INSTI dispensed date in the DTP, the INSTI drug exposure was termed unclassifiable.

2.2.5 Prevalence of Integrase Strand Transfer Inhibitor Resistance Mutations

Each ART-treated individual with newly detected study-defined INSTI resistance was counted once per mutation position in each calendar year they received ART in the DTP. Samples were defined as drug resistant to INSTIs if they contained cumulative mutations that result in intermediate- or high-level resistance to at least 1 ARV in the INSTI drug class, as defined by a score ≥30 based on the Stanford University HIV Drug Resistance Database Genotypic Resistance Interpretation Algorithm version 7.0.1

[63]. An individual's INSTI mutation contributed to the mutation count in the first year it was detected and automatically carried forward to all subsequent years the individual received ART in the DTP.

2.2.6 Statistical Analysis

Generalized additive models were used to model the nonlinear trend of the prevalence of PI-RT and INSTI drug resistance [123,124]. Statistical software R version 3.2.2 was used with the mgcv package [123], assuming a Beta distribution and a logit link. The year or percentage of patients tested for INSTI resistance were the explanatory variables being smoothed. Restricted maximum likelihood was used to estimate the parameters. A cubic regression spline was used to smooth the yearly trend in each of the outcomes.

2.3 Results

2.3.1 Characteristics and Yearly Number of Participants

Between January 1, 2009 and December 31, 2016, 9358 unique individuals received ART through the DTP in BC. Patients were predominantly male (83%) infected with HIV-1 subtype B (81%). The median age of participants at the year of first inclusion in the study was 46 years (25th–75th percentile [Q1–Q3] 38–53) (Table 2.1). In total, 3645 unique individuals (39%) received INSTI-containing therapy in the DTP during the study period. A patient may have been prescribed more than 1 INSTI during the study period. During this time period, there were 1546 individuals ever treated with raltegravir, 830 individuals ever treated with elvitegravir, and 1856 individuals ever treated with dolutegravir. Between 2009 and 2016, the number of individuals enrolled and receiving treatment through the DTP each year increased from 5587 to 7772 individuals (39% increase), and the proportion receiving an INSTI increased from 10% to 40% (542 to 3117 individuals) (Appendix I).

Table 2.1 Characteristics of individuals within the BC Drug Treatment Program between 2009 and 2016.

Variable	N=9358 n (%)
ART-treated individuals[a]	9358 (100)
INSTI-treated individuals[b]	3645 (39)
INSTI prescribed in the DTP[c]	
Raltegravir	1546 (17)
Elvitegravir	830 (9)
Dolutegravir	1856 (20)
Age, years, median (Q1-Q3)[d]	46 (38-53)
Sex	
Male	7768 (83)
Female	1590 (17)
HIV subtype	
B subtype	7542 (81)
non-B subtype	586 (6)
Unknown	1230 (13)
Number of years individuals contributed resistance data, median (Q1-Q3)	6 (2-8)
Number of years individuals contributed any type of data, median (Q1-Q3)	7 (4-8)
Number of ART-treated individuals ever tested for PI-RT resistance	7883 (84)
Number of physician-ordered resistance tests done per person, median (Q1-Q3)	2 (1-5)
Number of individuals with a single resistance test only, at some point in time	2647 (28)
Number of individuals with a single resistance test performed at baseline	2109 (23)
Number of INSTI-treated individuals ever tested for INSTI resistance	1244 (13)
Number of physician-ordered resistance tests done per person, median (Q1-Q3)	1 (1-2)
Number of individuals with a single resistance test only, at some point in time	752 (8)
Number of individuals with a single resistance test performed at baseline	482 (5)

[a]An individual contributed to the count of ART-treated if they were ever dispensed ART in the DTP during the study period; [b]An individual contributed to the count of INSTI-treated if they were ever dispensed INSTIs in the DTP during the study period; [c]An individual could be prescribed more than one INSTI during the study period; [d]Median age of individuals at first year of inclusion in the study.

2.3.2 Drug Resistance Testing

During the study period, the number of ART-treated individuals tested for PI-RT resistance increased from 4520 to 6614 individuals (81% to 85%) (Figure 2.1), and the number of all ART-treated individuals tested for INSTI resistance increased from 188 to 1440 individuals (3% to 19%) ($p < 0.0001$, $R^2 = 0.99$) (Appendix I). Among those treated with INSTIs each year, the number of individuals tested for INSTI resistance increased from 60 to 1059 individuals (11% to 34%) ($p < 0.0001$, $R^2 = 0.97$) (Figure 2.1).

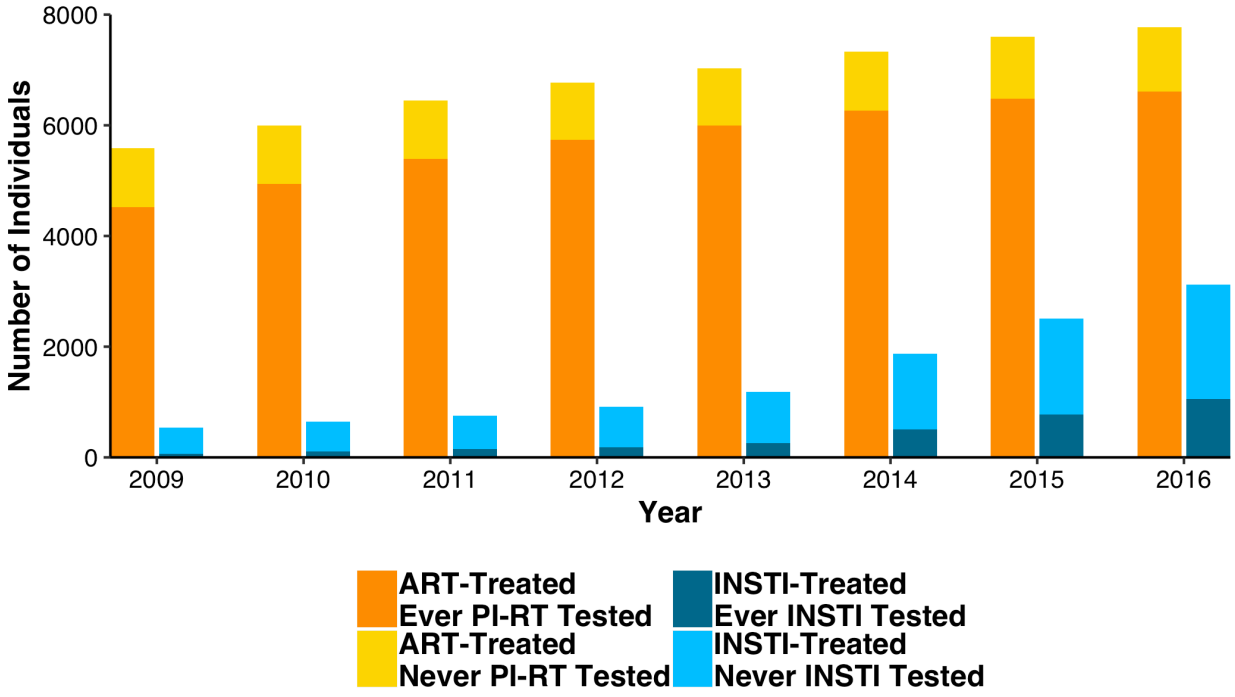


Figure 2.1 Antiretroviral therapy and drug resistance testing.
All active ART-treated individuals tested and not tested for PI-RT resistance and active INSTI-treated individuals tested and not tested for INSTI resistance as of December 31st of each year from 2009 to 2016 are indicated. Individuals contributed to the count of PI-RT or INSTI tested in the first year that a PI-RT or INSTI resistance test was performed, and this tested status was automatically carried forward to each subsequent year the individual was treated with ART in the DTP.

2.3.3 Prevalence of Resistance in All Antiretroviral Therapy-Treated Individuals

Figure 2.2 shows the prevalence of PI-RT and INSTI resistance in all ART-treated individuals treated with ART in BC in each calendar year from 2009 to 2016. The prevalence of study-defined PI-RT resistance in all ART-treated individuals declined significantly from 337 per 1000 ART-treated individuals in 2009 to 285 per 1000 ART-treated individuals in 2016 ($p < 0.0001$, $R^2 = 0.98$) (Figure 2.2). In contrast, the prevalence of study-defined INSTI resistance was lower than that of PI-RT resistance, but it increased from 1 per 1000 ART-treated individuals in 2009 to 7 per 1000 ART-treated individuals in 2016 ($p < 0.0001$, $R^2 = 0.98$) (Figure 2.2).

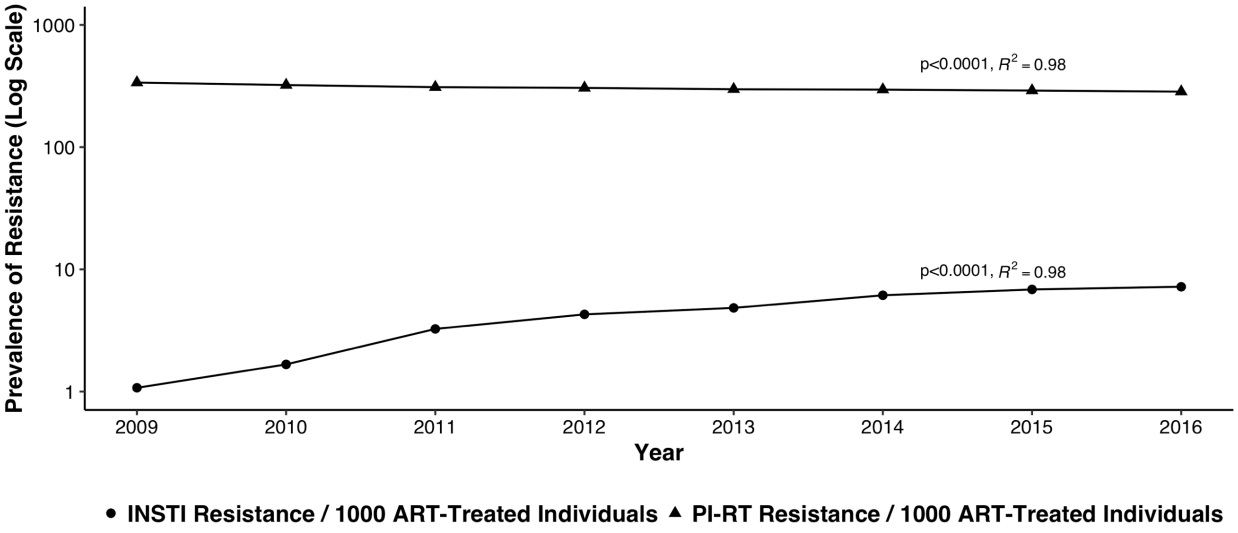


Figure 2.2 Prevalence of PI-RT and INSTI drug resistance.
The annual prevalence of PI-RT and INSTI drug resistance per 1000 ART-treated individuals between 2009 and 2016 within the DTP is shown. The trend shows a decrease in the prevalence of PI-RT resistance from 337 per 1000 ART-treated individuals in 2009 to 285 per 1000 ART-treated individuals in 2016 ($p < 0.0001$, $R^2 = 0.98$); the trend shows an increase in the prevalence of INSTI resistance from 1 per 1000 ART-treated individuals in 2009 to 7 per 1000 ART-treated individuals in 2016 ($p < 0.0001$, $R^2 = 0.98$).

2.3.4 Integrase Strand Transfer Inhibitor Resistance by Year of First Detection

From 2009 to 2016, 64 unique individuals (69% male) were newly identified as having study-defined INSTI resistance (Table 2.2). Median age was 47 years ([Q1–Q3] 40–53), and the majority were infected with HIV-1 subtype B (91%). Among these individuals, 88% (56 of 64) have also had mutations conferring PI-RT resistance. The observed number of newly identified cases of INSTI resistance ranged from 4 to 15 new cases per year during the study period, and it remained relatively stable at 6–9 cases per year after peaking in 2011 (Table 2.2). Patient characteristics of a subset of these 64 individuals have been previously described in detail elsewhere [57]. During the study period, there was an apparent shift from raltegravir-containing regimens to elvitegravir- and dolutegravir-containing regimens (Appendix II) and until 2014, most new cases of INSTI resistance were associated with the use of raltegravir (Table 2.2). In 2015 and 2016, the first 2 full years the 3 INSTIs were widely prescribed, 80% of new INSTI resistance cases (12 of 15) were attributed to elvitegravir and dolutegravir use compared with 2013 and 2014, where 88% of INSTI resistance cases (15 of 17) were associated with raltegravir use (Table 2.2). With decreasing use of raltegravir, most new cases of INSTI resistance have been associated with the use of elvitegravir. In 4 individuals, INSTI drug resistance mutations were documented in a sample drawn before treatment with an INSTI-containing regimen in the DTP. It is unknown whether these unclassifiable individuals had received prior INSTI treatment elsewhere (Table 2.2).

Table 2.2 Newly identified cases of INSTI resistance within the BC Drug Treatment Program between 2009 and 2016.

INSTI Drug Exposure[a]	Year								Total Resistance Cases per INSTI
	2009	2010	2011	2012	2013	2014	2015	2016	
Raltegravir	6	3	12	7	8	7	2	1	46
Elvitegravir	0	0	0	0	0	2	4	5	11
Dolutegravir	0	0	0	0	0	0	3	0	3
Unclassifiable[b]	0	1	3	0	0	0	0	0	4
Total per Year	6	4	15	7	8	9	9	6	**64**

[a]The INSTI a person was last exposed to prior to the first detection of INSTI resistance; [b]If a person had mutations conferring INSTI resistance before the first known INSTI dispensing date in the DTP, the INSTI drug exposure was termed unclassifiable.
Newly identified ART-treated individuals with mutations conferring INSTI resistance with a total score $\geq$30 by the Stanford University HIV Drug Resistance Database Genotypic Resistance Interpretation Algorithm v7.0.1 to at least one INSTI is displayed. Note that in Dec-2009, as shown in Appendix II, 475 individuals were prescribed raltegravir, 2 individuals were prescribed elvitegravir, and 0 individuals were prescribed dolutegravir, compared to Dec-2016, where 751 individuals were prescribed raltegravir, 579 individuals were prescribed elvitegravir, and 1467 individuals were prescribed dolutegravir.

2.3.5 Prevalence of Integrase Strand Transfer Inhibitor Resistance Mutations

The prevalence of specific INSTI resistance mutations for individuals with newly identified study-defined INSTI resistance from 2009 to 2016 is depicted in Figure 2.3. Nucleotide substitutions at codons 51, 66, 74, 92, 95, 97, 118, 121, 138, 140, 143, 145, 146, 147, 148, 151, 153, 155, 157, 163, 230, and 263 of the integrase gene were identified among individuals with study-defined INSTI resistance. Overall, there was an increase in the prevalence of INSTI resistance mutations counted at codons 66, 97, 140, 148, 155, and 263 of the integrase gene (Figure 2.3). By the end of the study, mutation 155H/S/T was detected in 18 individuals and had the highest prevalence of all detected mutations. The second most prevalent mutation, 263K, was identified in 14 individuals by the end of 2016, after increasing from 1 to 11 individuals between 2010 and 2011.

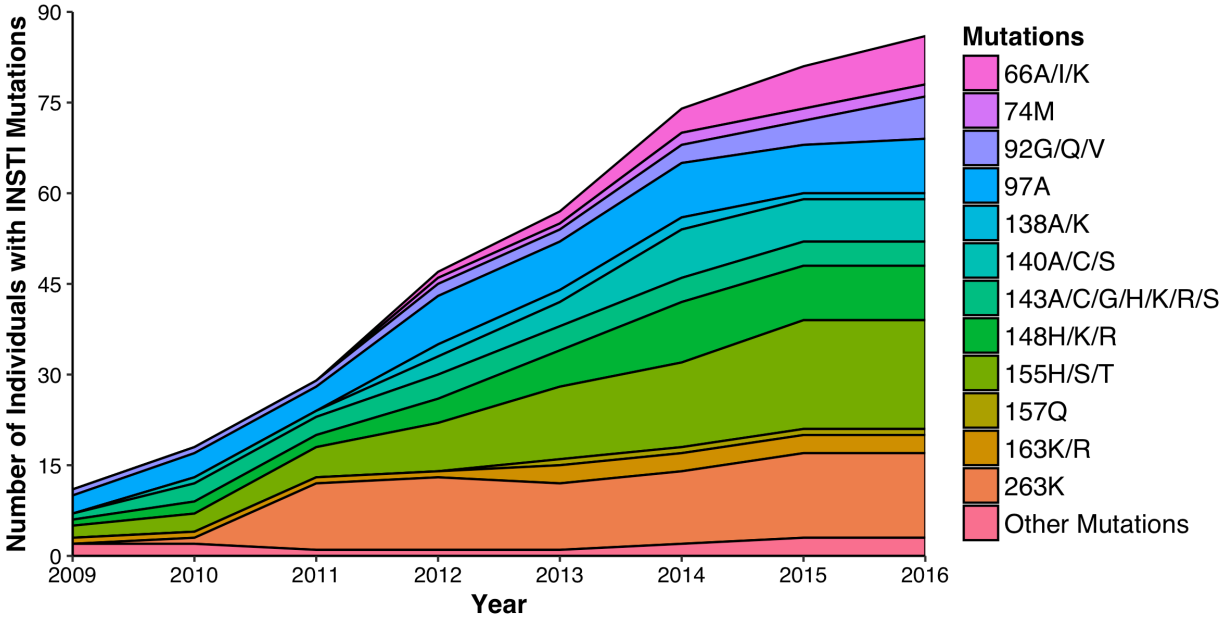


Figure 2.3 Prevalence of mutations conferring INSTI resistance within newly identified INSTI-resistant individuals.
The prevalence of INSTI resistance mutations in INSTI-resistant individuals within the DTP between 2009 and 2016 is shown. An individual's INSTI mutation contributed to the mutation count in the first year it was detected, and automatically carried forward to all subsequent years the individual received ART in the DTP. A person with mutations in three separate codons (e.g. 92, 95, and 97) in a year will contribute three counts for that year while a person with mutations in multiple positions within a codon only contributes count of one for that codon (e.g. mutations in positons 155H and 155S will only be counted once under 155H/S/T). The "Other Mutations" category represents the sum of mutation counts from 118R, 121Y, 145S, 146P, 147G, 151A/L, 153F/Y, 230R, 51Y, and 95K. Note: certain specific positions associated with INSTI resistance increased in prevalence over the time period.

2.4 Discussion

Antiretroviral-resistant strains of HIV have emerged for all ARVs used in clinical settings, including those of newer classes such as INSTIs [78]. In this study, we conducted a retrospective study of over 9000 HIV-1-positive, ART-treated individuals who participated in the BC DTP from 2009 to 2016. During this study period, there was a dramatic increase in the number of individuals receiving INSTIs raltegravir, elvitegravir, and dolutegravir, and a small but statistically significant increase in the prevalence of intermediate- to high-level INSTI resistance in the ART-treated population.

Similar to findings in the United States [125], we observed an increase in the proportion of ART-treated individuals tested for INSTI resistance. This increase may be due to fewer restrictions on INSTI resistance testing in BC in recent years as well as the partially automated INSTI resistance testing implemented in 2014 for individuals treated with ART in the DTP. This automated INSTI resistance testing system involves the addition of INSTI resistance tests to the first drug resistance test ever done in BC, and to any standard, physician-ordered PI-RT resistance test if the individual tested was currently or previously treated with INSTIs in BC. Nevertheless, although the number and proportion of INSTI-treated individuals tested for INSTI resistance increased each year during our study period, the proportion of INSTI resistance testing among INSTI-treated individuals has lagged behind the uptake of INSTI-containing therapy (11% in 2009; 34% in 2016) and lagged behind PI-RT resistance testing. This may be a result of the current physician-accessible ART guidelines, which do not support INSTI resistance testing for individuals unless there is evidence of virological failure and suspected INSTI resistance [50,114].

Over the study period, we observed a decrease in the prevalence of PI-RT resistance in our ART-treated population. A decreasing trend in PI-RT resistance has recently been reported in other clinical HIV-positive populations [126] and may be attributed to the introduction of newer ARVs with greater genetic barriers to resistance. For example, second-generation NNRTIs rilpivirine and etravirine are not

significantly impacted by single NNRTI resistance mutations [127] and have shown similar viral suppression efficacy to efavirenz [128]. In addition, increased production of novel fixed-dose combination products, which reduce pill burden and simplify regimens, may lead to better adherence and lower likelihood of ARV resistance [129,130].

The prevalence of INSTI resistance reported in our study is similar to a recent study observing the prevalence of transmitted INSTI resistance in a Swiss HIV cohort [131] but is lower than those reported in other North American and European cohorts [115,116,125,126]. This difference may in part be due to our method of defining drug resistance, which was restricted to individuals who had intermediate- to high-level INSTI resistance as defined by the Stanford University HIV Drug Resistance Database Genotypic Resistance Interpretation Algorithm version 7.0.1. Other studies have defined INSTI resistance as individual [115,116,125,131] and/or cumulative [116,125] presence of mutations. Nevertheless, there is an increase in the prevalence of intermediate- to high-level INSTI resistance within our population from 2009 to 2016 (1 to 7 per 1000 ART-treated individuals), likely due to the increase in the number of individuals receiving INSTI-containing therapies. In previous studies, we have identified <80% ARV adherence and CD4 cell count <200 cells/μL as strong risk factors of emergent INSTI resistance in DTP participants treated with INSTI-containing regimens [57], and these factors may contribute to the increase in the prevalence of INSTI resistance over time.

Since 2009, usage of elvitegravir and dolutegravir have risen considerably in the DTP since their introduction in BC in 2013 and 2014, respectively, whereas there has been a decline in the use of raltegravir, the first INSTI [34]. The observed increase in usage of dolutegravir may be due to its association with fewer adverse drug effects and its higher selective barrier for multi-drug resistance mutations compared with elvitegravir and raltegravir [132–134]. This shift in prescribed drug therapies coincides with the shift in INSTIs associated with newly detected cases of INSTI resistance. Over time in our ART-treated population, the observed number of new raltegravir-associated resistance cases decreased as use of raltegravir declined

(6 cases in 2009; 1 case in 2016). Elvitegravir-associated resistance accounted for the majority of new cases in 2015 and 2016, whereas only 3 new cases were associated with dolutegravir use, despite a marked increase in dolutegravir prescribing. Overall, the number of newly identified cases of individuals with study-defined INSTI resistance increased and then subsequently decreased during the study period.

Individuals with newly identified study-defined INSTI resistance appear to have mutations selected at codons 66, 97, 140, 148, 155, and 263. Mutation 66I is expected with increased elvitegravir usage [135]. Mutation 97A, in combination with other INSTI mutations, confers reduced susceptibility to raltegravir and dolutegravir [136,137]. Mutations 140A/S are associated with dolutegravir and raltegravir resistance, and in combination with mutation 148H, they are associated with reduced virological suppression [138]. Mutations 148H/R and 155H are mutations associated with 2 distinct mutational pathways that reduce susceptibility to all current INSTIs [139,140]. Mutation 263K is associated with decreased HIV-deoxyribonucleic acid integration, viral replication capability, and integrase enzyme capacity but may also confer low- to intermediate-level resistance to dolutegravir and elvitegravir [141,142]. The 263K mutation has previously been noted as an emergent INSTI mutation in treatment-experienced, INSTI-naive DTP individuals initiating raltegravir, elvitegravir, or dolutegravir for the first time [57] as well as in other treatment-experienced HIV-1-positive populations [132].

There are several limitations within our study. First, the incomplete coverage of INSTI resistance testing limits our ability to adequately characterize INSTI resistance mutations and may also underestimate the prevalence of INSTI resistance. We speculate that this poor coverage could be due to the relatively new nature of INSTIs and their significance might be overlooked. This study also observed cumulative mutations that confer intermediate- to high-level resistance to INSTIs, and this threshold does not include individual mutations that may confer low-level resistance and may be polymorphic or treatment-emergent. Most INSTI resistance tests are requested in response to virological failure rather than before initiation of ART therapy, and therefore, individuals with intermediate- to high-level resistance to INSTIs were included

in our study because they were more likely to be identified during routine clinical care. Without comprehensive longitudinal baseline INSTI resistance testing, the prevalence of individual low-level resistance mutations cannot be accurately estimated. As a result, we cannot determine the true prevalence of INSTI mutations present in our ART-treated population and our INSTI prevalence may be underestimated. Our prevalence of INSTI resistance may also be underestimated due to Sanger sequencing's inability to detect rare but clinically significant drug resistance mutations present at low viral populations [143]. In addition, our assessment of nucleotide substitutions within the integrase gene may limit the identification of key mutations outside of the integrase gene, which may confer high-level resistance to raltegravir, elvitegravir, and dolutegravir [144].

2.5 Conclusions

In summary, our results indicate that the prevalence of INSTI resistance in the DTP is low, but it is gradually increasing over time as INSTI prescribing increases. The ability to precisely characterize and determine the frequency of INSTI resistance is hampered by the limited usage of INSTI resistance testing. Although previous research suggests dolutegravir's genetic barrier to resistance is high, the genetic barrier of other currently prescribed INSTIs, raltegravir and elvitegravir, is comparable to other ARVs [134]. Integrase strand transfer inhibitor-treated individuals with suboptimal adherence may require more extensive monitoring to permit early detection of emergent INSTI resistance mutations. Further drug resistance monitoring is required to detect any changes in the prevalence of INSTI resistance with the introduction of newer INSTIs, such as bictegravir.

3 Chapter Three: Sensitive Detection of Pretreatment Minority Species HIV-1 NNRTI Resistance in Uganda: Limited Benefits and Analytical Pitfalls

3.1 Introduction

The prevalence of pretreatment resistance to widely used nonnucleoside reverse transcriptase inhibitors (NNRTIs) efavirenz and nevirapine for the treatment of human immunodeficiency virus type 1 (HIV-1), have risen considerably in recent years [78]. NNRTI-containing regimens have lower genetic barriers to resistance than alternative therapies and the emergence of NNRTI resistance may limit future therapeutic options [145]. In low- and middle-income countries such as Uganda, first-line therapy mainly constitutes NNRTI-containing regimens, and pretreatment low-, intermediate- or high-level resistance to NNRTIs have been detected in ≥10% of individuals, including children younger than 12 years of age, initiating antiretroviral therapy (ART) [78,146,147]. In response to increasing prevalence of NNRTI resistance in Uganda, the World Health Organization recommended a transition to dolutegravir-containing regimens, however this option currently excludes women of reproductive age due to potential emergence of neural tube defects during pregnancies [148–150].

Detection of potential drug resistance is imperative in guiding clinicians in the choice of initial and subsequent NNRTI-containing regimens in regions with high levels of pretreatment resistance [150]. Standard genotypic resistance tests rely on Sanger "bulk" population sequencing, which detect drug resistance-associated mutations (DRMs) present at ≥15-20% of the viral quasispecies population [151,152]. In contrast, next-generation sequencing methods have lower limits of detection and are capable of detecting potentially rare clinically-relevant minority species DRMs present at low viral thresholds (<15%) [143,153–155]. Furthermore, next-generation sequencing methods are associated with reduced cost per nucleotide-sequenced [156,157], becoming a supplemental sequencing methodology currently being

investigated for the purpose of genotypic resistance testing in low- and middle-income countries [150]. Common automated resistance interpretation algorithms which predict antiretroviral susceptibility include the Stanford University HIV Drug Resistance Database (HIVdb) Genotypic Resistance Interpretation Algorithm and the geno2pheno program [62,63]. Although these algorithms were designed to process Sanger-level sequencing data, nearly all next-generation sequencing studies to date looking at the relevance of minority species resistance levels, use consensus-based sequences and these algorithms to evaluate drug resistance data [66,158,159].

The clinical implications of minority species DRMs present at low viral populations is currently still debated. There is lack of consistent evidence to establish a definitive link between the presence of minority species DRMs and lower likelihood of virological suppression. While certain studies have shown that the presence of minority species DRMs prior to initiation of ART have a significant impact on treatment outcome [143,153,154], others have failed to find supportive evidence [160–162]. In this study, we examine the genotypic resistance profiles of samples from the Uganda AIDS Rural Outcomes (UARTO) cohort in order to determine the impact of pretreatment minority species NNRTI resistance on virological suppression among individuals initiating NNRTI-containing regimens in Uganda.

3.2 Methods

3.2.1 Population Description

Participants within this study were ART-naive patients initiating ART for the first time in Mbarara, Uganda as part of the UARTO cohort, previously described elsewhere [163]. Briefly, ART-naive patients were recruited from the Mbarara University of Science and Technology Immune Suppression Syndrome (ISS) clinic. Eligible participants were included in the UARTO cohort if they were $\geq$18 years of age, lived within 50 kilometers of the ISS clinic, and were initiating ART for the first time. UARTO participants were included in this study if they had pretreatment genotypic drug resistance testing, initiated NNRTI-

containing regimens, and had follow-up HIV-1 viral RNA load data available at 1-year post-treatment. Institutional Review Boards at Partners Healthcare, University of California San Francisco, Mbarara University of Science and Technology, and Uganda National Council for Science and Technology granted ethical approval for the UARTO cohort. The University of British Columbia Providence Health Care Research Ethics Board granted ethical approval for this retrospective study (H11-01642).

3.2.2 Genotypic Drug Resistance Testing

Next-generation sequencing was performed using an Illumina MiSeq platform (San Diego, California, United States). Initially, viral nucleic acids were extracted from 500 µL of plasma using the NucliSENS easyMAG from bioMérieux (Montreal, Canada). The reverse transcriptase gene was amplified by reverse transcription-polymerase chain reaction (RT-PCR) using the Expand High Fidelity PCR system from Roche Diagnostics (Laval, Canada). MiSeq library preparations occurred using the RT-PCR products, and codons 90-234 of the reverse transcriptase gene were amplified in a second PCR reaction, to include the most relevant NNRTI resistance-associated mutations [138], with primers incorporating Illumina indexing adaptors. Each sample then underwent dual indexing in a third low-cycle PCR step, according to the Illumina Nextera XT indexing procedure. MiSeq data were processed by in-house pipeline using Bowtie2 and SAMtools [61,164]. Reads were mapped to the HXB2 laboratory strain and the mapped-reads were then used to generate sequence-specific consensus sequences to which the sample reads were remapped. These methods have been described extensively elsewhere [165]. A threshold of 1%, 2%, 5%, 10%, and 20% were chosen to report minority bases present and to include them in the generation of consensus sequences at the different viral thresholds.

3.2.3 Drug Resistance Mutations and Drug Resistance

Drug resistance-associated mutations (DRMs) were defined as resistance-conferring nucleotide substitutions identified at codons 90-234 of the reverse transcriptase gene according to HIVdb version 8.5 [63]. Samples harbouring DRMs were defined as resistant to NNRTIs if they contained one or more

mutations which result in cumulative low-, intermediate-, or high-level resistance to NNRTIs etravirine, rilpivirine, efavirenz, or nevirapine, as defined by a score ≥15 based on the HIVdb v8.5 [63]. Minority species DRMs were defined as resistance-conferring nucleotide substitutions detected in <20% of the virus population.

3.2.4 Prevalence of Pretreatment Drug Resistance

Individuals contributed to the prevalence count of pretreatment resistance if they had study-defined pretreatment NNRTI resistance detected within their pretreatment genotypic drug resistance tests and were subsequently treated with NNRTI-containing regimens during the study period. The prevalence of pretreatment NNRTI resistance was calculated for the duration of the study period.

3.2.5 Statistical Analysis

A linear-log regression model was used to model the trend of the prevalence of NNRTI resistance at different viral thresholds. The impact of resistance observed at the different study viral thresholds of 1%, 2%, 5%, 10%, and 20% on detectable viral RNA load at 1-year post-treatment initiation were assessed using univariable and multivariable binomial logistic regression models, stratified by the different viral threshold cut-offs. Age at enrolment, baseline CD4 count and viral RNA load were treated as continuous variables, and all other variables including sex and presence of study-defined pretreatment NNRTI resistance were treated as categorical variables. Detectable viral load was assessed using viral RNA load data at 1-year post-treatment initiation and was defined as viral RNA load ≥400 copies/mL while virological suppression was defined as viral RNA load <400 copies/mL. Unadjusted odds ratios (uORs), adjusted odds ratios (aORs), and 95% confidence intervals were determined. Statistically significant findings were defined as $p < 0.05$.

3.3 Results

3.3.1 Characteristics of Study Population

In total, 234 unique individuals from the UARTO cohort were included in this study with pretreatment sample collection years ranging between 2005 and 2013. Patients were predominantly female (68%) infected with HIV-1 subtype A (60%) followed by subtype D (25%). The median age of participants at year of first enrollment in UARTO was 33 years (25^{th} – 75^{th} percentile [Q1-Q3] 26-40). Individuals were treated with efavirenz- or nevirapine-containing regimens with the majority of participants receiving efavirenz (65%) (Table 3.1).

Table 3.1 Characteristics of individuals.

Variable	All (N=234)
NNRTI-treated individuals	234 (100)
Individuals with detectable HIV-1 viral RNA load 1-year post-treatment initiation	66 (28)
Initial antiretroviral treatment	
3TC/AZT/EFV	22 (9)
3TC/AZT/NVP	69 (30)
3TC/d4T/EFV	1 (0.4)
3TC/d4T/NVP	7 (3)
3TC/TDF/EFV	125 (53.4)
3TC/TDF/NVP	6 (3)
FTC/TDF/NVP	1 (0.4)
FTC/TDF/EFV	3 (1)
Baseline	
Age, years, median (Q1-Q3)	33 (26-40)
HIV-1 viral RNA load, median (Q1-Q3), log copies/mL	4.9 (4.4-5.5)
CD4+ cell count, median (Q1-Q3), cells/mm^3	241 (155-261)
Sex	
Male	74 (32)
Female	160 (68)

Variable	All (N=234)
HIV-1 subtype	
A	140 (60)
B	13 (6)
C	16 (7)
D	58 (25)
CRF01_AE	2 (1)
CRF10_CD	5 (2)
Earliest date of sample collection	20-Oct-2005
Latest date of sample collection	23-Aug-2013

3.3.2 Prevalence of NNRTI Resistance at Different Viral Thresholds

The prevalence of study-defined pretreatment NNRTI resistance detected at viral thresholds 1%, 2%, 5%, 10%, and 20% is shown in Table 3.2. Pretreatment NNRTI resistance was lowest (27 individuals, 12%) at 20% viral threshold and highest (51 individuals, 22%) at 1% viral threshold. Overall, the prevalence of NNRTI resistance decreased as cut-off of viral threshold increased ($p < 0.05$, $R^2 = 0.97$). The prevalence of intermediate- to high-level NNRTI resistance to future prescribed treatment of efavirenz or nevirapine, as defined by a score $\geq$30 on the HIVdb, was 5% at 20% viral threshold within the study population (Table 3.2).

Table 3.2 Prevalence of pretreatment NNRTI resistance observed at different viral thresholds.

Viral Threshold (%)	Prevalence of Study-Defined NNRTI Resistance[a] (%)	Prevalence of Intermediate- to High-Level NNRTI Resistance[b] (%) to Future Prescribed Treatment
1	22	12
2	18	8
5	16	6
10	14	5
20	12	5

[a]The prevalence of low-, intermediate-, or high-level NNRTI resistance (score $\geq$15) to etravirine, rilpivirine, efavirenz, or nevirapine.
[b]The prevalence of intermediate- to high-level NNRTI resistance (score $\geq$30) to future prescribed treatment of efavirenz or nevirapine.

3.3.3 Pretreatment NNRTI Resistance not a Significant Predictor of Virological Failure

Stratified by the different viral threshold, observed study-defined pretreatment NNRTI resistance was not a significant predictor of detectable viral RNA load 1-year post-treatment initiation, at a significance value of 5% (Table 3.3). Furthermore, the presence of study-defined minority species NNRTI resistance (<20% viral threshold) did not increase odds of detectable viral RNA load 1-year post-treatment initiation (Figure 3.1). Higher age was the only significant predictor variable identified as a protective variable in the univariable and multivariable logistic regression models (Table 3.3).

Table 3.3 Univariable and multivariable logistic regression analyses for detectable HIV-1 viral RNA load 1-year post-treatment initiation among ART-naive individuals with pretreatment NNRTI resistance observed at 1%, 2%, 5%, 10%, and 20% viral thresholds.

Variable	Univariable			Multivariable		
	p-value	OR	95% CI	p-value	OR	95% CI
1% Viral threshold						
Harbouring one or more DRM(s) which confer a total score of ≥15 on Stanford's HIVdb	0.75	1.14	(0.5-2.3)	0.81	1.1	(0.5-2.3)
Baseline CD4 count	0.12	0.78	(0.6-1.1)	0.15	0.80	(0.6-1.1)
Baseline HIV-1 viral RNA load	0.17	1.24	(0.9-1.8)	0.35	1.15	(0.9-1.6)
Age at enrollment	0.03*	0.97	(0.9-1.0)	0.03*	0.96	(0.9-1.0)
Sex (female)	0.72	0.90	(0.5-1.7)	0.47	0.79	(0.4-1.5)
2% Viral threshold						
Harbouring one or more DRM(s) which confer a total score of ≥15 on Stanford's HIVdb	0.95	1.00	(0.5-2.1)	0.84	0.93	(0.4-1.9)
Baseline CD4 count	0.12	0.78	(0.6-1.1)	0.15	0.80	(0.6-1.1)
Baseline HIV-1 viral RNA load	0.17	1.24	(0.9-1.8)	0.36	1.14	(0.9-1.6)
Age at enrollment	0.03*	0.97	(0.9-1.0)	0.02*	0.96	(0.9-1.0)
Sex (female)	0.72	0.90	(0.5-1.7)	0.50	0.80	(0.4-1.5)

Variable	Univariable			Multivariable		
	p-value	OR	95% CI	p-value	OR	95% CI
5% Viral threshold						
Harbouring one or more DRM(s) which confer a total score of ≥15 on Stanford's HIVdb	1.00	1.00	(0.4-2.1)	0.84	0.92	(0.4-2.0)
Baseline CD4 count	0.12	0.78	(0.6-1.1)	0.15	0.80	(0.6-1.1)
Baseline HIV-1 viral RNA load	0.17	1.24	(0.9-1.8)	0.36	1.14	(0.9-1.6)
Age at enrollment	0.03*	0.97	(0.9-1.0)	0.02*	0.96	(0.9-1.0)
Sex (female)	0.72	0.90	(0.5-1.7)	0.50	0.80	(0.4-1.5)
10% Viral threshold						
Harbouring one or more DRM(s) which confer a total score of ≥15 on Stanford's HIVdb	0.68	1.2	(0.5-2.6)	0.87	1.10	(0.4-2.4)
Baseline CD4 count	0.11	0.78	(0.6-1.1)	0.15	0.80	(0.6-1.1)
Baseline HIV-1 viral RNA load	0.17	1.24	(0.9-1.8)	0.34	1.14	(0.9-1.6)
Age at enrollment	0.03*	0.97	(0.9-1.0)	0.03*	0.96	(0.9-1.0)
Sex (female)	0.72	0.90	(0.5-1.7)	0.48	0.80	(0.4-1.5)
20% Viral threshold						
Harbouring one or more DRM(s) which confer a total score of ≥15 on Stanford's HIVdb	0.53	1.30	(0.5-3.0)	0.64	1.20	(0.5-2.9)
Baseline CD4 count	0.12	0.78	(0.6-1.1)	0.16	0.80	(0.6-1.1)
Baseline HIV-1 viral RNA load	0.17	1.24	(0.9-1.8)	0.34	1.15	(0.9-1.6)
Age at enrollment	0.03*	0.97	(0.9-1.0)	0.03*	0.96	(0.9-1.0)
Sex (female)	0.72	0.90	(0.5-1.7)	0.48	0.79	(0.4-1.5)

Results based on univariable and multivariable logistic regression models with p-value significant ≤ 0.05; Significant p-value indicated with asterisk (*); Dependent variable is detectable viral RNA load at 1-year post-treatment initiation; Units: Baseline CD4 count, cells/mm^3; Baseline viral RNA load, copies/μL; Age at enrollment, years.
Abbreviations: OR, odds ratio; CI, 95% confidence interval.

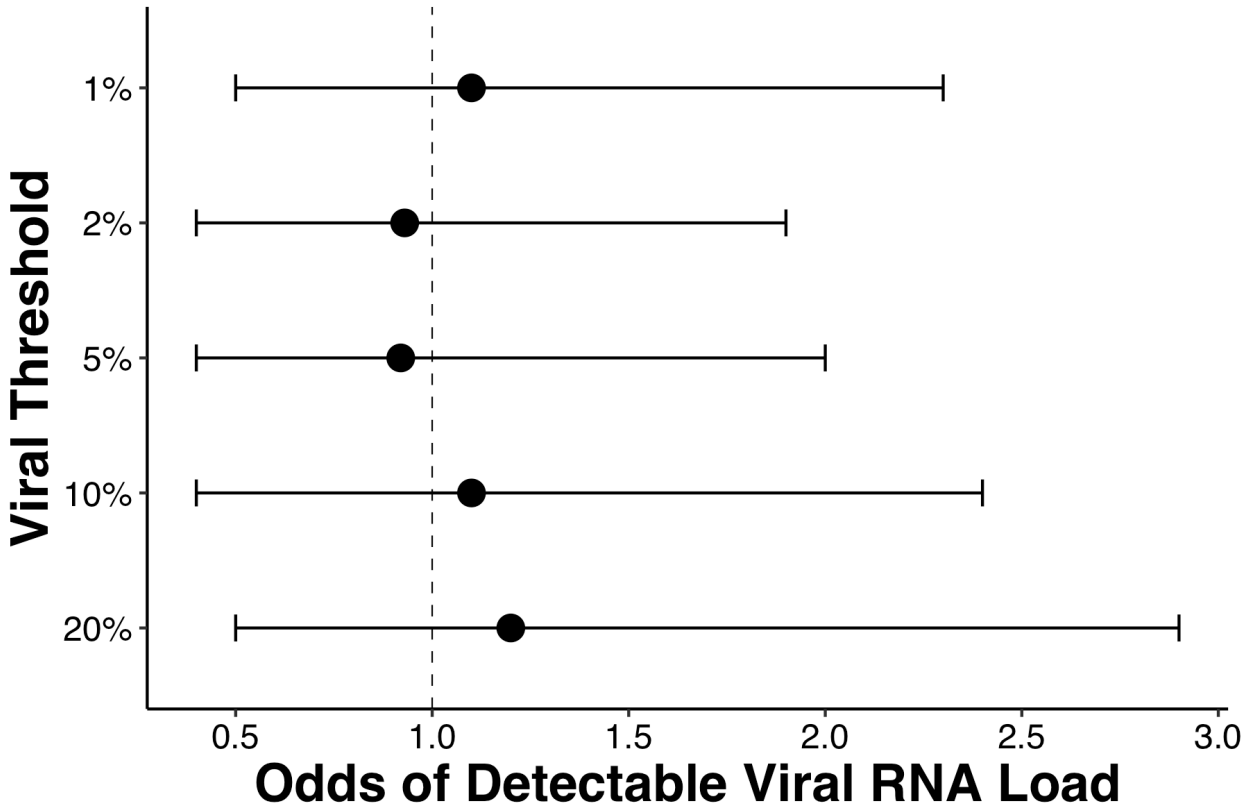


Figure 3.1 Presence of pretreatment NNRTI resistance did not significantly increase odds of detectable HIV-1 viral RNA load 1-year post-treatment initiation.
Results are based on multivariable logistic regression analyses of the impact of study-defined pretreatment NNRTI resistance, evaluated at different viral thresholds (1%, 2%, 5%, 10%, and 20%), on detectable viral RNA load 1-year post-treatment initiation. The presence of pretreatment NNRTI resistance was not a significant predictor of detectable viral RNA load 1-year post-treatment initiation at any of the viral thresholds examined ($p > 0.05$). Individuals harbouring DRMs at higher viral thresholds had higher odds of having detectable viral RNA load (>400 copies/mL) at 1-year post-therapy initiation (20% viral threshold: adjusted odds ratio [aOR] 1.20; 95% confidence interval [CI] 0.5-2.9; $p = 0.64$) compared to lower viral thresholds (10% viral threshold: aOR 1.10; CI 0.4-2.4; $p = 0.87$), although this was not statistically significant. Higher age was a statistically significant protective variable (20% viral threshold; aOR 0.96; CI 0.9-1.0; $p = 0.03$).

3.3.4 NNRTI Mutations

Nucleotide substitutions at codons 101, 103, 108, 138, 179, 181, 188, 190, 221, and 230 of the reverse transcriptase gene were identified among individuals with study-defined pretreatment NNRTI resistance (Figure 3.2). An individual could have contributed to more than one mutation count at a specific codon (e.g. an individual with mutations K103N/R/S contributed to the count of K103N, K103R, and

K103S, separately). As viral threshold decreased, there were increasing variation and mixture nucleotide combinations detected at codons in the reverse transcriptase gene (Appendix III). Overall, there was high prevalence of E138A (11 individuals at 20% viral threshold; 14 individuals at 1% viral threshold) and K103N (7 individuals at 20% viral threshold; 10 individuals at 1% viral threshold) detected within individuals with study-defined pretreatment NNRTI resistance (Figure 3.2).

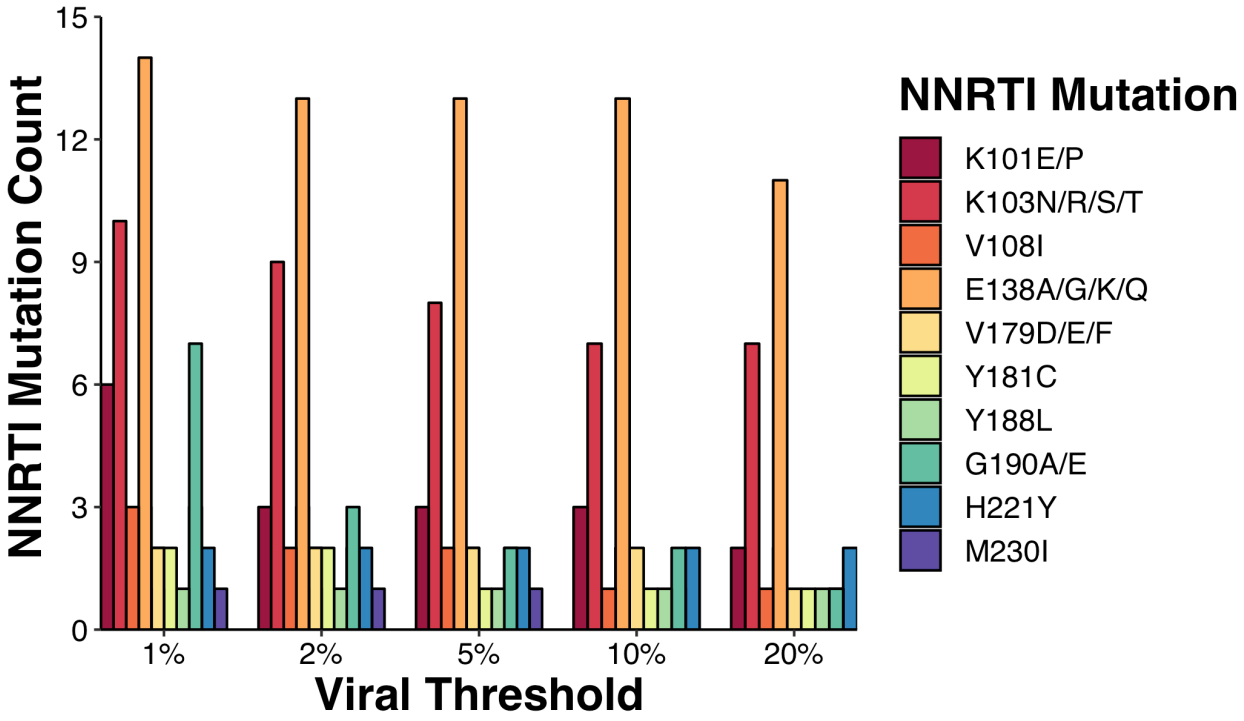


Figure 3.2 High prevalence of E138A and K103N resistance-associated mutations observed at different viral thresholds.
High prevalence of E138A and K103N continuously observed at viral thresholds of 1%, 2%, 5%, 10%, and 20% among ART-naive individuals with study-defined pretreatment NNRTI resistance. As viral threshold decreased, the number of individuals with observed NNRTI mutations increased and subsequently, the number of individuals observed with study-defined pretreatment NNRTI resistance increased.

3.3.5 Increasing Mixture Nucleotides at Low Viral Thresholds Subject to Artefactual Interpretations

During inspection of the specific NNRTI DRMs present at the different viral thresholds, it became apparent that at times, DRMs appeared to actually decrease in observed prevalence at lower minority species populations within individuals. This physically impossible observation triggered a closer examination of the consensus sequences produced at the different viral thresholds and their resistance interpretation.

In total, there were 10 cases of minority species DRMs observed at higher viral thresholds absent at lower viral thresholds within individuals' resistance profiles. In total, there were six, three, and one DRM(s) present at higher viral thresholds absent at 1%, 2%, and 5% viral threshold, respectively (Figure 3.3A). The 10 instances of absent minority species DRMs were added manually to each respective patient's resistance profile prior to data analysis and are summarized in Table 3.4.

The 10 cases of absent minority species DRMs represent an artefact due to the interpretation of mixture nucleotide combinations within consensus sequences by Stanford's HIVdb (Appendix IV). The interpretation of increasing numbers of mixture nucleotides at low viral thresholds resulted in the following: 1) failure to identify DRMs, which were present in the consensus-based sequence, due to increasing numbers of emerging amino acids (e.g. YWK at position 188 results in amino acids L/F/Q/H/Y and none are recognized as a resistance-associated amino acid change), or 2) complete or partial alignment failure of the consensus sequence with reference sequences.

Partial alignment failure resulted in nucleotide trimming of the consensus sequence and consequences observed were the removal of DRMs and failure to interpret DRMs present in untrimmed regions. Although full-length consensus sequences (N=234, codons 90-234) at 1%, 2%, 5%, 10%, and 20% viral thresholds were provided to Stanford's HIVdb, as viral threshold decreased, the number of full-length

sequences detected decreased (20% viral threshold, n=234; 1% viral threshold, n=152) (Figure 3.3B). Similar results were observed during resistance interpretation of nucleotide mixtures present at codon 184, a nucleoside reverse transcriptase inhibitor (NRTI) resistance-associated codon, in an HIV HXB2 backbone sequence (GenBank accession number: K03455) by the geno2pheno algorithm (Table 3.5). Mixture nucleotide combinations were manually added to codon 184 of the reverse transcriptase gene within an HXB2 viral sequence and subsequently analyzed by geno2pheno.

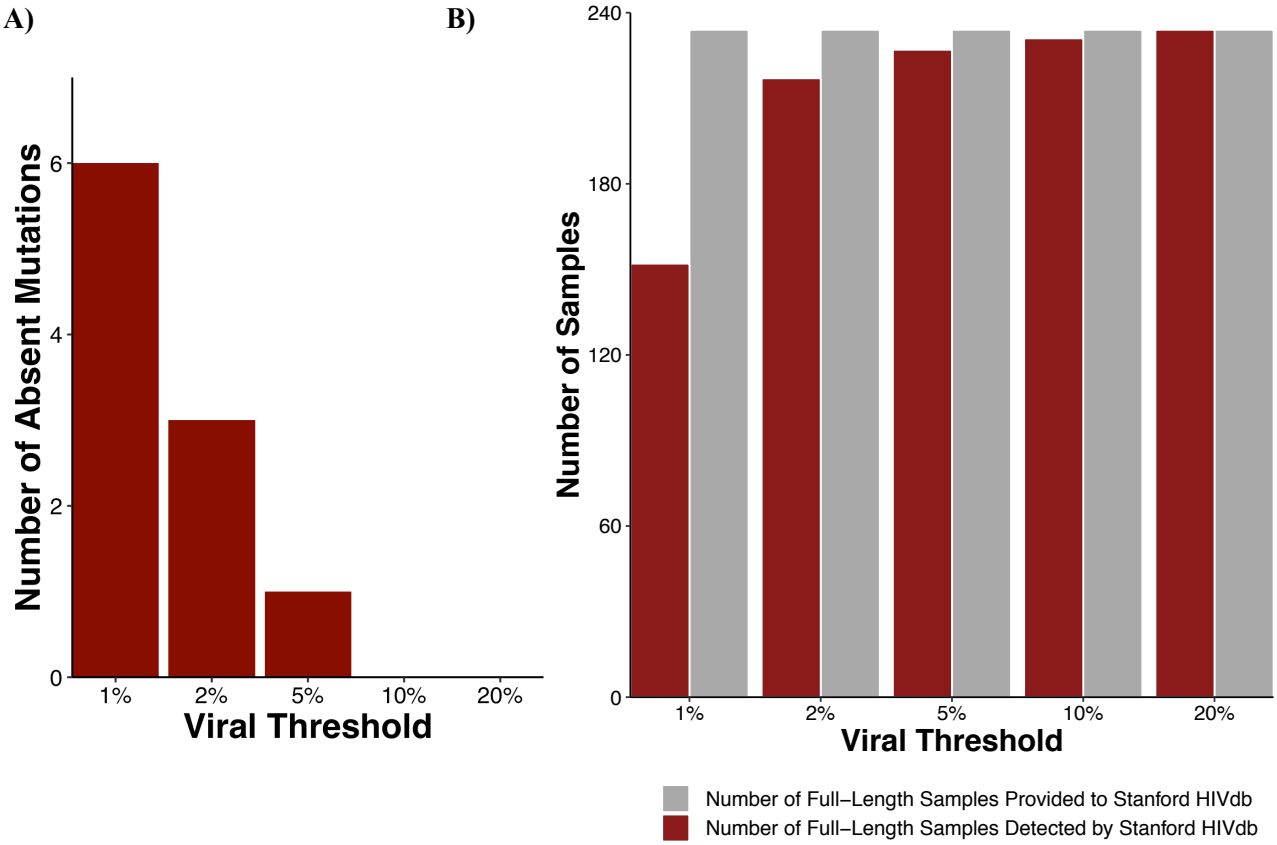


Figure 3.3 Ambiguous interpretation of increasing mixture nucleotides at low viral thresholds resulted in under-reporting of drug resistance-associated mutations.
A) There were 10 cases of minority species drug resistance-associated mutations (DRMs) detected at higher viral thresholds not detected at lower viral thresholds in individuals with study-defined pretreatment NNRTI resistance, an artefactual result of increasing presence of mixtures nucleotides at resistance-associated codons interpreted using standard interpretation algorithms. In total, there were six, three, and one minority species DRM not identified at 1%, 2%, and 5% viral threshold, respectively. **B)** Increasing mixture nucleotide combinations at low viral thresholds result in sequence trimming by Stanford's HIVdb which may subsequently lead to the removal of present DRMs and/or the inability to identify present DRMs in untrimmed regions.

Table 3.4 Summaries of absent drug resistance-associated mutations at low viral thresholds.

Instance	Individual	Amino Acid	Nucleotide Mixture Combinaton[a]	Detectable Viral Threshold[b]	Reason
1	P1	G190E	GRA	5%	Full-length sequence not detected (90-177 recognized at 5% viral threshold); Although GRA present in the submitted and in the detected, trimmed consensus sequence, it was unable to be identified as G190E through Stanford's HIVdb.
2	P1	G190E	GRA	2%	Full-length sequence not detected (90-174 recognized at 2% viral threshold); Although GRA present in the submitted and in the detected, trimmed consensus sequence, it was unable to be identified as G190E through Stanford's HIVdb.
3	P2	V179F	DTW	2%	Full-length sequence not detected (90-177 recognized at 2% viral threshold); DTW nucleotide combination not recognized because sequence length was trimmed. Sequence was not trimmed as a result of DTW mixture.
4	P3	V179F	DTW	2%	Full-length sequence not detected (90-177 recognized at 2% viral threshold); DTW nucleotide combination not recognized because sequence length was trimmed. Sequence was not trimmed as a result of DTW mixture.
5	P1	G190E	GDW	1%	Full-length sequence not detected (90-170 recognized at 1% viral threshold); GDW nucleotide combination not recognized because sequence length was trimmed.
6	P2	V179F	DTW	1%	No reverse transcriptase gene could be aligned.

Instance	Individual	Amino Acid	Nucleotide Mixture Combinaton[a]	Detectable Viral Threshold[b]	Reason
7	P4	Y181C	KRK	1%	Full-length sequence detected; KRK presents a nucleotide combination that is not recognized by Stanford's HIVdb. KRK results in 8 combinations of nucleotides resulting in amino acids C/D/E/G/Y and stop codons. Most likely, the algorithm is unable to decipher the resistance resulting from one of the 8 combinations (1/8 result in Y181C).
8	P5	H221Y	HAT	1%	Full-length sequence not detected (90-166 recognized at 1% viral threshold); HAT nucleotide combination was unable to be recognized because sequence length was trimmed.
9	P3	V179F	DTD	1%	Full-length sequence not detected (90-154 recognized at 1% viral threshold); Although DTD presents a nucleotide combination that results in 9 combinations of nucleotides resulting in amino acids F/I/L/M/V and 1/9 results in V179F, the sequence was trimmed as a result to also remove this nucleotide combination as poor quality.
10	P6	Y188L	YWK	1%	Full-length sequence detected; YWK presents a nucleotide combination that is not recognizable by Stanford's HIVdb. YWK results in 8 combinations of nucleotides resulting in amino acids L/F/Q/H/Y and stop codons. Most likely, the algorithm is unable to decipher the resistance resulting from 3 of the 8 combinations (3/8 result in Y188L).

[a]Nucleotide combination detected at the relevant codon in the consensus sequence.
[b]Detectable viral threshold refers to the viral threshold with the absent DRM which was subsequently added manually.

Table 3.5 Summaries of geno2pheno resistance interpretation of nucleotide combinations at codon 184 of the reverse transcriptase gene.

Nucleotide Combination[a]	Expected Amino Acid(s)	Interpreted Amino Acid	Interpretation
GTG	M184V	M184V	3TC resistant (56-fold) (ddI, ABC partial)
STG	M184L/V	M184V	3TC resistant (56-fold) (ddI, ABC partial)
VTG	M184L/M/V	M184V	3TC resistant (56-fold) (ddI, ABC partial)
GHM	M184A/D/E/V	M184V	3TC resistant (56-fold) (ddI, ABC partial)
GHV	M184A/D/E/V	Absent	3TC resistant (12-fold) (ddI partial)
GWS	M184D/E/V	M184V	3TC resistant (56-fold) (ddI, ABC partial)
BTG	M184L/V	M184V	3TC resistant (56-fold) (ddI, ABC partial)
BMG	M184A/E/P/Q/S/*[b]	Wildtype	3TC resistant (12-fold) (ddI partial)
SWS	M184D/E/H/L/Q/V	M184V	3TC resistant (56-fold) (ddI, ABC partial)
GNN	M184A/D/E/G/V	Absent	3TC resistant (12-fold) (ddI partial)
VHG	M184A/E/K/L/M/P/Q/ T/V	Absent	3TC resistant (12-fold) (ddI partial)
VMG	M184A/E/K/P/Q/T	Wildtype	3TC resistant (12-fold) (ddI partial)
NNN	M184A/C/D/E/F/G/H/ I/K/L/M/N/P/Q/R/S/T/ V/W/*	Absent	3TC resistant (12-fold) (ddI partial)

[a]Mixture nucleotide combinations were manually added to codon 184 of the reverse transcriptase gene within an HXB2 viral sequence and subsequently analyzed by geno2pheno.
[b]Asterisk symbolizes a stop codon.

3.4 Discussion

The rising prevalence of pretreatment drug resistance remains a threat to the global management of HIV. In this study, we investigated the prevalence of pretreatment low-, intermediate-, and high-level NNRTI resistance at different viral thresholds in a Ugandan population of over 200 individuals using next-generation sequencing and subsequently assessed the predictive virological effect of NNRTI resistance 1-year post-NNRTI treatment initiation. The prevalence of study-defined pretreatment NNRTI resistance found at 20% viral threshold was 12% and increased to 22% at 1% viral threshold. Study-defined pretreatment NNRTI resistance at the viral thresholds tested (1%, 2%, 5%, 10%, and 20%) did not significantly increase likelihood of virological failure 1-year post-treatment initiation. Findings of concern highlight potential complications arising from the use of common resistance interpretation algorithms with consensus-based sequences produced at low viral thresholds.

Our observed prevalence of pretreatment NNRTI resistance at the Sanger-detectable viral threshold of 20% (prevalence of resistance 12%) is similar to those previously reported in a Ugandan population in the PharmAccess African Studies to Evaluate Resistance Monitoring (PASER-M) cohort (prevalence of resistance 13%) [166], but lower than those recently reported in Uganda's national pretreatment drug resistance survey (prevalence of resistance 15%) [78]. Although our definition of NNRTI resistance, defined as a score $\geq$15 based on the Stanford's HIVdb v8.5 to either efavirenz, nevirapine, etravirine, or rilpivirine, is different from Uganda's pretreatment drug resistance survey, which defined NNRTI resistance as a score $\geq$15 based on the Stanford's HIVdb v8.3 to either efavirenz or nevirapine, this most likely is not the reason for the difference in prevalence. Instead, the difference in prevalence may be attributed to the inclusion criteria of the study population for Uganda's pretreatment drug resistance survey which included individuals initiating ART more recently (2016), individuals younger than 18 years of age, and individuals with prior antiretroviral exposure [78]. Even so, our results further support previously identified prevalence of NNRTI resistance in eastern African countries [167].

Nearly 30% of our study population (n=66) had detectable viral RNA load 1-year post-treatment initiation, and these values are similar to those identified in the PASER-M cohort [166]. The presence of pretreatment NNRTI resistance at viral thresholds tested was not a significant predictor of increased likelihood of detectable viral RNA load at 1-year post-treatment initiation. Our findings are comparable to a study assessing the impact of pretreatment minority drug resistance detected at various viral thresholds above 2%, on virological suppression in Kenya [168]. In contrast, a large multicenter study evaluating the effects of Sanger-detectable pretreatment low-, intermediate-, and high-level NNRTI resistance on virological suppression 1-year post-treatment initiation in sub-Saharan Africa identified a strong association between pretreatment resistance that lead to a partially-active ART, and virological failure [166].

The NNRTIs efavirenz and nevirapine are commonly used globally due to their high efficacy and low cost [75,169]. However mutation K103N, which confers high-level cross resistance within the NNRTI class, is selected through efavirenz and nevirapine use [170]. Our analysis of pretreatment samples from the ART-naive Ugandans in this study revealed potential, transmitted high-level resistance to their upcoming prescription of efavirenz- or nevirapine-containing therapy, as a result of the detection of mutation K103N. Although mutation E138A is selected with the use of second-generation NNRTIs etravirine and rilpivirine, it is a polymorphic mutation more frequently detected in subtype C viruses compared to subtype B viruses [171,172]. Etravirine and rilpivirine are not used extensively in Uganda as first-line therapy, but the identification of E138A may present future treatment complications with rilpivirine or etravirine in the Ugandan population, which may be generalized to regions with high prevalence of non-subtype B virus [173].

With increasing versatility of next-generation sequencing technologies and their investigation for use in low- and middle-income countries [150], the use of next-generation sequencing with current common resistance interpretation algorithms such as Stanford's HIVdb or geno2pheno, may lead to underrepresented

mutations and subsequently, underreporting of resistance. For example, during our analysis we accounted for unidentified minority species DRMs present at high viral thresholds yet absent at low viral thresholds, and manually added them to the resistance profiles of each individual patient, thereby reporting the true prevalence of study-defined NNRTI resistance. Although our reported prevalence of pretreatment NNRTI resistance would not have differed at 5%, 10%, and 20% viral thresholds, the prevalence of study-defined NNRTI resistance would have been underreported at 1% and 2% viral thresholds. It is important to note that while we encountered absent NNRTI mutations as viral threshold decreased (Appendix IV. A and B), potential false-positive DRMs may also be detected in consensus-based sequences (Appendix IV. C). For example, a consensus-based sequence of viruses with nucleotides "AAG" and "CGC" at codon 103 of the reverse transcriptase gene will result in mixture nucleotides "MRS", and their subsequent translation into possible amino acids, H/K/N/Q/R/S, will create fictional resistance-associated amino acid substitutions, K103H, K103N, and K103S, not present in the original viral population. This variation in resistance interpretation may be avoided if: 1) consensus sequences obtained at viral thresholds <10% are not used for resistance interpretation, 2) resistance interpretation algorithms incorporate increased guidelines specifying restrictions for the use of sequences with high prevalence of mixture nucleotides, or 3) individual read sequences are translated prior to the formation of consensus sequences.

There are a number of limitations within this study. First, our assessment of NNRTI DRMs are excluded to a segment of the reverse transcriptase gene and may limit our ability to adequately determine the true prevalence of NNRTI resistance as we do not account for potential NNRTI DRMs outside of codons 90-234. Secondly, our definition of NNRTI resistance comprises of cumulative scores of DRMs which confer low-, intermediate-, or high-level resistance to NNRTIs and therefore we did not identify individual DRMs which may be polymorphic with scores <15 on Stanford's HIVdb. In addition, the samples examined within this study were collected between 2005-2013 and the identified low, intermediate-, and high-level NNRTI resistance may not have significant relevance to the current prevalence of NNRTI resistance in

Uganda. Finally, our analyses of virological response do not account for adherence to treatment, a crucial variable in the assessments of individual response to treatment and viral suppression status.

3.5 Conclusion

In conclusion, although the observed prevalence of study-defined pretreatment NNRTI resistance increased as viral threshold decreased, the presence of NNRTI resistance at 1%, 2%, 5%, 10%, and 20% viral threshold failed to predict detectable viral RNA load at 1-year post-treatment initiation. This lack of effect may be due to compensatory mechanisms of the backbone NRTI regimen, due to their relatively different mechanism of action from NNRTIs. Perhaps the most significant finding of this study highlights the potential complications of the use of resistance interpretation algorithms in examining consensus sequences assembled at low viral populations. The use of consensus sequences with Sanger-level resistance interpretation algorithms in previous studies observing the impact of minority species resistance levels may have created artefacts which lead to over- or under-reporting of resistance levels, and these studies should perhaps be revisited.

4 Chapter Four: Phylogenetic Surveillance of Travel-Related Zika virus Infections Through Whole-Genome Sequencing Methods

4.1 Introduction

In 2018, the World Health Organization (WHO) held the second annual review of R&D Blueprint, identifying the Zika virus (ZIKV) as a pathogen that should be prioritized for research and development in public health emergency context, due to its epidemic potential and its lack of sufficient treatment [174]. First identified in Uganda in 1947, ZIKV is a member of the *Flavivirus* genus within the *Flaviviridae* family and is an arthropod vector-borne virus spread primarily through infected *Aedes* mosquitoes [175]. Sequencing analyses involving whole-genome and gene-specific analyses have identified three lineages; Asian, East African, and West African [175]. Phylogenetic and molecular clock analyses have confirmed the Asian lineage responsible for the recent sporadic spread of ZIKV outside of Africa and Asia: Yap Island, Federated States of Micronesia in 2007; French Polynesia in 2013; Brazil in 2015 [176–178]. The emergence of ZIKV in recent years has been associated with increased incidence of the neurological conditions Guillain-Barré syndrome and meningoencephalitis, and prenatal microcephaly [179–182]. Similar to infections by other members of the *Flavivirus* genus, dengue virus and chikungunya virus, individuals infected with ZIKV are generally asymptomatic, with only one in five infected individuals showing non-specific symptoms such as a mild fever, rash, and conjunctivitis [179,183].

The majority of ZIKV infections in regions without significant prevalence of ZIKV vector *Aedes* mosquitos, such as Canada, are travel-acquired infections [184]. Routine surveillance to identify and track new cases of ZIKV infections currently rely on suspicions of ZIKV infection by healthcare providers, and additional traditional laboratory confirmation through real time reverse transcription-polymerase chain reaction (RT-PCR) amplification and antibody-based tests [93,179]. However, ZIKV may be misdiagnosed

as other, closely related infections due the non-specific nature of ZIKV symptoms. Moreover, laboratory confirmation through real time RT-PCR can be limited in low-resource settings while serological antibody-based tests may be cross-reactive among *Flaviviruses* leading to further misdiagnosis of ZIKV infections [93].

Whole-genome sequencing (WGS) methods have previously been used to analyze the evolution and genomic variability of viruses such as hepatitis C virus (HCV) [185], and are progressively being used during viral outbreaks in an effort to identify transmission patterns [186–189]. Sequencing of ZIKV presents an alternative or supplementary method of ZIKV surveillance and may allow insight into the geographic origins of infections, transmission patterns, and genomic diversification. In this study, through phylogenetic analysis of whole-genome sequences derived from patients with confirmed travel-acquired ZIKV infection, we attempt to identify the origins of travel-acquired ZIKV infections and, through intraspecies and intrafamilial comparative analyses with HCV, determine ZIKV's genomic variability.

4.2 Methods

4.2.1 Study Population

Specimens from five subjects with confirmed travel-acquired ZIKV infection (putatively from Belize, Mexico, an undisclosed Caribbean region, Barbados, and Panama) were obtained from the British Columbia Centre for Disease Control (BCCDC) Public Health Laboratory and the samples had a range of cycle threshold (Ct) values (21 – 33).

Three longitudinal whole-genome HCV sequences were provided from one subject living with HCV genotype 1a strain from the Vancouver Injection Drug Users Study (VIDUS), a prospective ongoing cohort established in 1996 involving people who inject illicit drugs in Vancouver's Downtown Eastside neighbourhood. The VIDUS participants provide behavioural and demographic data through completion of

questionnaires at bi-annual study visits which also involve HCV and human immunodeficiency virus (HIV) antibody testing [190].

The University of British Columbia Providence Health Care Research Ethics Board granted ethical approval for this study (H16-02865). All experiments were performed in accordance with institutional guidelines and regulations. A waiver of consent was obtained as research was conducted on anonymized leftover, stored clinical specimens and there were no direct benefits to the participants as infections had previously been diagnosed and reported.

4.2.2 Whole-Genome Sequencing

Confirmation of ZIKV infection and viral nucleic acid extraction occurred at the BCCDC while RT-PCR, WGS, and data analysis occurred at the BC Centre for Excellence in HIV/AIDS (BC-CfE) in Vancouver BC, Canada. Detailed description of the methodology and primer sequences used can be found elsewhere [191]. Briefly, sequencing of ZIKV genome was performed on an Illumina MiSeq platform (San Diego, California, United States) using a previously published procedure designed to overcome some of the limitations of low viral RNA copy number and partially degraded samples by amplifying several short amplicons to create a tiling path across the ZIKV genome [191]. This methodology involves a multiplex PCR system which amplifies alternating regions of the ZIKV genome using two separate primer pools. The purpose of this is to amplify approximately 400 nucleotide regions which overlap by 100 nucleotides, providing consistent coverage throughout the genome. In total, 35 primers in two separate primer pools were provided by Dr. Nick Loman from the University of Birmingham to span the ZIKV genome (10,807 base-pair [bp]). KAPA Hyper Library preparation kit from Kapa Biosystems (Wilmington, Massachusetts, United States) and duel-indexed adapters from Illumina were used for MiSeq library construction [165].

4.2.3 Data Analysis

Sequencing analysis

Sequences generated from the MiSeq were quality trimmed and primers were removed using Trimmomatic, aligned to each respective reference sequence using BWA-MEM, and processed with SAMtools [60,61]. Paired-end reads were mapped to a set of 17 phylogenetically distinct whole-genome reference sequences obtained from GenBank using the default settings for local alignment. This was followed by a re-alignment to the single reference with the highest overall read count for each sample and, subsequently using custom Python scripts, consensus sequences were obtained from MPILEUP files using base quality and mapping quality thresholds of 15. For HCV sequences, consensus sequences were similarly generated using an alternative reference database with full-length HCV genomes (n=3,162) from the Los Alamos HCV database [192].

Phylogenetic analysis

Phylogenetic analysis was performed to investigate geographic clustering of travel-acquired ZIKV cases. Whole-genome sequences were compiled into a reference set and were retrieved from GenBank (nucleotide search details: ‘zika[All Fields] AND "Zika virus"[porgn:__txid64320] AND (viruses[filter] AND ("9000"[SLEN] : "11000"[SLEN]))). In total, 864 nucleotide sequences ranging in length (9,000-11,000 bp) were obtained in FASTA format. Nucleotide sequences with partial genomes, large repeated regions of ambiguous nucleotides, sequences with missing geographic origins, and confirmed non-ZIKV sequences were excluded from the dataset. In total, 362 sequences were used as reference sequences from different countries of interest (n= 39) (Appendix V). The United Nations geoscheme was used to identify geographical regions of the countries of interest [193].

Patient sample and reference sequences were aligned with the multiple sequence aligner MUSCLE [70] using the parameters for large dataset alignment. Sequences were trimmed to WHO ZIKV reference sequence (GenBank accession number: KX369547) [194] for consistency, using alignment viewing

program Aliview [195]. Phylogenetic trees were inferred in a maximum likelihood framework under a General Time Reversible (GTR) substitution model with gamma-distributed rate heterogeneity using RAxML version 8.2.12 [72]. The reliability of the phylogenetic topology was estimated using a bootstrap analysis with 1000-pseudoreplicates. Figure 4.1 illustrates the method of detecting travel-acquired cases through sequence similarity.

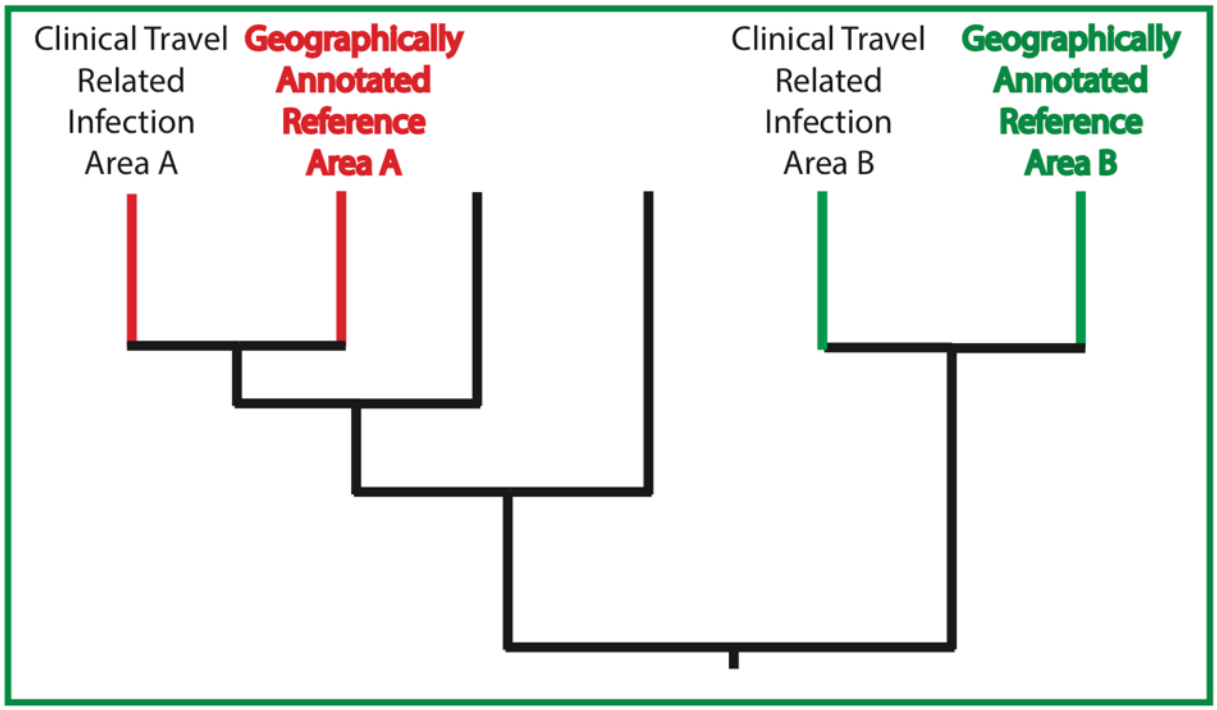


Figure 4.1 Outline of phylogenetic analysis and the identification of origins of infection.
Predicted phylogenetic association between ZIKV lineages derived from travel-related clinical cases and geographic reference sequences. Clinical travel cases from areas A and B are predicted to be observed in the phylogeny most closely related to reference sequences from the same areas.

Genomic diversity of ZIKV travel-acquired cases

Variation across the ZIKV genome was measured for each nucleotide position with a minimum read depth of 100. To compare the overall fraction of minority variants with the dominant variant at each position, the average percentage of all variants for each sample was calculated at each position and plotted in R using the ggplot2 package [196]. Variants present at >99% and >1% of the reads mapped at a nucleotide

position were classified as dominant and minority variants, respectively. Variants present at $\leq$ 1% were classified as noise.

Further, a Shannon diversity index was used to characterize and compare the intrafamilial genomic diversity of ZIKV to the HCV sequences acquired from one individual sampled longitudinally at three timepoints. Shannon diversities were then calculated per position from the nucleotide proportions listed above using custom Python scripts (Appendix VI). A sliding window approach was used for visualization purposes where the diversities were summed for each 100 bp window. These were then plotted in R using the ggplot2 package [196].

Patristic distance threshold for phylogeographic analysis

Patristic distances for the phylogenetic tree were obtained using the Python package DendroPy [197]. Countries with smaller intra-country patristic distances relative to inter-country patristic distances indicated country-specific sequences were more closely related to each other compared with other between-country sequences. The countries for each sequence and its closest relative were also identified and compared.

Genomic variation of whole-genome ZIKV sequences

To assess genomic diversity of current circulating ZIKV strains based on geographical region, variable nucleotides unique to each country were identified. A sequence alignment trimmed to protein-coding regions of ZIKV was performed on the genomes dated after 2000 from the reference dataset, as well as on the travel-related ZIKV samples. A consensus sequence was obtained for each country that had >1 whole-genome sequence (n=24) and countries with a single representative sequence were also included in the analysis (n=12). Countries with reference sequences acquired prior to the year 2000 were excluded (n=5) to better investigate the current circulating ZIKV strain. Unique positions were initially identified if they were present in at least 50% of the reference sequences corresponding to a specific country, while not

identified in consensus sequences corresponding to any other country in the dataset. Subsequently, these variants were further screened to assess whether they were unique relative to any individual sequences from alternative countries by comparing each variant to all nucleotides at the respective position.

4.3 Results

4.3.1 Whole-Genome Sequencing Results from Travel-Acquired ZIKV Infections

Samples with higher ZIKV Ct values had higher number of human reads and lower percentage of ZIKV reads (Pearson's correlation: -0.99, $p < 0.05$), consistent with a lower absolute amount of ZIKV in the samples. Higher Ct values were also associated with lower overall depth of ZIKV coverage (Pearson's correlation: -0.88, $p < 0.05$) (Table 4.1). Each sample's consensus sequence ranged in length from 8-10.5 kilo-base-pair (kbp). Median depth of coverage of all samples was 24,000 reads (IQR: 17,000-25,000). Although some contigs had low depth of coverage (fewer than 10 reads), they still provided sufficient genome coverage for the regions sequenced.

Table 4.1 Sequencing results of five samples with confirmed travel-acquired ZIKV infection.

Sample[a]	Ct Value[b]	No. of Reads[c]	No. of Human Reads	No. of ZIKV Reads	% ZIKV Virus	Median Depth of Coverage[d]
Sample 1	33	3,342,514	737,467	1,442,159	43	673
Sample 2	28	3,052,866	7,863	2,532,761	83	25143
Sample 3	30	3,088,142	73,386	2,427,755	79	23884
Sample 4	26	2,046,216	1,774	1,672,980	82	16832
Sample 5	21	3,282,070	4,388	2,629,877	80	32939

[a]Samples were provided by the BCCDC and were anonymized again at the BC-CfE
[b]Cycle threshold (Ct) was obtained through quantitative RT-PCR
[c]Total number of human, virus, and random reads obtained per sample
[d]Median depth of coverage was calculated using the median number of reads across the genome for each sample

4.3.2 Phylogenetic Analysis of Travel-Acquired ZIKV Infections

Phylogenetic analysis of the resulting consensus sequences with whole-genome reference sequences obtained from GenBank confirmed the clinically indicated region of ZIKV infection for one (Sample 2) of the five samples (Figure 4.2). Sample 1 and Sample 4 were missing whole-genome reference sequences from clinically indicated areas of infection. However, they clustered within close geographical proximity to neighboring regions (Sample 1 had suspected travel-acquired ZIKV from Belize and clustered closely to samples from neighboring country Nicaragua; Sample 4 had suspected travel-acquired ZIKV from Barbados and clustered closely to samples from Colombia) (Figure 4.2). Although Sample 3 clustered closely with samples from the United States of America (USA), it did not have a known country of suspected infection and therefore phylogenetic analysis could not confirm or refute the suspected country of infection. Sample 5 failed to cluster to reference sequences from Panama, its suspected area of infection despite having full genome coverage. Relatively short branch length of the GenBank sequences and the travel-acquired ZIKV sequences implied limited genetic divergence between sequences from recent ZIKV cases [69].

Figure 4.2 Phylogenetic analysis of five samples with confirmed Zika infection collected by the BCCDC.

Phylogenetic analysis of consensus sequences from travel-related clinical cases of ZIKV infection with

geographically annotated reference sequences confirmed the suspected regions of ZIKV infection for one (Sample 2) of the five samples. Two samples, Sample 1 (Belize) and Sample 4 (Barbados), were missing whole-genome sequences from suspected areas of infection. However, they clustered within close geographical proximity to neighboring regions. Values for bootstraps greater than 70 are shown at the nodes. Scalebar shows the genetic difference through number of substitutions per site.

4.3.3 Genomic Diversity of Travel-Acquired ZIKV Infections

Nucleotide variation

Nucleotide variation was measured for each Canadian sample as the average proportion of minority variants relative to the WHO ZIKV reference sequence. Shown in Figure 4.3, is the nucleotide variation averaged over the five Canadian samples sequenced. Although minority variants were present throughout the ZIKV genome, of the 10,108 positions with at least 100 bp coverage, only 504 (4.9%) displayed any evidence of variation. In general, an overall relatively low nucleotide variation was observed in the patient sample set despite the samples apparently having five different origins of ZIKV infections.

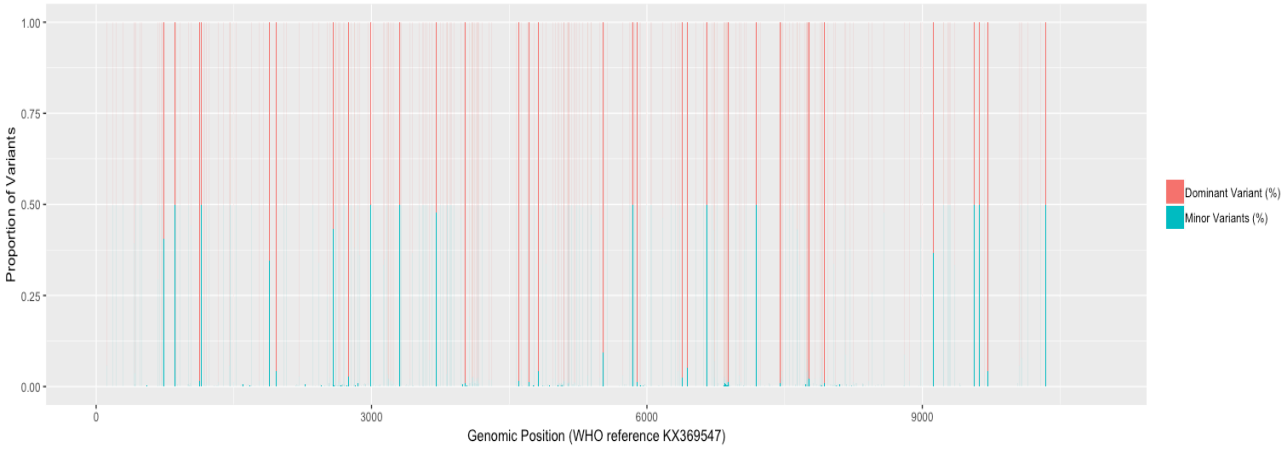


Figure 4.3 Nucleotide variation of samples.
Nucleotide variation was measured for each sample as the average proportion of dominant variant (consensus) to minor variants. Shown in Figure 4.3, is the nucleotide variation averaged over the samples sequenced in the study. In general, an overall relatively low nucleotide variation was observed despite allegedly having five different origins of infections. This suggests that high sequence similarity may present an issue in identifying origins of ZIKV infections.

Mean Shannon diversity index

Diversity across the ZIKV genome was also calculated using the mean Shannon diversity index for all of the Canadian samples (Figure 4.4) and the five ZIKV samples showed a relatively low diversity index when compared to the HCV samples.

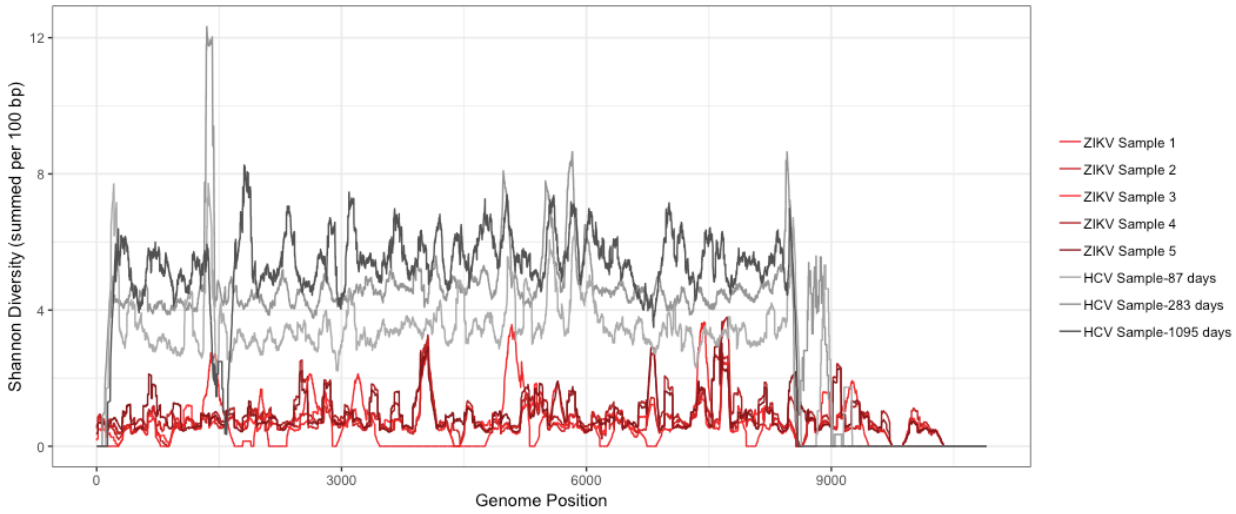


Figure 4.4 Shannon diversity index of ZIKV and HCV samples.
The Shannon diversity index for ZIKV samples (n=5) and HCV samples (n=3) shows a relatively low diversity index of ZIKV samples compared to HCV samples.

4.3.4 Country-Specific Genomic Diversity

Patristic distance

Due to the challenges for geographic placement of the limited dataset of five ZIKV-acquired infection sequences, the remaining reference sequences selected in this study were also examined to determine the limitations of the phylogeographic signal in general for ZIKV consensus sequences. Here, the accuracy of the identification of each respective country for a given reference sequence was examined using a country-specific patristic distance threshold. For each country, specific thresholds were generated where the largest within-country phylogenetic distance (patristic distance) was compared with all between-country phylogenetic distances. When applying these country specific thresholds, only seven countries (Cambodia, Senegal, Peru, Uganda, South Korea, Nigeria, and Malaysia) were shown to have within-country patristic distances that were smaller than those distances between all other countries. Of these, four

(Malaysia, South Korea, Peru, and Cambodia) belonged to the Asian lineage responsible for the recent epidemic. When examining the countries of the closest relatives for each of the 367 samples in the phylogenetic tree, only 252 (68.7%) were isolated from the same country (a sample's closest relative being another sample from the same country).

Nucleotide variation

To better understand the extent of ZIKV's genetic variability as ZIKV infections spread between continents in the 21st century, nucleotide diversity of whole-genome sequences isolated after the year 2000 were quantified, aligned and rooted to the WHO reference (Figure 4.5). The number of variable positions decreased as ZIKV dispersed from Asia to the Americas. The number of variants unique to each country, and not present in any sequences related to any other country, is further depicted in Figure 4.6. With geographical expansion and north- and southward spread of ZIKV, the number of nucleotides unique to each country decreased. In Asia, the highest number of unique variants were found in Micronesia (n=37) followed by Indonesia (n=34) and the lowest were found in Singapore (n=1). In the Americas, the highest number of unique variants were found in Ecuador (n=8), and the lowest were found in Venezuela (n=1) (Figure 4.6).

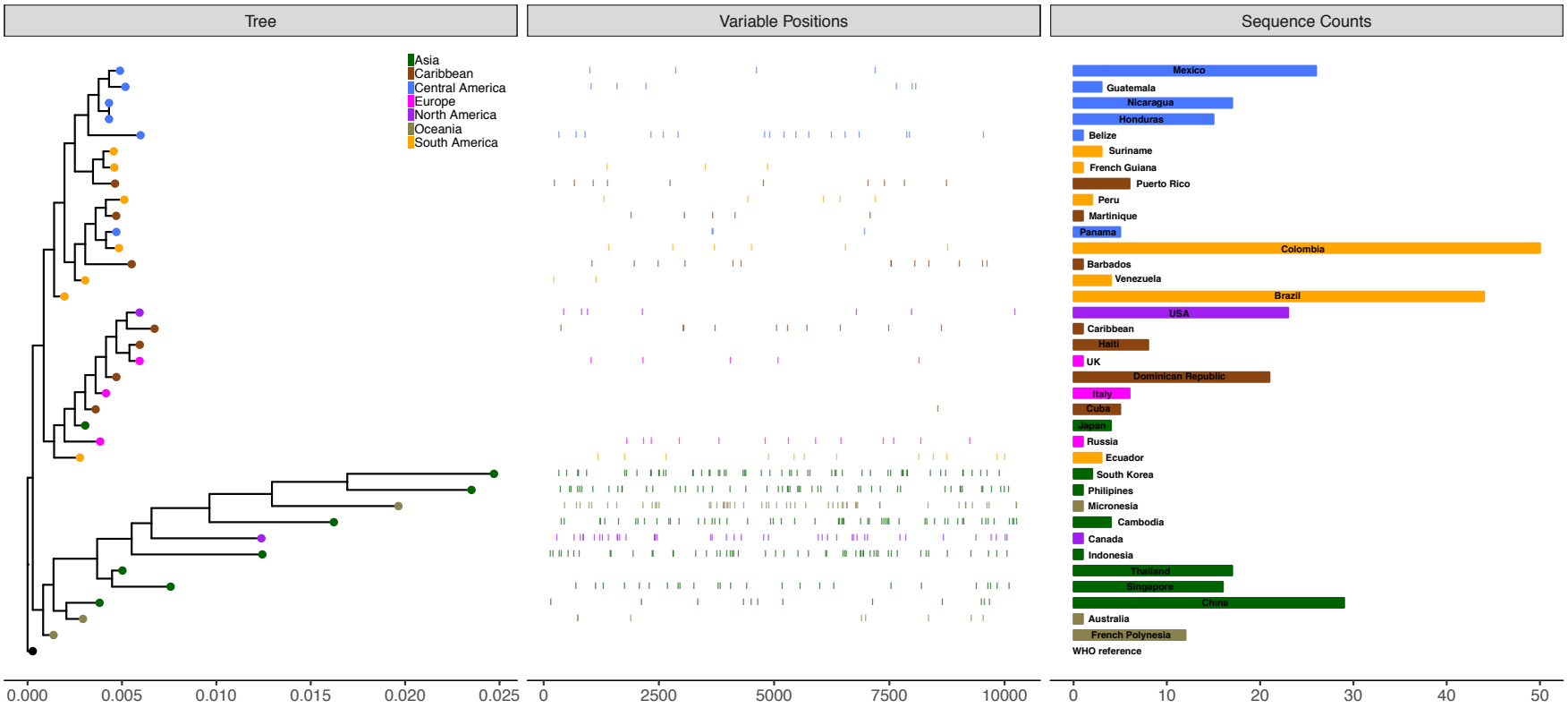


Figure 4.5 Nucleotide variability of ZIKV sequences after the year 2000 in dataset.
Analysis of the genomic diversity of circulating ZIKV strains after the year 2000 reveal reduced genetic variability as the geographical expansion of ZIKV occurred from Asia to the Americas. Whole-genome sequences were aligned and rooted to the WHO reference. Unique positions were identified if they were present in at least 50% of the sequences corresponding to a specific country but not present in the consensus sequence corresponding to any other country.

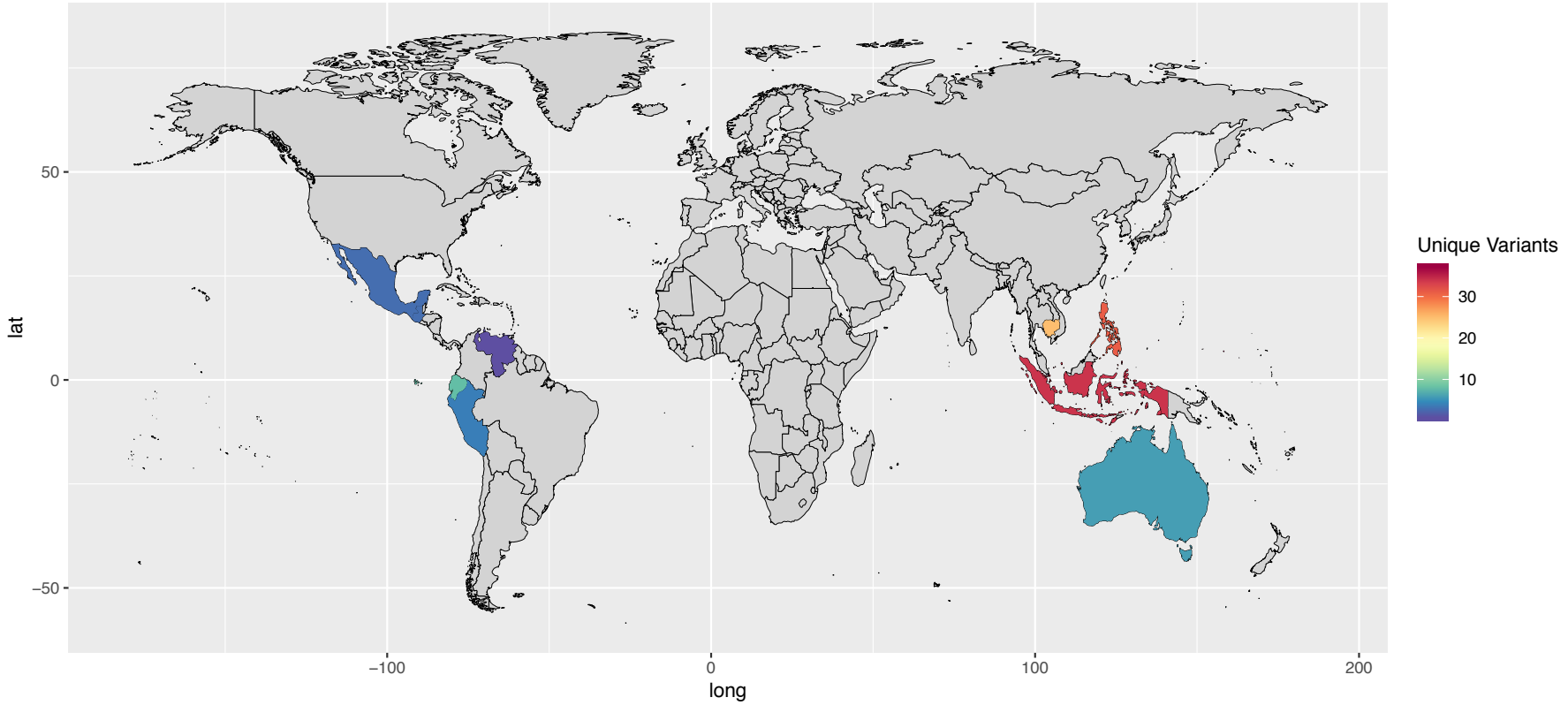


Figure 4.6 Nucleotide variants unique to each country.
The number of nucleotide positions unique to a country is shown. With geographical expansion into the Western Hemisphere and north- and southward spread of ZIKV, the number of variants identified in whole-genome sequences decreased.
Abbreviations: lat, latitude; long, longitude.

4.4 Discussion

Surveillance of ZIKV is essential to the development of public health guidelines to prevent recurring and newly-emergent ZIKV outbreaks, which are associated with growing numbers of neurological abnormalities and birth malformations [198,199]. Currently, ZIKV is detected through serological testing and real time RT-PCR, and these methods may not always be reliable [200,201]. In this proof-of-concept study, we performed WGS on five Canadian samples from the BCCDC with confirmed travel-acquired ZIKV infections, and aimed to verify the original countries of infections. Our findings suggest that although phylogenetic analysis of WGS data can generally identify the broad geographical regions of ZIKV infection, it may not necessarily be used to pinpoint the exact country of infection due to the current circulating ZIKV and its low genomic variability, incomplete and biased coverage of reference sequences in databanks, and complications arising from extensive global travels.

While phylogenetic analysis did correctly verify the country associated with ZIKV infection for one of the five samples, two samples (Sample 1 and Sample 4) were missing whole-genome reference sequences from regions of suspected ZIKV infections. Lack of WGS data from suspected regions of ZIKV infections on public, global databases does not necessarily indicate absence of ZIKV transmission in these populations. Limited resources in countries with suspected ZIKV infection [202] and underreporting of ZIKV infections due to asymptomatic patients [203] may contribute to the lack of data available on online resources even in the presence of local ZIKV transmission. Furthermore, inconsistent annotations of WGS data on public sequence databases such as GenBank, and the lack of curated and standardized available reference sequences from countries with suspected ZIKV infections [204] can hinder accurate ZIKV identification and complicate surveillance attempts. For example, the generic qualifier “country” available for uploaded nucleotide sequences on GenBank is defined as the location pertaining to the isolation of the sequenced organism [205], which may be interpreted as the location where sequencing of an organism took place as opposed to the location where an infection may have occurred. This is likely the case of Sample 3

in our sample set, which clusters within a group of samples with "country" specified as the USA. However, upon closer inspection, some of these ZIKV infections from the USA cluster were likely travel-related. For example, although the country for sample sequence KX842449 is recorded as USA, the sample is from an individual recently returning from Cuba. Additionally, the WHO reference sequence (10,769 bp) used in many publications does not cover the full ZIKV genome (10,807 bp) and its protein length differs from the NCBI reference genome (NC_012532) [204]. This may in turn complicate subsequent ZIKV protein annotations. Sequence databases may benefit from adopting additional features such as "suspected region of infection" for sequences of infectious pathogens, in order to maximize sequence utility and research reproducibility. Additional analysis performed in this study also showed that >30% of the reference sequences shared their closest relative with samples isolated from different countries further suggesting that in general, phylogeographic analysis of ZIKV may present unique challenges relative to other more rapidly evolving viruses such as HCV.

Low genomic variability of ZIKV may present additional obstacles during global surveillance of ZIKV. Although our five ZIKV samples had five different sources of infections, there were high sequence similarities between the samples and the WHO reference sequence with respect to the identification of minority, non-consensus nucleotides. The diversity of ZIKV was also substantially lower relative to HCV suggesting limited evolutionary selective pressures. Large segments of highly conserved regions of the genome, lack of a RNA dependent RNA polymerase with similar fidelity of RNA synthesis compared to HCV, and lack of drug pressure may be reasons for the low genomic variability observed in our ZIKV samples [206,207]. Fortunately, low genomic variation is likely an advantage in vaccine and therapeutic developments [208].

While investigating the genetic diversity of all ZIKV sequences based on country and global region where each sample was identified and/or sequenced, a fewer number of unique nucleotides were identified and lower variation were observed in the countries located in the Western Hemisphere. Our results further

corroborate and support previous findings and predictions that indicate the evolutionary progression of ongoing epidemics may decrease as time passes [209]. Sequences from USA and neighboring regions such as Haiti and Dominican Republic have high sequence similarity, compared to sequences from the Asian regions such as South Korea, Thailand, or Cambodia. Our findings reinforce the notion of splitting of the Asian lineage into two distinct lineages, American and Asian, as suggested by Gubler *et al*. 2017 [210].

There are several limitations within our study that may be addressed by future research. Firstly, the low number of patient samples within our dataset may impact the validity of our findings regarding genomic variability of ZIKV. With a larger sample size or a greater global sampling, we will likely be able to better interpret the use of phylogenetic analysis in identifying origins of travel-acquired ZIKV infections. Perhaps one of the most important limitations is the absence of whole-genome reference sequences available from countries associated with travel-related ZIKV infection. Although we were unable to confirm the country of infection for two of five samples due to missing reference sequences, we were able to identify the broad geographical region of infection. Low genomic variation of ZIKV can potentially distort phylogeographical inference.

4.5 Conclusion

Surveillance of ZIKV is dependent on healthcare providers and their vigilant and persistent inspection of individuals who may have been exposed to ZIKV. Even individuals who are asymptomatic can potentially transmit ZIKV to unsuspecting partners through sexual transmission [211,212]. Continual surveillance of ZIKV is essential to early detection of ZIKV outbreaks, particularly in areas where *Aedes* mosquito populations are extant and growing. While WGS of ZIKV may present obstacles during epidemic surveillance, mostly due to lack of standardized available data during phylogeographic analysis and low ZIKV genomic diversity, it provides a unique method of identifying broad geographical regions of an infection and can also provide insight into the genetic variability of a circulating virus.

5 Chapter Five: General Discussion and Conclusion

5.1 Thesis Summary

5.1.1 Prevalence of Human Immunodeficiency Virus-1 Integrase Strand Transfer Inhibitor Resistance in British Columbia, Canada Between 2009 and 2016: A Longitudinal Analysis

The prescription of integrase strand transfer inhibitors (INSTIs) raltegravir, elvitegravir, and dolutegravir has increased dramatically since their introduction in Canada. Chapter Two describes the increasing trend of the prevalence of intermediate- to high-level INSTI resistance, as defined by a score ≥30 based on the Stanford University HIV Drug Resistance Database (HIVdb) Genotypic Resistance Interpretation Algorithm, in the ART-treated population within the DTP between 2009-2016. Although the INSTI dolutegravir, has been shown to have a high genetic barrier to resistance, INSTIs raltegravir and elvitegravir have comparable resistance patterns to NNRTIs, and INSTI resistance should be interpreted with similar caution.

Our results indicate the prevalence of INSTI resistance was low but gradually increasing in the DTP (1 to 7 per 1000 ART-treated individuals) between 2009 and 2016. The proportion of INSTI-treated individuals receiving Sanger-based, INSTI resistance testing in BC increased over time but was overall significantly lower than the proportion of antiretroviral therapy (ART)-treated individuals receiving protease and reverse transcriptase inhibitor (PI-RT) resistance testing. Furthermore, our results suggest an increasing prevalence of mutations at specific codons associated with INSTI resistance being selected within individuals with study-defined INSTI resistance. By the end of the study period, mutation 155H/S/T followed by mutation 263K had the highest, observed prevalence in individuals with study-defined INSTI resistance. Although the number of newly identified individuals with study-defined INSTI resistance per year increased and subsequently decreased during the study period, this most likely is due to the introduction and expanding use of dolutegravir.

Due to current INSTI resistance testing guidelines, the ability to correctly identify the frequency of study-defined INSTI resistance in this study was limited by incomplete INSTI resistance testing, and the prevalence of resistance may be underestimated.

5.1.2 Sensitive Detection of Pretreatment Minority Species HIV-1 NNRTI Resistance in Uganda: Limited Benefits and Analytical Pitfalls

The paradigm of HIV treatment is shifting from nonnucleoside reverse transcriptase inhibitor (NNRTI)-containing regimens to INSTI- and specifically dolutegravir-containing regimens, largely in regions with high prevalence of NNRTI resistance, such as Uganda. However, this shift in treatment recommendations at the time of this thesis writing, excludes women of reproductive age due to potential development of neural tube defects during pregnancies, which has been observed in individuals receiving dolutegravir-containing treatments [150]. Chapter Three investigated the prevalence of pretreatment NNRTI resistance as defined by a score $\geq$15 based on Stanford's HIVdb, at different viral populations or thresholds of 1%, 2%, 5%, 10%, and 20%, using next-generation sequencing and assessed the predictive virological effect of NNRTI resistance 1-year post-treatment initiation.

The prevalence of study-defined pretreatment NNRTI resistance found at Sanger-level of detection of 20% viral threshold was 12%, similar to reported, observed prevalence of NNRTI resistance obtained from the World Health Organization (WHO) [78]. Study-defined pretreatment NNRTI resistance at the viral thresholds tested (1%, 2%, 5%, 10%, and 20%) did not significantly increase likelihood of virological detection 1-year post-treatment initiation. The most prevalent NNRTI DRMs identified in individuals with study-defined NNRTI resistance were E138A and K103N, and while E138A confers low-level resistance to rilpivirine and is an expected polymorphism in non-subtype B HIV-positive individuals, K103N is capable of conferring high-level resistance to efavirenz and nevirapine, the main NNRTIs prescribed in this Ugandan population.

Perhaps the most significant findings of Chapter Three emphasize the potential complications arising from the use of common resistance interpretation algorithms such as Stanford's HIVdb or the geno2pheno program with consensus-based sequences produced at low viral thresholds. There were 10 cases of drug resistance-associated mutations detected within consensus sequences produced at higher viral thresholds not being detected within consensus sequences produced at lower viral thresholds. This is an artefactual result of increasing mixture nucleotide combinations in consensus-based sequences produced at low viral thresholds, subsequently interpreted using resistance interpretation algorithms. Although this potential shortcoming of resistance interpretation algorithms may be avoided if certain precautions such as translation of read sequences prior to consensus-based sequence formation are taken, past and current research observing the impact and frequency of resistance detected at low viral thresholds using next-generation sequencing technologies should be examined with assumptions of potential misrepresentations of resistance.

5.1.3 Phylogenetic Surveillance of Travel-Related Zika Virus Infections Through Whole-Genome Sequencing Methods

In 2018, the World Health Organization identified the ZIKV as a pathogen that should be prioritized for public health research due to its epidemic potential. The majority of ZIKV infections in regions without endemic *Aedes* mosquito populations are travel-related. In Chapter Four, a proof-of-concept study involving the use of whole-genome sequencing (WGS) was used to examine the evolutionary dynamics of travel-acquired ZIKV infections diagnosed in Canada and to explore the limitations of phylogeographic analysis. Using next-generation sequencing, whole-genome ZIKV sequences were obtained from five Canadians with diagnosed ZIKV infections and phylogeographic analysis was performed to investigate geographic clustering. Genomic variability of ZIKV samples was assessed and for context, compared with HCV samples.

Phylogenetic analysis confirmed the suspected region of ZIKV infection for one of five samples and one sample failed to cluster with sequences from its suspected country of infection. Travel-acquired ZIKV samples depicted low genomic variability relative to HCV samples. A floating patristic distance threshold classified all pre-2000 ZIKV sequences into separate clusters, while only sequences obtained from Asian countries after-2000 were similarly classifiable.

Our findings indicate ZIKV is increasing in its geographical distribution and decreasing in genomic variability in the samples obtained from countries in the Western Hemisphere. Our data further supports previous suggestions of a potentially new lineage of ZIKV in the Americas similar to the Asian ZIKV lineage, yet distinct from the African ZIKV lineage. Low genomic variability, incomplete coverage of reference sequences in databanks such as GenBank, and complications arising from extensive global travels further complicate surveillance interpretation of both ZIKV transmission and evolution, concepts significant in the monitoring of global ZIKV outbreaks.

5.2 Limitations

There are limitations within this thesis, many of which are presented in each individual Chapter. However, it is important to note that within this thesis as a whole, Chapter Two and Chapter Three investigate HIV drug resistance while Chapter Four explores ZIKV diversity and the limits of viral surveillance using next-generation sequencing. Together, these chapters may present potential complications in comprehending the extent to which sequencing technologies may increase our knowledge of viral pathogens due to their variability in research availability and genetic diversity.

While HIV and ZIKV are both viral RNA pathogens, HIV has been studied extensively over the last four decades as a result of numerous international collaborations intent on decreasing the incidence and prevalence of HIV [213], while ZIKV is considered a relatively neglected disease becoming more relevant

over recent years [214]. Therefore, examining of the applications of sequencing technologies for a virus such as HIV, which has established, curated surveillance systems in place may not be comparable to a virus such as ZIKV. Nevertheless, although we identified low genomic variability a reason for the limited use of sequencing in examining phylogeographical patterns of ZIKV, this is not to say that sequencing may not be useful in understanding other aspects of ZIKV.

In addition, extrapolation of the sequencing applications performed in this thesis may not be appropriate for other institutions or other viruses. The research performed in Chapter Two, Chapter Three, and Chapter Four occurred in resource-rich settings with available in-house bioinformatic pipelines, factors not always present during outbreaks or in resource-limited settings. In addition, the genomic variability of viruses such as ZIKV, are constantly changing. Methods and applications used in this study may be required to be adapted to the changing genomic dynamics of viruses in the future. The use of template-specific primers may be replaced with non-specific, random primers currently used in metagenomic studies [37]. Moreover, HIV is a global disease with more than 37 million individuals infected worldwide [215]. While the study populations in Chapter Two and Chapter Three span varying subtypes of HIV-1 – including subtypes A, B, C, and D – study participants in Chapter Two and Chapter Three were individuals $\geq$19 years of age and $\geq$18 years of age, respectively, and generalizations of drug resistance levels from these Chapters, may not be appropriate for younger individuals living with HIV.

5.3 Conclusion: Applications and Future Directions

As infectious viral pathogens increase in geographical distribution, monitoring of the viral genetic code remains a means to characterize viral pathogens, and identify transmission patterns in the process. This thesis presents examples of widely used sequencing methodologies, Sanger sequencing and next-generation sequencing, in testing hypotheses concerning the evolution and transmission of currently circulating viral pathogens, HIV and ZIKV.

Drug resistance remains a threat to the effectiveness of HIV treatment, and regardless of the extent of genetic barriers, antiretrovirals (ARVs) including INSTIs are not impervious to resistance. Genotypic resistance testing for the presence of PI- and/or RT-associated resistance is routinely done at baseline and at times of virological failure in BC. Current ART guidelines do not recommend INSTI resistance testing prior to initiation of treatment unless there is suspected INSTI resistance [55,108]. High prevalence of INSTI polymorphic substitutions, which can reduce INSTI susceptibility by conferring low-level INSTI resistance, have previously been detected in clinical populations within European countries (>14%) [216–218]. However, a nation-wide study observing INSTI resistance patterns in the United States revealed lower proportion of individuals with polymorphic mutations capable of conferring low-level INSTI resistance (<7%) and instead, higher proportion of individuals with mutations capable of conferring high-level raltegravir- and elvitegravir-associated resistance (>14%) [125]. In Chapter Two, there was an observed increase in the prevalence of INSTI resistance within the individuals receiving INSTI-containing regimens in BC. Considering the increasing global use of INSTIs and decreasing costs associated with genotypic resistance tests, modifications to current ART guidelines to include INSTI resistance testing in INSTI-treated populations may be suitable especially in particular settings with known presence of INSTI resistance [216]. At institutions such as the British Columbia Centre for Excellence in HIV/AIDS (BC-CfE), INSTI resistance testing is becoming increasing incorporated into baseline genotypic resistance tests performed [219].

Certain scientists speculate that the global Joint United Nations Programme on HIV/AIDS (UNAIDS) "90-90-90" effort – which seeks to have 90% of individuals living with HIV aware of their HIV status, 90% of diagnosed HIV infections treated with ART, and 90% of individuals treated with ART have undetectable viral RNA load [220] – is unattainable by the 2020 deadline [221,222]. The UNAIDS 2030 campaign to eliminate the AIDS epidemic however is currently within reach, with the reported number of AIDS-related deaths in 2017 decreasing by 54% since peaking in 2004 [215]. Still, mathematical modeling predictions warn if pretreatment NNRTI resistance levels continue to increase and reach levels above 10%

in sub-Saharan Africa, the 2030 campaign will also be unachievable [78]. Furthermore, though genotypic resistance testing is part of routine clinical care in many countries, it is not widely used in many regions of the world including sub-Saharan Africa, which carry the majority of the HIV burden [84,223]. As a result, drug resistance may be overlooked and potentially impact patient treatment outcome. Although in Chapter Three, pretreatment drug resistance in Uganda was not identified to increase odds of detectable viral RNA load 1-year post-treatment initiation, the observed level of Sanger-detectable pretreatment NNRTI resistance was above 10%, potentially a threat to the elimination of AIDS. Increasing the frequency and affordability of worldwide genotypic resistance testing is feasible with newer sequencing methodologies such as next-generation sequencing, which reduce cost and increase data output [84]. Efforts should be made to expand accessibility of genotypic resistance testing within regions with known and increasing prevalence of drug resistance to impede potential rise of resistance. With advancements in the field of next-generation sequencers, the findings in Chapter Three emphasize potential complications emerging from the use of Sanger-based resistance interpretation algorithms such as Stanford's HIVdb and geno2pheno, to analyze next-generation sequencing minority-level consensus sequencing data. The use of next-generation sequencing for multiplex resistance surveillance of large populations is cost-effective and its use is supported by international researchers in resource-limited settings [55,84]. However, the level of bioinformatic support required may be problematic and misrepresentations of true levels of resistance can occur, if Sanger-detectable viral thresholds are not used during resistance interpretation by current, standard resistance interpretation algorithms.

The vast variety of applications of sequencing technologies in the field of virology should be perceived with limitations and is dependent on characteristics unique to each virus. Emerging viruses, such as ZIKV, are unpredictable in their transmission, geographical distribution, and pathogenicity [224,225]. While genetic sequencing of HIV can reveal drug resistance-associated mutations and provide a means of quality assurance within laboratory settings, similar concepts cannot be applied to ZIKV due to its lack of treatment, low genomic variability, and vast geographical spread. Phylogeographical analysis of ZIKV in

Chapter Four revealed that viruses such as ZIKV present new obstacles to researchers who develop and depend on automated, standardized methodologies of viral diagnostics and surveillance. In addition, ZIKV's cross-reactivity with other similar viruses, its asymptomatic presence, and its low sequence divergence signify a necessity for a broad-spectrum curated viral surveillance system that can monitor ZIKV cases and as a by-product, temporal changes in viral pathogenicity and infectivity, recognizing overlapping and increasing viral geographical spread. Future directions of viral research should investigate the utility of single-molecule sequencing platforms in examining the transmission and evolution of viruses in real-time during viral outbreaks and within resource-limited settings, and additionally, assess the tools required to interpret vast amount of data generated in the process.

Bibliography

1. Goodwin S, McPherson JD, McCombie WR. Coming of age: Ten years of next-generation sequencing technologies. Nat Rev Genet. 2016;17(6):333–51.
2. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409:860–921.
3. National Human Genome Research Institute. A Brief History of the Human Genome Project - National Human Genome Research Institute (NHGRI) [Internet]. 2012 [cited 2019 Mar 8]. Available from: https://www.genome.gov/12011239/a-brief-history-of-the-human-genome-project/
4. Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. Genomics. 2016;107(1):1–8.
5. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977 Dec;74(12):5463–7.
6. Nelson DL, Cox MM. Nucleotides and Nucleic Acids. In: Lenhninger Principles of Biochemistry. 6th ed. New York: W. H. Freeman and Company; 2013. p. 281–313.
7. Chidgeavadze ZG, Beabealashvilli RS, Atrazhev AM, Kukhanova MK, Azhayev A V, Krayevsky AA. 2',3'-Dideoxy-3' aminonucleoside 5'-triphosphates are the terminators of DNA synthesis catalyzed by DNA polymerases. Nucleic Acids Res. 1984 Feb 10;12(3):1671–86.
8. Verma M, Kulshrestha S, Puri A. Genome Sequencing. In: Bioinformatics Volume 1: Data, Sequencing Analysis, and Evolution. Humana Press, New York, NY; 2017. p. 3–33.
9. Franc LTC, Carrilho E, Kist TBL. A review of DNA sequencing techniques. Q Rev Biophys. 2002;35(2):169–200.
10. Chen DY, Swerdlow HP, Harke HR, Zhang JZ, Dovichi NJ. Low-cost, high-sensitivity laser-induced fluorescence detection for DNA sequencing by capillary gel electrophoresis. J Chromatogr. 1991 Oct 18;559(1–2):237–46.

11. von Bubnoff A. Next-Generation Sequencing: The Race Is On. Cell. 2008;132(5):721–3.

12. Rusk N. Torrents of sequence. Nat Methods. 2011;8(1):44.

13. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437(7057):376–80.

14. Pettersson E, Lundeberg J, Ahmadian A. Generations of sequencing technologies. Genomics. 2009;93(2):105–11.

15. QIAGEN. Pyrosequencing Technology and Platform Overview [Internet]. [cited 2019 Mar 12]. Available from: https://www.qiagen.com/ca/service-and-support/learning-hub/technologies-and-research-topics/pyrosequencing-resource-center/technology-overview/

16. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011;475:348–52.

17. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008;456(7218):53–9.

18. Illumina Inc. An introduction to Next-Generation Sequencing Technology [Internet]. 2017 [cited 2019 Mar 12]. Available from: https://www.illumina.com/documents/products/illumina_sequencing_introduction.pdf

19. Guo J, Xu N, Li Z, Zhang S, Wu J, Kim DH, et al. Four-color DNA sequencing with 3'-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. Proc Natl Acad Sci. 2008;105(27):9145–50.

20. Vierstraete A. Next Generation Sequencing [Internet]. 2018 [cited 2019 Mar 14]. Available from: https://users.ugent.be/~avierstr/index.html

21. Buermans HPJ, den Dunnen JT. Next generation sequencing technology: Advances and applications. Biochim Biophys Acta - Mol Basis Dis. 2014;1842(10):1932–41.

22. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. Science (80-). 2009;323(5910):133–8.

23. Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H. Continuous base identification for single-molecule nanopore DNA sequencing. Nat Nanotechnol. 2009;4(4):265–70.

24. Lu H, Giordano F, Ning Z. Oxford Nanopore MinION Sequencing and Genome Assembly. Vol. 14, Genomics, Proteomics and Bioinformatics. 2016. p. 265–79.

25. Schlaberg R, Simmon KE, Fisher MA. A systematic approach for discovering novel, clinically relevant bacteria. Emerg Infect Dis. 2012;18(3):422–30.

26. Wang L, Fu S, Cao Y, Zhang H, Feng Y, Yang W, et al. Discovery and genetic analysis of novel coronaviruses in least horseshoe bats in southwestern China. Emerg Microbes Infect. 2017;6(3:e14).

27. Schaer J, Reeder DAM, Vodzak ME, Olival KJ, Weber N, Mayer F, et al. Nycteria parasites of Afrotropical insectivorous bats. Int J Parasitol. 2015;45(6):375–84.

28. Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. N Engl J Med. 2018;379(15):1403–14.

29. Tostevin A, White E, Dunn D, Croxford S, Delpech V, Williams I, et al. Recent trends and patterns in HIV-1 transmitted drug resistance in the United Kingdom. HIV Med. 2017;18(3):204–13.

30. McDermott PF, Tyson GH, Kabera C, Chen Y, Li C, Folster JP, et al. Whole-genome sequencing for detecting antimicrobial resistance in nontyphoidal Salmonella. Antimicrob Agents Chemother. 2016;60(9):5515–20.

31. Chartier-Harlin MC, Crawford F, Houlden H, Warren A, Hughes D, Fidani L, et al. Early-onset Alzheimer's disease caused by mutations at codon 717 of the β-amyloid precursor protein gene. Nature. 1991;352:844–6.

32. Ma M, Ru Y, Chuang LS, Hsu NY, Shi LS, Hakenberg J, et al. Disease-associated variants in different categories of disease located in distinct regulatory elements. BMC Genomics. 2015;16(Suppl 8: S3).

33. Pollard MO, Porter T, Gurdasani D, Sandhu MS, Mentzer AJ. Long reads: their purpose and place.

Hum Mol Genet. 2018;27(R2):R234–41.

34. Gwinn M, MacCannell D, Armstrong GL. Next-Generation Sequencing of Infectious PathogensNext-Generation Sequencing of Infectious PathogensNext-Generation Sequencing of Infectious Pathogens. JAMA. 2019 Mar 5;321(9):893–4.

35. Prasso JE, Deng JC. Postviral Complications: Bacterial Pneumonia. Clin Chest Med. 2016/12/13. 2017 Mar;38(1):127–38.

36. Granerod J, Ambrose HE, Davies NWS, Clewley JP, Walsh AL, Morgan D, et al. Causes of encephalitis and differences in their clinical presentations in England: A multicentre, population-based prospective study. Lancet Infect Dis. 2010;10(12):835–44.

37. Gu W, Miller S, Chiu CY. Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection. Annu Rev Pathol Mech Dis. 2019 Jan 24;14(1):319–38.

38. Souf S. Recent advances in diagnostic testing for viral infections. Biosci Horizons. 2016;9:hzw010.

39. Bleeker-Rovers CP, Vos FJ, De Kleijn EMHA, Mudde AH, Dofferhoff TSM, Richter C, et al. A prospective multicenter study on fever of unknown origin: The yield of a structured diagnostic protocol. Medicine (Baltimore). 2007;86(1):26–38.

40. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature. 2016;530:228–32.

41. Briese T, Kapoor A, Mishra N, Jain K, Kumar A, Jabado OJ, et al. Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. MBio. 2015;6(5):e01491-15.

42. Gardy J, Loman NJ, Rambaut A. Real-time digital pathogen surveillance - the time is now. Genome Biol. 2015;16(1):155.

43. Baillie GJ, Galiano M, Agapow P-M, Myers R, Chiam R, Gall A, et al. Evolutionary Dynamics of Local Pandemic H1N1/2009 Influenza Virus Lineages Revealed by Whole-Genome Analysis. J Virol. 2012 Jan 1;86(1):11–8.

44. Gastanaduy PA, Redd SB, Clemmons NS, Lee AD, Hickman CJ, Rota PA, et al. Manual for the

Surveillance of Vaccine-Preventable Diseases-Chapter 7: Measles [Internet]. 2018 [cited 2019 Mar 12]. Available from: https://www.cdc.gov/vaccines/pubs/surv-manual/chpt07-measles.html

45. Puddicombe D, Sahni V, Naus M, Gardy JL, Henry B, Amlani A, et al. Whole-Genome Sequencing of Measles Virus Genotypes H1 and D8 During Outbreaks of Infection Following the 2010 Olympic Winter Games Reveals Viral Transmission Routes. J Infect Dis. 2015 Jul 6;212(10):1574–8.

46. Grabowski MK, Redd AD. Molecular tools for studying HIV transmission in sexual networks. Curr Opin HIV AIDS. 2014 Mar;9(2):126–33.

47. Boeras DI, Hraber PT, Hurlston M, Evans-Strickfaden T, Bhattacharya T, Giorgi EE, et al. Role of donor genital tract HIV-1 diversity in the transmission bottleneck. Proc Natl Acad Sci. 2011;108:E1156-63.

48. Beerenwinkel N, Günthard HF, Roth V, Metzner KJ. Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data. Front Microbiol. 2012;3:329.

49. Moya A, Holmes EC, González-Candelas F. The population genetics and evolutionary epidemiology of RNA viruses. Nat Rev Microbiol. 2004;2(4):279–88.

50. Panel on Antiretroviral Guidelines for Adults and Adolescents DHHS. Guidelines for the Use of Antiretroviral Agents in Adults and Adolescents Living with HIV [Internet]. October. 2017 [cited 2018 Nov 12]. Available from: https://aidsinfo.nih.gov/contentfiles/lvguidelines/adultandadolescentgl.pdf

51. Lowe CF, Merrick L, Harrigan PR, Mazzulli T, Sherlock CH, Ritchie G. Implementation of next-generation sequencing for hepatitis B virus resistance testing and genotyping in a clinical microbiology laboratory. J Clin Microbiol. 2016;54(1):127–33.

52. Fourati S, Pawlotsky JM. Virologic tools for HCV drug resistance testing. Viruses. 2015;7(12):6346–59.

53. Liu TF, Shafer RW. Web Resources for HIV Type 1 Genotypic-Resistance Test Interpretation. Clin Infect Dis. 2006 Jun 1;42(11):1608–18.

54. British Columbia Centre for Excellence in HIV/AIDS Pharmacovigilance Initiative. Drug resistance emerging during therapy with dolutegravir and other integrase inhibitors [Internet]. 2015 [cited 2019 Mar 27]. Available from: http://cfenet.ubc.ca/sites/default/files/uploads/publications/centredocs/bccfe_safetyalert_no5_2015_rev2018.pdf

55. Günthard HF, Calvez V, Paredes R, Pillay D, Shafer RW, Wensing AM, et al. Human Immunodeficiency Virus Drug Resistance: 2018 Recommendations of the International Antiviral Society-USA Panel. Clin Infect Dis. 2019;68(2):177–87.

56. You J, Wang H, Huang X, Qin Z, Deng Z, Luo J, et al. Therapy-emergent drug resistance to integrase strand transfer inhibitors in HIV-1 patients: A subgroup meta-analysis of clinical trials. PLoS One. 2016;11(8):e0160087.

57. Lepik KJ, Harrigan PR, Yip B, Wang L, Robbins MA, Zhang WW, et al. Emergent drug resistance with integrase strand transfer inhibitor-based regimens. AIDS. 2017 Jun 19;31(10):1425–34.

58. Nekrutenko A, Taylor J. Next-generation sequencing data interpretation: Enhancing reproducibility and accessibility. Nat Rev Genet. 2012;13(9):667–72.

59. Woods CK, Brumme CJ, Liu TF, Chui CKS, Chu AL, Wynhoven B, et al. Automating HIV drug resistance genotyping with RECall, a freely accessible sequence analysis tool. J Clin Microbiol. 2012 Jun 1;50(6):1936–42.

60. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589–95.

61. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2079–9.

62. Beerenwinkel N, Däumer M, Oette M, Korn K, Hoffmann D, Kaiser R, et al. Geno2pheno: Estimating phenotypic drug resistance from HIV-1 genotypes. Nucleic Acids Res. 2003;31(13):3850–5.

63. Stanford University HIV Drug Resistance Database. HIVdb Program: Mutation List Analysis -

HIV Drug Resistance Database [Internet]. 2018 [cited 2018 May 17]. Available from: https://hivdb.stanford.edu/hivdb/by-mutations/

64. Poon AFY, Kirkby D, Martin E, Richard H L. MiCall [Internet]. [cited 2019 Jul 13]. Available from: https://github.com/cfe-lab/MiCall

65. National Microbiology Laboratory of Canada. HyDRA [Internet]. [cited 2019 Jul 13]. Available from: https://hydra.canada.ca/pages/about?lang=en-CA

66. Ji H, Enns E, Brumme CJ, Parkin N, Howison M, Lee ER, et al. Bioinformatic data processing pipelines in support of next-generation sequencing-based HIV drug resistance testing: the Winnipeg Consensus. J Int AIDS Soc. 2018;21(10):e25193.

67. Colijn C, Gardy J. Phylogenetic tree shapes resolve disease transmission patterns. Evol Med Public Heal. 2014;2014(1):96–108.

68. Thézé J, Li T, du Plessis L, Bouquet J, Kraemer MUG, Somasekar S, et al. Genomic Epidemiology Reconstructs the Introduction and Spread of Zika Virus in Central America and Mexico. Cell Host Microbe. 2018;23:855–64.

69. Lam TT-Y, Hon C-C, Tang JW. Use of phylogenetics in the molecular epidemiology and evolutionary studies of viral infections. Crit Rev Clin Lab Sci. 2010 Feb 1;47(1):5–49, 11–3.

70. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7.

71. Larkin MA, Blackshields G, Brown NP, Chenna R, Mcgettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23(21):2947–8.

72. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3.

73. Kumar S, Stecher G, Peterson D, Tamura K. MEGA-CC: Computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. Bioinformatics. 2012;28(20):2685–6.

74. Gallo RC, Montagnier L. The Discovery of HIV as the Cause of AIDS. N Engl J Med. 2003 Dec

11;349(24):2283–5.

75. U.S. Department of Health & Human Services. A Timeline of HIV and AIDS [Internet]. [cited 2019 Jul 9]. p. 1–17. Available from: https://www.hiv.gov/hiv-basics/overview/history/hiv-and-aids-timeline

76. Crum NF, Riffenburgh RH, Wegner S, Agan BK, Tasker SA, Spooner KM, et al. Comparisons of causes of death and mortality rates among HIV-infected persons: Analysis of the pre-, early, and late HAART (Highly Active Antiretroviral Therapy) eras. J Acquir Immune Defic Syndr. 2006;41(2):194–200.

77. Wainberg MA, Friedland G. Public health implications of antiretroviral therapy and HIV drug resistance. J Am Med Assoc. 1998;279(24):1977–83.

78. World Health Organization. HIV drug resistance report 2017 [Internet]. Geneva; 2017 [cited 2018 May 12]. Available from: http://www.who.int/hiv/pub/drugresistance/hivdr-report-2017/en/

79. Bebenek K, Wilson SH, Kunkelg T a. Error-prone Polymerization by HIV- 1 Reverse Transcriptase. J Biol Chem. 1993;268(14):10324–34.

80. Smith EC. The not-so-infinite malleability of RNA viruses: Viral and cellular determinants of RNA virus mutation rates. PLOS Pathog. 2017;13(4):e1006254.

81. Coffin J, Swanstrom R. HIV pathogenesis: dynamics and genetics of viral populations and infected cells. Cold Spring Harb Perspect Med. 2013;3(1):a012526.

82. Tang MW, Shafer RW. HIV-1 Antiretroviral Resistance: Scientific Principles and Clinical Applications. Drugs. 2012;72(9):e1–25.

83. Hernán MA. The effect of combined antiretroviral therapy on the overall mortality of HIV-infected individuals. AIDS. 2010;24(1):123–37.

84. Inzaule SC, Hamers RL, Paredes R, Yang C, Schuurman R, de Wit TFR. The evolving landscape of HIV drug resistance diagnostics for expanding testing in resource-limited settings. AIDS Rev. 2017;19(4):219–30.

85. World Health Organization. Hepatitis C [Internet]. 2018 [cited 2019 Mar 28]. Available from:

https://www.who.int/news-room/fact-sheets/detail/hepatitis-c

86. Chen SL, Morgan TR. The natural history of hepatitis C virus (HCV) infection. Int J Med Sci. 2006;3(2):47–52.

87. Kumar A, Rajput MK, Paliwal D, Yadav A, Chhabra R, Singh S. Genotyping & diagnostic methods for hepatitis C virus: A need of low-resource countries. Indian J Med Res. 2018;147(5):445–55.

88. World Health Organization. GLOBAL HEPATITIS REPORT, 2017 [Internet]. 2017 [cited 2019 Mar 28]. p. 51–5. Available from: https://www.who.int/hepatitis/publications/global-hepatitis-report2017/en/

89. Wyles DL, Luetkemeyer AF. Understanding Hepatitis C Virus Drug Resistance: Clinical Implications for Current and Future Regimens. Top Antivir Med. 2017;25(3):103–9.

90. Pierson TC, Diamond MS. The emergence of Zika virus and its new clinical syndromes. Nature. 2018;560(7720):573–81.

91. Musso D, Gubler DJ. Zika Virus. Clin Microbiol Rev. 2016 Jul 30;29(3):487–524.

92. Liu Y, Liu J, Du S, Shan C, Nie K, Zhang R, et al. Evolutionary enhancement of Zika virus infectivity in Aedes aegypti mosquitoes. Nature. 2017;545(7655):482–6.

93. Singh RK, Dhama K, Karthik K, Tiwari R, Khandia R, Munjal A, et al. Advances in diagnosis, surveillance, and monitoring of zika virus: An update. Front Microbiol. 2018;8:2677.

94. Brachman PS. Infectious diseases—past, present, and future. Int J Epidemiol. 2003 Oct 1;32(5):684–6.

95. Gulis G, Fujino Y. Epidemiology, population health, and health impact assessment. J Epidemiol. 2015 Mar 5;25(3):179–80.

96. Simonin Y, Loustalot F, Desmetz C, Foulongne V, Constant O, Fournier-Wirth C, et al. Zika Virus Strains Potentially Display Different Infectious Profiles in Human Neural Cells. EBioMedicine. 2016 Oct;12:161–9.

97. Houlihan CF, Larke NL, Watson-Jones D, Smith-McCune KK, Shiboski S, Gravitt PE, et al.

Human papillomavirus infection and increased risk of HIV acquisition. A systematic review and meta-analysis. AIDS. 2012 Nov 13;26(17):2211–22.

98. Hotez PJ. Neglected Tropical Diseases in the Anthropocene: The Cases of Zika, Ebola, and Other Infections. PLoS Negl Trop Dis. 2016;10(4):e0004648.

99. Centers for Disease Control and Prevention. Years of Ebola Virus Disease Outbreaks [Internet]. 2018 [cited 2019 Mar 29]. Available from: https://www.cdc.gov/vhf/ebola/history/chronology.html

100. Fallah MP, Skrip LA, Gertler S, Yamin D, Galvani AP. Quantifying Poverty as a Driver of Ebola Transmission. PLoS Negl Trop Dis. 2015;9(12):e0004260.

101. World Health Organization. Statement on the October 2018 meeting of the IHR Emergency Committee on the Ebola virus disease outbreak in the Democratic Republic of the Congo. 2018.

102. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. Nature. 2008;451(7181):990–3.

103. Ali S, Gugliemini O, Harber S, Harrison A, Houle L, Ivory J, et al. Environmental and Social Change Drive the Explosive Emergence of Zika Virus in the Americas. PLoS Negl Trop Dis. 2017;11(2):e0005135.

104. Lennox JL, DeJesus E, Lazzarin A, Pollard RB, Madruga JVR, Berger DS, et al. Safety and efficacy of raltegravir-based versus efavirenz-based combination therapy in treatment-naive patients with HIV-1 infection: a multicentre, double-blind randomised controlled trial. Lancet. 2009;374(9692):796–806.

105. Sax PE, DeJesus E, Mills A, Zolopa A, Cohen C, Wohl D, et al. Co-formulated elvitegravir, cobicistat, emtricitabine, and tenofovir versus co-formulated efavirenz, emtricitabine, and tenofovir for initial treatment of HIV-1 infection: A randomised, double-blind, phase 3 trial, analysis of results after 48 weeks. Lancet. 2012;379(9835):2439–48.

106. World Health Organization. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: Recommendations for a public health approach [Internet]. 2016

[cited 2018 May 12]. Available from: http://www.who.int/hiv/pub/arv/arv-2016/en/

107. European AIDS Clinical Society. European AIDS Clinical Society Guidelines v9.0 [Internet]. 2017 [cited 2018 Nov 15]. p. 12–3. Available from: http://www.eacsociety.org/files/2018_guidelines-9.1-english.pdf

108. Saag MS, Benson CA, Gandhi RT, Hoy JF, Landovitz RJ, Mugavero MJ, et al. Antiretroviral Drugs for Treatment and Prevention of HIV Infection in Adults: 2018 Recommendations of the International Antiviral Society–USA Panel Antiretroviral Drugs for Treatment and Prevention of HIV Infection in Adults Antiretroviral Drugs for Treat. JAMA. 2018 Jul 24;320(4):379–96.

109. Quashie PK, Mesplède T, Wainberg MA. Evolution of HIV integrase resistance mutations. Curr Opin Infect Dis. 2013;26(1):43–9.

110. Anstett K, Brenner B, Mesplede T, Wainberg MA. HIV drug resistance against strand transfer integrase inhibitors. Retrovirology. 2017 Dec 5;14(1):36.

111. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, et al. Prevention of HIV-1 infection with early antiretroviral therapy. N Engl J Med. 2011 Aug 11;365(6):493–505.

112. Barbour JD, Hecht FM, Wrin T, Liegler TJ, Ramstead CA, Busch MP, et al. Persistence of primary drug resistance among recently HIV-1 infected adults. AIDS. 2004 Aug 20;18(12):1683–9.

113. Primary Care Guidelines Panel of the BC Centre for Excellence in HIV/AIDS. Primary Care Guidelines for the Management of HIV/AIDS in British Columbia On behalf of the Primary Care Guidelines Panel BC Centre for Excellence in HIV/AIDS [Internet]. 2015 [cited 2018 Sep 1]. Available from: http://www.cfenet.ubc.ca/sites/default/files/uploads/primary-care-guidelines/primary-care-guidelines_015-09-15.pdf

114. Committee for Drug Evaluation and Therapy of the British columbia Centre for Excellence in HIV/AIDS. Therapeutic Guidelines for Antiretroviral (ARV) Treatment of Adult HIV Infection [Internet]. 2015 [cited 2018 May 15]. Available from:

http://www.cfenet.ubc.ca/sites/default/files/uploads/Guidelines/bccfe-art-guidelines-Oct_14_2015.pdf

115. Ji H, Patterson A, Taylor T, Rank C, Halverson J, Capina R, et al. Prevalence of Primary Drug Resistance Against HIV-1 Integrase Inhibitors in Canada. J Acquir Immune Defic Syndr. 2018 May 1;78(1):e1–3.

116. Zoufaly A, Kraft C, Schmidbauer C, Puchhammer-Stoeckl E. Prevalence of integrase inhibitor resistance mutations in Austrian patients recently diagnosed with HIV from 2008 to 2013. Infection. 2017;45(2):165–70.

117. Fulcher JA, Du Y, Zhang T, Sun R, Landovitz RJ. Emergence of Integrase Resistance Mutations During Initial Therapy Containing Dolutegravir. Clin Infect Dis. 2018;67(5):791–4.

118. Menza TW, Billock R, Samoff E, Eron JJ, Dennis AM. Pretreatment integrase strand transfer inhibitor resistance in North Carolina from 2010-2016. AIDS. 2017;31(16):2235–44.

119. BC Centre for Excellence in HIV/AIDS. Drug Treatment Program [Internet]. BC Centre for Excellence in HIV/AIDS. 2018 [cited 2018 May 12]. Available from: http://www.cfenet.ubc.ca/drug-treatment-program

120. Patterson S, Cescon A, Samji H, Cui Z, Yip B, Lepik KJ, et al. Cohort Profile: HAART Observational Medical Evaluation and Research (HOMER) Cohort. Int J Epidemiol. 2015 Feb;44(1):58–67.

121. Gonzalez-Serna A, Min JE, Woods C, Chan D, Lima VD, Montaner JSG, et al. Performance of HIV-1 drug resistance testing at low-level viremia and its ability to predict future virologic outcomes and viral evolution in treatment-naive individuals. Clin Infect Dis. 2014;58(8):1165–73.

122. Shahid A, Zhang WW, Montoya V, Cheung PK, Oliveira N, Sidhu MS, et al. Systematic determination of in vitro phenotypic resistance to HIV-1 integrase strand transfer inhibitors from clinical samples. bioRxiv. 2019 Jan 1;621755.

123. Wood S. Generalized Additive Models: An Introduction with R. Chapman & Hall/CRC; 2006.

124. Clark M. Generalized Additive Models: Getting Started with Additive Models in R. Center for

Social Research, University of Notre Dame. 2013.

125. Hurt CB, Sebastian J, Hicks CB, Eron JJ. Resistance to HIV integrase strand transfer inhibitors among clinical specimens in the united states, 2009-2012. Clin Infect Dis. 2014;58(3):423–31.

126. Davy-Mendez T, Eron JJ, Brunet L, Zakharova O, Dennis AM, Napravnik S. New antiretroviral agent use affects prevalence of HIV drug resistance in clinical care populations. AIDS. 2018;32(17):2593–603.

127. Teeranaipong P, Sirivichayakul S, Mekprasan S, Ohata PJ, Avihingsanon A, Ruxrungtham K, et al. Role of Rilpivirine and Etravirine in Efavirenz and Nevirapine-Based Regimens Failure in a Resource-Limited Country: A Cross- Sectional Study. PLoS One. 2016;11(4):e0154221.

128. Cohen CJ, Molina J-M, Cassetti I, Chetchotisakd P, Lazzarin A, Orkin C, et al. Week 96 efficacy and safety of rilpivirine in treatment-naive, HIV-1 patients in two Phase III randomized trials. AIDS. 2013;27(November 2012):939–50.

129. Legorreta A, Yu A, Chernicoff H, Gilmore A, Jordan J, Rosenzweig JC. Adherence to combined Lamivudine+Zidovudine versus individual components: A community-based retrospective medicaid claims analysis. AIDS Care - Psychol Socio-Medical Asp AIDS/HIV. 2005;17(8):938–48.

130. Kauf TL, Davis KL, Earnshaw SR, Anne Davis E. Spillover adherence effects of fixed-dose combination HIV therapy. Patient Prefer Adherence. 2012;6:155–64.

131. Scherrer AU, Yang W-L, Kouyos RD, Böni J, Yerly S, Klimkait T, et al. Successful Prevention of Transmission of Integrase Resistance in the Swiss HIV Cohort Study. J Infect Dis. 2016 Aug 1;214(3):399–402.

132. Cahn P, Pozniak AL, Mingrone H, Shuldyakov A, Brites C, Andrade-Villanueva JF, et al. Dolutegravir versus raltegravir in antiretroviral-experienced, integrase-inhibitor-naive adults with HIV: Week 48 results from the randomised, double-blind, non-inferiority SAILING study. Lancet. 2013;382(9893):700–8.

133. Lepik KJ, Yip B, Ulloa A, Wang L, Toy J, Akagi L, et al. Adverse drug reactions to integrase

strand transfer inhibitors. AIDS. 2018 Feb 24;32(7):1.

134. Brenner BG, Wainberg MA. Clinical benefit of dolutegravir in HIV-1 management related to the high genetic barrier to drug resistance. Virus Res. 2017;239:1–9.

135. Mesplède T, Wainberg MA. Resistance against integrase strand transfer inhibitors and relevance to HIV persistence. Viruses. 2015;7(7):3703–18.

136. Fransen S, Gupta S, Danovich R, Hazuda D, Miller M, Witmer M, et al. Loss of Raltegravir Susceptibility by Human Immunodeficiency Virus Type 1 Is Conferred via Multiple Nonoverlapping Genetic Pathways. J Virol. 2009;83(22):11440–6.

137. Hardy I, Brenner B, Quashie P, Thomas R, Petropoulos C, Huang W, et al. Evolution of a novel pathway leading to dolutegravir resistance in a patient harbouring N155H and multiclass drug resistance. J Antimicrob Chemother. 2015;70(2):405–11.

138. Wensing AM, Calvez V, Günthard HF, Johnson VA, Paredes R, Pillay D, et al. 2017 Update of the Drug Resistance Mutations in HIV-1. Top Antivir Med. 2017;24(4):132–3.

139. Hazuda D, Miller M, Nguyen B, Zhao J. Resistance to the HIV-integrase inhibitor raltegravir: analysis of protocol 005 a phase II study in patients with triple-class resistant HIV-1 infection. In: Antivir Ther. Barbados, West Indies; 2007. p. S10.

140. Gatell JM, Katlama C, Grinsztejn B, Eron JJ, Lazzarin A, Vittecoq D, et al. Long-term efficacy and safety of the HIV integrase inhibitor raltegravir in patients with limited treatment options in a Phase II study. J Acquir Immune Defic Syndr. 2010;53(4):456–63.

141. Mesplède T, Quashie PK, Osman N, Han Y, Singhroy DN, Lie Y, et al. Viral fitness cost prevents HIV-1 from evading dolutegravir drug pressure. Retrovirology. 2013;10(1).

142. Blanco JL, Varghese V, Rhee SY, Gatell JM, Shafer RW. HIV-1 integrase inhibitor resistance and its clinical implications. J Infect Dis. 2011;203(9):1204–14.

143. Cozzi-Lepri A, Noguera-Julian M, Di Giallonardo F, Schuurman R, Däumer M, Aitken S, et al. Low-frequency drug-resistant HIV-1 and risk of virological failure to first-line NNRTI-based ART: a multicohort European case–control study using centralized ultrasensitive 454

pyrosequencing. J Antimicrob Chemother. 2015 Mar 1;70(3):930–40.

144. Malet I, Subra F, Charpentier C, Collin G, Descamps D, Calvez V, et al. Mutations located outside the integrase gene can confer resistance to HIV-1 integrase strand transfer inhibitors. MBio. 2017;8(5):e00922-17.

145. Luber AD. Genetic Barriers to Resistance and Impact on Clinical Response. J Int AIDS Soc. 2005;7(69).

146. HIV Market Report: The state of the HIV treatment, testing, and prevention markets in low- and middle-income countries, 2017-2022 [Internet]. Boston; 2018 [cited 2019 Apr 15]. p. 1–28. Available from: https://clintonhealthaccess.org/content/uploads/2018/09/2018-HIV-Market-Report_FINAL.pdf

147. The Republic of Uganda Ministry of Health. Consolidated Guidelines for Prevention and Treatment of HIV in Uganda [Internet]. 2018 [cited 2019 Apr 5]. p. 1–153, 34–41. Available from: https://elearning.idi.co.ug/pluginfile.php/5675/mod_page/content/19/Uganda HIV Guidelines - September 2018.pdf

148. World Health Organization. Statement on DTG [Internet]. Geneva; 2018 [cited 2019 Apr 5]. p. 1–2. Available from: https://www.who.int/medicines/publications/drugalerts/Statement_on_DTG_18May_2018final.pdf

149. Nakkazi E. Changes to dolutegravir policy in several African countries. Lancet. 2018;392(10143):199.

150. World Health Organization. 2018 Progress Report: Global Action Plan on HIV Drug Resistance 2017-2021 [Internet]. Geneva; 2018 [cited 2019 May 10]. Available from: http://apps.who.int/iris/bitstream/handle/10665/273049/WHO-CDS-HIV-18.12-eng.pdf?ua=1

151. Larder BA, Kohli A, Kellam P, Kemp SD, Kronick M, Henfrey RD. Quantitative detection of HIV-1 drug resistance mutations by automated DNA sequencing. Nature. 1993;365:671–3.

152. Palmer S, Kearney M, Maldarelli F, Halvas EK, Bixby CJ, Bazmi H, et al. Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are

missed by standard genotype analysis. J Clin Microbiol. 2005;43(1):406–13.

153. Li JZ, Paredes R, Ribaudo H, Svarovskaia ES, Metzner KJ, Kozal MJ, et al. Minority HIV-1 Drug Resistance Mutations and the Risk of NNRTI-based Antiretroviral Treatment Failure: A Systematic Review and Pooled Analysis. JAMA. 2011;305(13):1327–35.

154. Simen BB, Simons JF, Hullsiek KH, Novak RM, MacArthur RD, Baxter JD, et al. Low-Abundance Drug-Resistant Viral Variants in Chronically HIV-Infected, Antiretroviral Treatment–Naive Patients Significantly Impact Treatment Outcomes. J Infect Dis. 2009;199(5):693–701.

155. Stella-Ascariz N, Arribas JR, Paredes R, Li JZ. The Role of HIV-1 Drug-Resistant Minority Variants in Treatment Failure. J Infect Dis. 2017;216(suppl_9):S847–50.

156. Dudley DM, Chin EN, Bimber BN, Sanabani SS, Tarosso LF, Costa PR, et al. Low-cost ultra-wide genotyping using roche/454 pyrosequencing for surveillance of HIV drug resistance. PLoS One. 2012;7(5):e36494.

157. Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. Comparison of next-generation sequencing systems. J Biomed Biotechnol. 2012;2012.

158. Avila-Ríos S, García-Morales C, Matías-Florentino M, Tapia-Trejo D, Hernández-Álvarez BF, Moreira-López SE, et al. HIV drug resistance in antiretroviral treatment-naïve individuals in the largest public hospital in Nicaragua, 2011-2015. PLoS One. 2016;11(10):e0164156.

159. Fisher RG, Smith DM, Murrell B, Slabbert R, Kirby BM, Edson C, et al. Next generation sequencing improves detection of drug resistance mutations in infants after PMTCT failure. J Clin Virol. 2015;62:48–53.

160. Charpentier C, Lee GQ, Rodriguez C, Visseaux B, Storto A, Fagard C, et al. Highly frequent HIV-1 minority resistant variants at baseline of the ANRS 139 TRIO trial had a limited impact on virological response. J Antimicrob Chemother. 2015;70(7):2090–6.

161. Peuchant O, Thiébaut R, Capdepont S, Lavignolle-Aurillac V, Neau D, Morlat P, et al. Transmission of HIV-1 minority-resistant variants and response to first-line antiretroviral therapy. AIDS. 2008;22(12):1417–23.

162. Metzner KJ, Rauch P, Braun P, Knechten H, Ehret R, Korn K, et al. Prevalence of key resistance mutations K65R, K103N, and M184V as minority HIV-1 variants in chronically HIV-1 infected, treatment-naïve patients. J Clin Virol. 2011;50(2):156–61.

163. Haberer JE, Musinguzi N, Boum Y, Siedner MJ, Mocello AR, Hunt PW, et al. Duration of antiretroviral therapy adherence interruption is associated with risk of virologic rebound as determined by real-time adherence monitoring in rural Uganda. J Acquir Immune Defic Syndr. 2015;70(4):386–92.

164. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.

165. Lapointe HR, Dong W, Lee GQ, Bangsberg DR, Martin JN, Mocello AR, et al. HIV Drug Resistance Testing by High-Multiplex "Wide" Sequencing on the MiSeq Instrument. Antimicrob Agents Chemother. 2015 Nov;59(11):6824–33.

166. Hamers RL, Schuurman R, Sigaloff KCE, Wallis CL, Kityo C, Siwale M, et al. Effect of pretreatment HIV-1 drug resistance on immunological, virological, and drug-resistance outcomes of first-line antiretroviral treatment in sub-Saharan Africa: A multicentre cohort study. Lancet Infect Dis. 2012;12(4):307–17.

167. Gupta RK, Gregson J, Parkin N, Haile-Selassie H, Tanuri A, Andrade Forero L, et al. HIV-1 drug resistance before initiation or re-initiation of first-line antiretroviral therapy in low-income and middle-income countries: a systematic review and meta-regression analysis. Lancet Infect Dis. 2018 Mar 1;18(3):346–55.

168. Hassan AS, Bibby DF, Mwaringa SM, Agutu CA, Ndirangu KK, Sanders EJ, et al. Presence, persistence and effects of pretreatment HIV-1 drug resistance variants detected using next generation sequencing: A Retrospective longitudinal study from rural coastal Kenya. PLoS One. 2019;14(2):e0210559.

169. Eaton EF, Tamhane A, Saag M, Mugavero MJ, Kilgore ML. Cost considerations in the current antiretroviral era. AIDS. 2016;30(14):2115–9.

170. Bacheler LT, Anton ED, Kudish P, Baker D, Bunville J, Krakowski K, et al. Human immunodeficiency virus type 1 mutations selected in patients failing efavirenz combination therapy. Antimicrob Agents Chemother. 2000;44(9):2475–84.

171. Sluis-Cremer N, Jordan MR, Huber K, Wallis CL, Bertagnolio S, Mellors JW, et al. E138A in HIV-1 reverse transcriptase is more common in subtype C than B: Implications for rilpivirine use in resource-limited settings. Antiviral Res. 2014;107:31–4.

172. Sluis-Cremer N. The emerging profile of cross-resistance among the nonnucleoside HIV-1 reverse transcriptase inhibitors. Viruses. 2014;6(8):2960–73.

173. Adams J, Patel N, Mankaryous N, Tadros M, Miller CD. Nonnucleoside reverse transcriptase inhibitor resistance and the role of the second-generation agents. Ann Pharmacother. 2010;44(1):157–65.

174. World Health Organization. R&D Blueprint: List of Blueprint priority diseases. 2018.

175. Faye O, Freire CCM, Iamarino A, Faye O, de Oliveira JVC, Diallo M, et al. Molecular evolution of Zika virus during its emergence in the 20(th) century. PLoS Negl Trop Dis. 2014;8(1):e2636.

176. Musso D, Nilles EJ, Cao-Lormeau VM. Rapid spread of emerging Zika virus in the Pacific area. Clin Microbiol Infect. 2014;20(10):595–6.

177. Ai J-W, Zhang Y, Zhang W. Zika virus outbreak: “a perfect storm”. Emerg Microbes Infect. 2016 Mar 9;5(3):e21.

178. Gong Z, Gao Y, Han G-Z. Zika Virus: Two or Three Lineages? Trends Microbiol. 2016 Jul 1;24(7):521–2.

179. Petersen LR, Jamieson DJ, Powers AM, Honein MA. Zika Virus. N Engl J Med. 2016;374(16):1552–63.

180. Another kind of Zika public health emergency. Lancet. 2017;389(10069):573.

181. Mlakar J, Korva M, Tul N, Popović M, Poljšak-Prijatelj M, Mraz J, et al. Zika Virus Associated with Microcephaly. N Engl J Med. 2016;374(10):951–8.

182. Carod-Artal FJ. Neurological complications of Zika virus infection. Expert Rev Anti Infect Ther.

2018 May;16(5):399–410.

183. Ginier M, Neumayr A, Gunther S, Schmidt-Chanasit J, Blum J. Zika without symptoms in returning travellers: What are the implications? Travel Med Infect Dis. 2016;14(1):16–20.

184. Government of Canada. Zika virus: For health Professionals [Internet]. 2018 [cited 2018 Oct 10]. Available from: https://www.canada.ca/en/public-health/services/diseases/zika-virus/health-professionals.html?id=surveillance

185. Hedskog C, Chodavarapu K, Ku KS, Xu S, Martin R, Miller MD, et al. Genotype- and Subtype-Independent Full-Genome Sequencing Assay for Hepatitis C Virus. J Clin Microbiol. 2015 Jul 1;53(7):2049–59.

186. Arias A, Watson SJ, Asogun D, Tobin EA, Lu J, Phan MVT, et al. Rapid outbreak sequencing of Ebola virus in Sierra Leone identifies transmission chains linked to sporadic cases. Virus Evol. 2016 Jan 22;2(1):vew016.

187. Houlihan CF, Frampton D, Ferns RB, Raffle J, Grant P, Reidy M, et al. Use of Whole-Genome Sequencing in the Investigation of a Nosocomial Influenza Virus Outbreak. J Infect Dis. 2018 Sep 22;218(9):1485–9.

188. Metsky HC, Matranga CB, Wohl S, Schaffner SF, Freije CA, Winnicki SM, et al. Zika virus evolution and spread in the Americas. Nature. 2017;546:411–5.

189. Grubaugh ND, Ladner JT, Kraemer MUG, Dudas G, Tan AL, Gangavarapu K, et al. Genomic epidemiology reveals multiple introductions of Zika virus into the United States. Nature. 2017;546:401–5.

190. the VIDUS project [Internet]. British Columbia Centre on Substance Use. 2016 [cited 2018 Dec 1]. Available from: https://www.bccsu.ca/vidus/

191. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc. 2017 May 24;12(6):1261–76.

192. Kuiken C, Yusim K, Boykin L, Richardson R. The Los Alamos hepatitis C sequence database.

Bioinformatics. 2005 Feb 1;21(3):379–84.

193. The United Nations Statistics Division. Methodology: Standard country or area codes for statistical use (M49)-Geographic Regions [Internet]. 1999. Available from: https://unstats.un.org/unsd/methodology/m49/

194. Trosemeier J-H, Musso D, Blumel J, Theze J, Pybus OG, Baylis SA. Genome Sequence of a Candidate World Health Organization Reference Strain of Zika Virus for Nucleic Acid Testing. Genome Announc. 2016 Sep;4(5):e00917-16.

195. Larsson A. AliView: A fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. 2014;30(22):3276–8.

196. Wickam H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2016.

197. Sukumaran J, Holder MT. DendroPy: a Python library for phylogenetic computing. Bioinformatics. 2010 Apr 25;26(12):1569–71.

198. Miner JJ, Cao B, Govero J, Smith AM, Fernandez E, Cabrera OH, et al. Zika Virus Infection during Pregnancy in Mice Causes Placental Damage and Fetal Demise. Cell. 2016;165(5):1081–91.

199. Brasil P, Pereira JP, Moreira ME, Ribeiro Nogueira RM, Damasceno L, Wakimoto M, et al. Zika Virus Infection in Pregnant Women in Rio de Janeiro. N Engl J Med. 2016 Dec 15;375(24):2321–34.

200. Lanciotti RS, Kosoy OL, Laven JJ, Velez JO, Lambert AJ, Johnson AJ, et al. Genetic and serologic properties of Zika virus associated with an epidemic, Yap State, Micronesia, 2007. Emerg Infect Dis. 2008 Aug;14(8):1232–9.

201. Colombo TE, Terzian ACB, Júnior JPA, Parreira R, Cabrera EMS, Santos INP dos, et al. Zika detection: comparison of methodologies. Brazilian J Microbiol. 2018 Jan 1;49(1):144–7.

202. Gardner LM, Bóta A, Gangavarapu K, Kraemer MUG, Grubaugh ND. Inferring the risk factors behind the geographical spread and transmission of Zika in the Americas. PLoS Negl Trop Dis. 2018;12(1):e0006194.

203. Pacheco O, Beltrán M, Nelson CA, Valencia D, Tolosa N, Farr SL, et al. Zika Virus Disease in Colombia — Preliminary Report. N Engl J Med. 2016;

204. Theys K, Libin P, Dallmeier K, Pineda-Pena A-C, Vandamme A-M, Cuypers L, et al. Zika genomics urgently need standardized and curated reference sequences. PLoS Pathog. 2017 Sep;13(9):e1006528.

205. European Bioinformatics Institute. International Nucleotide Sequence Database Collaboration: The DDBJ/ENA/GenBank Feature Table Definition. 2017.

206. Echeverría N, Moratorio G, Cristina J, Moreno P. Hepatitis C virus genetic variability and evolution. World J Hepatol. 2015;7(6):831–45.

207. Selisko B, Papageorgiou N, Ferron F, Canard B. Structural and functional basis of the fidelity of nucleotide selection by flavivirus RNA-dependent RNA polymerases. Viruses. 2018;10(2):E59.

208. Lemaire D, Barbosa T, Rihet P. Coping with genetic diversity: The contribution of pathogen and human genomics to modern vaccinology. Brazilian J Med Biol Res. 2012;45(5):376–85.

209. Faria NR, Do Socorro Da Silva Azevedo R, Kraemer MUG, Souza R, Cunha MS, Hill SC, et al. Zika virus in the Americas: Early epidemiological and genetic findings. Science (80-). 2016;352(6283):345–9.

210. Gubler DJ, Vasilakis N, Musso D. History and Emergence of Zika Virus. J Infect Dis. 2017;216(suppl 10):S860–7.

211. Musso D, Roche C, Robin E, Nhan T, Teissier A, Cao-Lormeau VM. Potential sexual transmission of zika virus. Emerg Infect Dis. 2015;21(2):359–61.

212. Deckard DT, Chung WM, Brooks JT, Smith JC, Woldai S, Hennessey M, et al. Male-to-Male Sexual Transmission of Zika Virus — Texas, January 2016. MMWR Morb Mortal Wkly Rep. 2016;65(14):372–4.

213. Barré-Sinoussi F, Ross AL, Delfraissy JF. Past, present and future: 30 years of HIV research. Nat Rev Microbiol. 2013;11:877–83.

214. Wikan N, Smith DR. Zika virus: History of a newly emerging arbovirus. Lancet Infect Dis.

2016;16(7):e119–26.

215. Unaids. Fact sheet - Latest global and regional statistics on the status of the AIDS epidemic. [Internet]. 2017 [cited 2019 Jun 14]. Available from: http://www.unaids.org/sites/default/files/media_asset/UNAIDS_FactSheet_en.pdf

216. Saag MS, Günthard HF, Smith DM. Baseline Genotype Testing to Assess Drug Resistance before Beginning HIV Treatment - Reply. JAMA - J Am Med Assoc. 2018;320(20):2153–4.

217. Casadellà M, van Ham PM, Noguera-Julian M, van Kessel A, Pou C, Hofstra LM, et al. Primary resistance to integrase strand-transfer inhibitors in Europe. J Antimicrob Chemother. 2015;70(10):2885–8.

218. Ambrosioni J, Nicolás D, Manzardo C, Agüero F, Blanco JL, Mosquera MM, et al. Integrase strand-transfer inhibitor polymorphic and accessory resistance substitutions in patients with acute/recent HIV infection. J Antimicrob Chemother. 2017;72(1):205–9.

219. BC Centre for Excellence in HIV/AIDS. Therapeutic Guidelines: Management of Acute HIV Infections [Internet]. 2018. Available from: http://cfenet.ubc.ca/sites/default/files/uploads/Guidelines/Management-of-Acute-HIV-Infections-[16-MAY-2018].pdf

220. UNAIDS. 90-90-90: An ambitious treatment target to help end the AIDS epidemic [Internet]. 2014 [cited 2019 Jun 13]. Available from: https://www.unaids.org/sites/default/files/media_asset/90-90-90_en.pdf

221. Levi J, Raymond A, Pozniak A, Vernazza P, Kohler P, Hill A. Can the UNAIDS 90-90-90 target be achieved? A systematic analysis of national HIV treatment cascades. BMJ Glob Heal. 2016;1:e000010.

222. Kirby T. The UK reaches UNAIDS 90-90-90 targets. Lancet. 2018;392(10163):2427.

223. Fettig J, Swaminathan M, Murrill CS, Kaplan JE. Global epidemiology of HIV. Infect Dis Clin North Am. 2014;28(3):323–37.

224. Drexler M. Zika epidemic forcing scientists to rethink assumptions about human biology

[Internet]. HARVARD T.H. CHAN. Boston; 2016 [cited 2019 Jan 28]. Available from: https://www.hsph.harvard.edu/news/features/zika-epidemic-microcephaly-guillain-barre/

225. Samarasekera U, Triunfol M. Concern over Zika virus grips the world. Lancet. 2016;387(10018):521–4.

Appendices

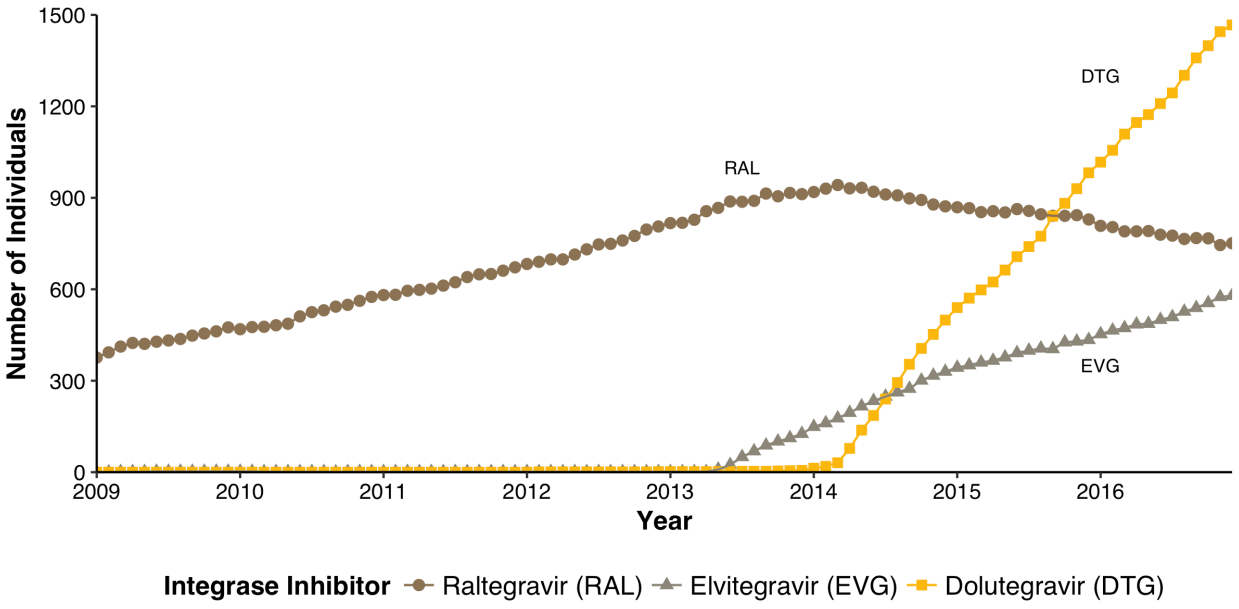
Appendix I. Yearly number of individuals within the BC Drug Treatment Program between 2009 and 2016.

The yearly number of ART-treated individuals within the DTP between 2009 and 2016 is displayed. In each year, an individual contributed to the count of ART-treated if they were dispensed ART, and the subset INSTI-treated if they were dispensed any INSTI (raltegravir, elvitegravir, or dolutegravir) in the calendar year. Individuals found to have study-defined PI-RT or INSTI resistance were counted as PI-RT resistant or INSTI resistant, respectively, and contributed to the count of prevalence of PI-RT or INSTI resistance per 1000 ART-treated individuals for each year.

Year	All ART-treated individuals	All ART-treated individuals				
		INSTI-treated, n (%)	Tested for INSTI resistance, n (%)	INSTI resistant, n (%)	Tested for PI-RT resistance, n (%)	PI-RT resistant, n (%)
2009	5587	542 (10)	188 (3)	6 (0.1)	4520 (81)	1885 (34)
2010	5991	649 (11)	259 (4)	10 (0.2)	4945 (83)	1930 (32)
2011	6450	751 (12)	337 (5)	21 (0.3)	5397 (84)	1997 (31)
2012	6773	916 (14)	417 (6)	29 (0.4)	5739 (85)	2069 (31)
2013	7028	1182 (17)	523 (7)	34 (0.5)	5993 (85)	2092 (30)
2014	7333	1873 (26)	751 (10)	45 (0.6)	6264 (85)	2167 (30)
2015	7595	2510 (33)	1051 (14)	52 (0.7)	6482 (85)	2206 (29)
2016	7772	3117 (40)	1440 (19)	56 (0.7)	6614 (85)	2211 (28)

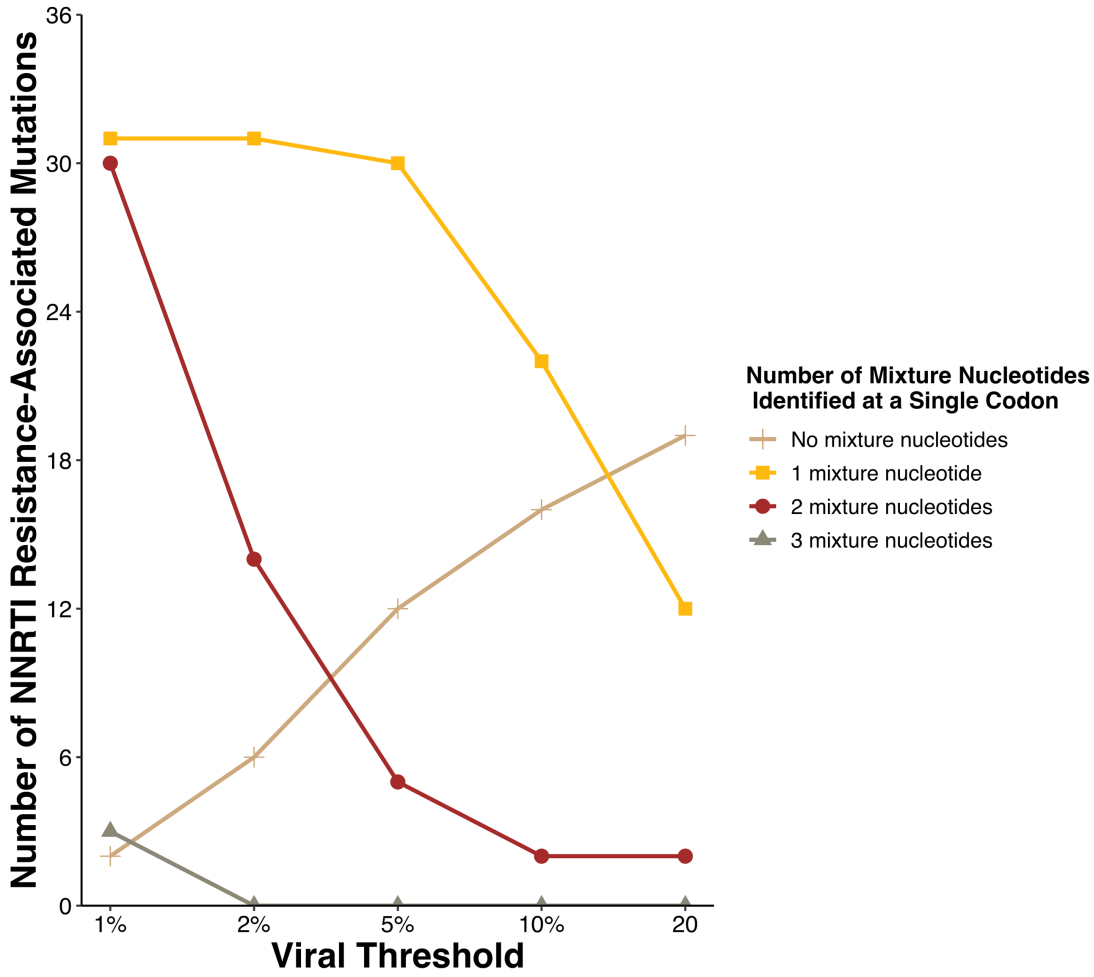
Appendix II. Trends of raltegravir, elvitegravir, and dolutegravir usage.

The figure shows the change in usage of raltegravir (RAL), elvitegravir (EVG), and dolutegravir (DTG) within the Drug Treatment Program between 2009 and 2016.



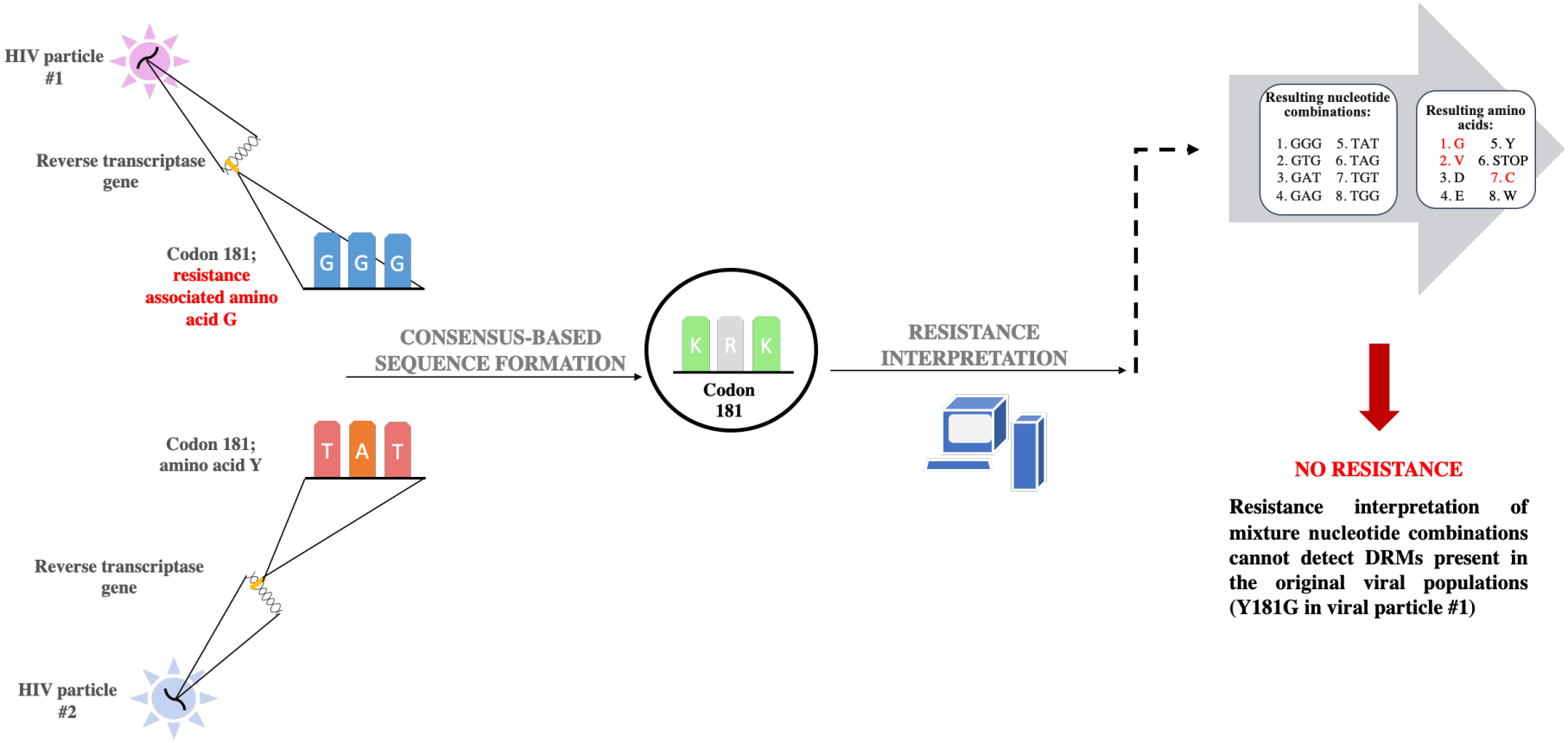
Appendix III. Increasing number of nucleotide mixture combinations at codons associated with NNRTI resistance observed in consensus-based sequences produced at low viral thresholds.

As viral threshold decreased, there were increasing number of mixture nucleotides observed at NNRTI drug resistance-associated codons in the reverse transcriptase gene. While at viral threshold of 20%, the majority of codons associated with NNRTI resistance contained no mixture nucleotides within their tri-nucleotide sequence, at the low viral threshold of 1%, the majority of codons associated with NNRTI resistance contained at least 1 mixture nucleotide within their tri-nucleotide sequence.



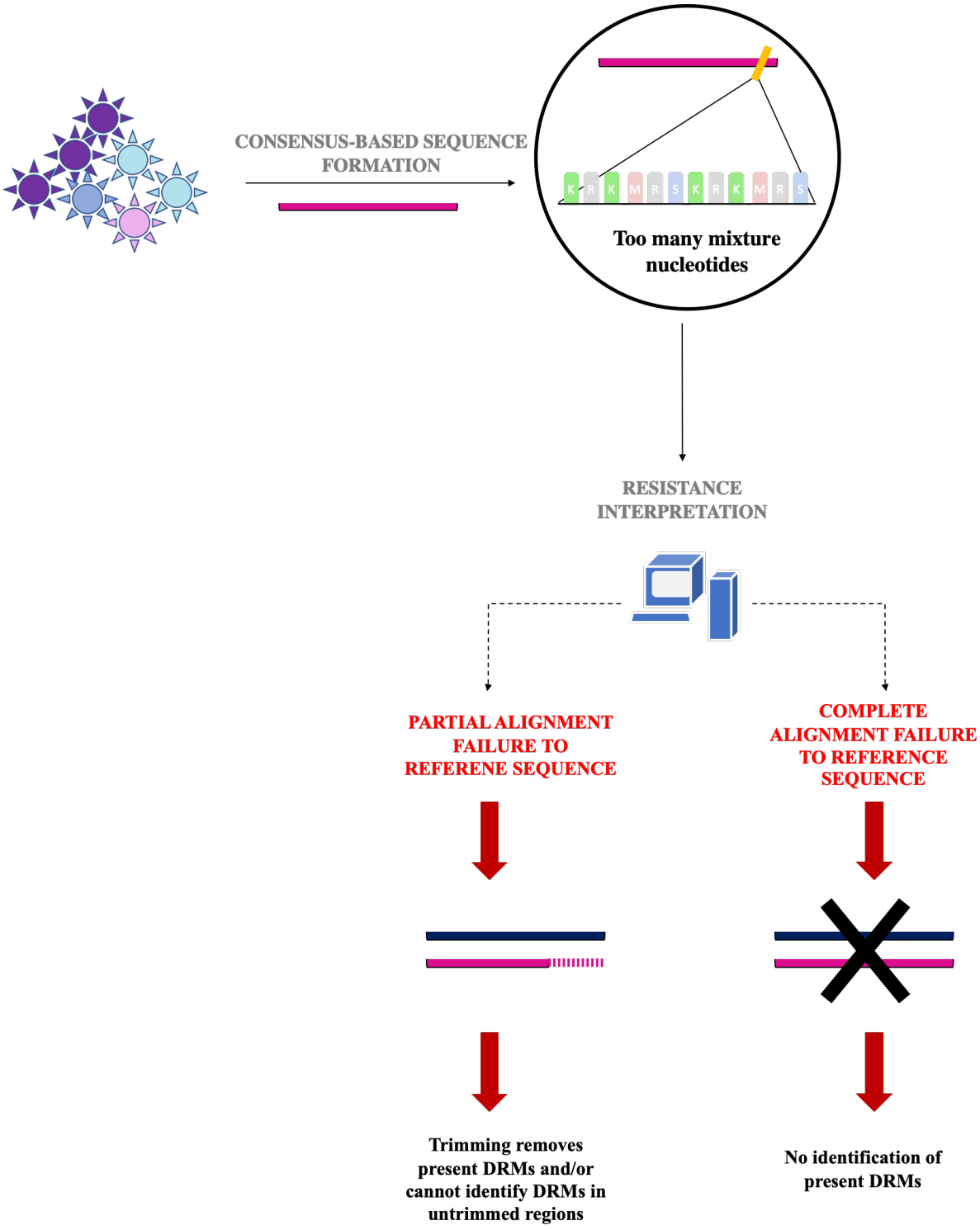
Appendix IV. Potential complications arising from the use of consensus-based sequences with resistance interpretation algorithms.

A) A viral population with both a virus with nucleotides "GGG" and nucleotides "TAT" at codon 181 of the reverse transcriptase gene, will result in mixture nucleotides "KRK". Resistance interpretation of KRK results in different nucleotide combinations and various amino acids, some of which are associated with resistance (red). However, Stanford's HIVdb cannot detect the resistance amino acids, and resistance is therefore not identified at codon 181 in the consensus-based sequence.



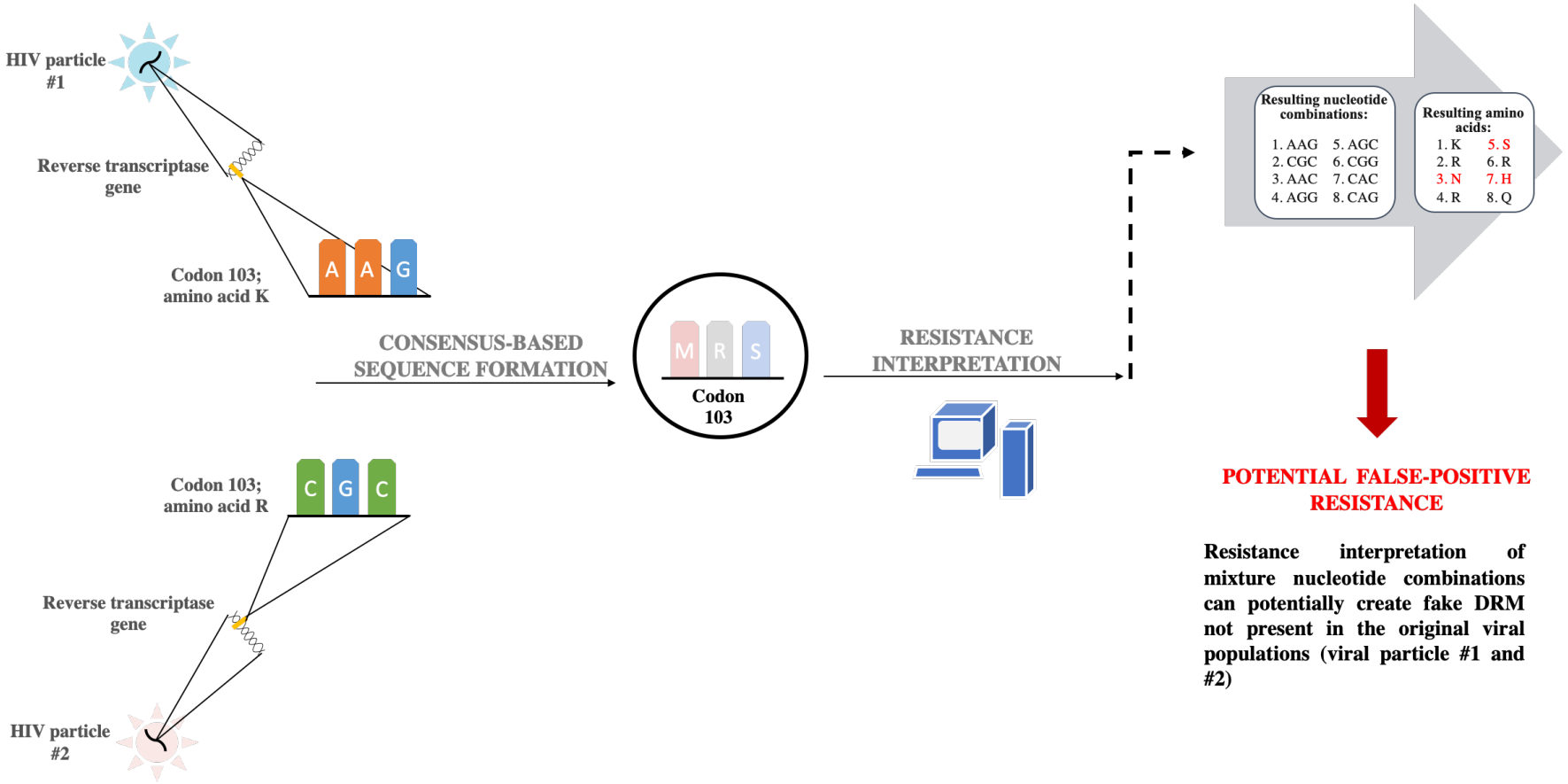
Appendix IV. Potential complications arising from the use of consensus-based sequences with resistance interpretation algorithms. Continued.

B) A consensus-based sequence may contain high numbers of mixture nucleotides which can result in complete or partial alignment failure to reference genes during resistance interpretation, resulting in no identification of present DRMs, or trimming of consensus-based sequences and removing present DRMs in the process.



Appendix IV. Potential complications arising from the use of consensus-based sequences with resistance interpretation algorithms. Continued.

C) A viral population with both a virus with nucleotides "AAG" and nucleotides "CGC" at codon 103 of the reverse transcriptase gene, will result in mixture nucleotides "MRS". Resistance interpretation of MRS results in different nucleotide combinations and various amino acids, some of which are associated with resistance (in red). Inspection of the resulting amino acids from the consensus-based sequence may identify resistance-associated amino acids (in red) that are not in the original viral population, although this complication was not identified in the analysis or results.



Appendix V. Whole-genome ZIKV reference set collected from GenBank.

Accession Number	Country	Date of Isolation
KU963574	NIGERIA	1968
HQ234500	NIGERIA	1968
HQ234501	SENEGAL	1984
KU955595	SENEGAL	1984
KU955592	SENEGAL	1984
KU955591	SENEGAL	1984
KX601166	SENEGAL	1984
KY348860	SENEGAL	1984
KF268948	CENTRAL-AFRICAN-REPUBLIC	1976
MH130104	UGANDA	1947
MH130097	UGANDA	1947
MH130105	UGANDA	1947
MH130102	UGANDA	1947
KU955594	UGANDA	1947
MH130100	UGANDA	1947
MH130094	UGANDA	1947
MH130109	UGANDA	1947
MH130101	UGANDA	1947
MH130095	UGANDA	1947
MH130106	UGANDA	1947
MH130103	UGANDA	1947
MH130098	UGANDA	1947
MH130096	UGANDA	1947
MH130108	UGANDA	1947
MH130099	UGANDA	1947
MH130107	UGANDA	1947
KX377336	MALAYSIA	1966
KX694533	MALAYSIA	1966
HQ234499	MALAYSIA	1966
KX601167	MALAYSIA	1966
KU681082	PHILIPPINES	2012
KY120353	SOUTH-KOREA	2016
KY553111	SOUTH-KOREA	2016
EU545988	MICRONESIA	2007
MG645981	THAILAND	2006
MH368551	CAMBODIA	2016

MH158236	CAMBODIA	2010
KU955593	CAMBODIA	2010
JN860885	CAMBODIA	2010
KY272987	THAILAND	2016
KF993678	CANADA	2013
KX694532	THAILAND	2013
KY328290	CHINA	2016
KX051562	THAILAND	2015
MH119185	THAILAND	2016
KX051561	THAILAND	2013
KU179098	INDONESIA	2014
KX051560	THAILAND	2013
MG807647	THAILAND	2017
KY126351	THAILAND	2016
MH013290	THAILAND	2017
LC369584	THAILAND	2017
MG548661	THAILAND	2016
MG548660	THAILAND	2016
MF996804	THAILAND	2017
MF692778	THAILAND	2016
MG807646	THAILAND	2016
KY241778	SINGAPORE	2016
KY241777	SINGAPORE	2016
KY241779	SINGAPORE	2016
KY241780	SINGAPORE	2016
KY241776	SINGAPORE	2016
MH255601	SINGAPORE	2016
KY241783	SINGAPORE	2016
KY241781	SINGAPORE	2016
KY241782	SINGAPORE	2016
KY241787	SINGAPORE	2016
KY241775	SINGAPORE	2016
KY241786	SINGAPORE	2016
KY241785	SINGAPORE	2016
KY241784	SINGAPORE	2016
KU681081	THAILAND	2014
LC219720	JAPAN	2016
MF801384	HONDURAS	2016
KU744693	VENEZUELA	2016
KY241788	SINGAPORE	2016

MF801414	MEXICO	2016
MF801417	MEXICO	2016
KX827268	USA	2016
MF098771	MEXICO	2017
MH157208	MEXICO	2016
MH157213	MEXICO	2016
KY631494	MEXICO	2015
KY631493	MEXICO	2015
MF801398	MEXICO	2016
MF801413	MEXICO	2016
KX856011	MEXICO	2016
MH157202	MEXICO	2016
KY120349	MEXICO	2016
MF801395	MEXICO	2016
KY120348	MEXICO	2016
KX247632	MEXICO	2015
MF434516	NICARAGUA	2016
MF801397	MEXICO	2016
MF801396	MEXICO	2016
KY785442	HONDURAS	2016
KY765324	NICARAGUA	2016
KY785448	HONDURAS	2016
MF801406	MEXICO	2016
MF434522	NICARAGUA	2016
MF801378	GUATEMALA	2016
MF801403	MEXICO	2016
KY014306	HONDURAS	2016
MF988734	SINGAPORE	2017
MH063262	CUBA	2017
MF159531	CUBA	2017
MF801418	MEXICO	2016
KY693676	HONDURAS	2016
MF434517	NICARAGUA	2016
KU870645	USA	2016
KY014319	HONDURAS	2016
KY765322	NICARAGUA	2016
KY765317	NICARAGUA	2016
KY765321	NICARAGUA	2016
KY765318	NICARAGUA	2016
MF434521	NICARAGUA	2016

KY014312	HONDURAS	2016
KY328289	HONDURAS	2016
MF801410	MEXICO	2016
MF593625	CHINA	2016
KY927808	CHINA	2016
MF801387	HONDURAS	2016
MF801402	MEXICO	2016
MF801412	MEXICO	2016
MF801426	NICARAGUA	2016
KU501217	GUATEMALA	2015
KU501216	GUATEMALA	2015
KY765326	NICARAGUA	2016
KY765323	NICARAGUA	2016
KY765320	NICARAGUA	2016
KX421195	NICARAGUA	2016
KX421194	NICARAGUA	2016
KX906952	HONDURAS	2016
KX694534	HONDURAS	2015
KX262887	HONDURAS	2016
KY765327	NICARAGUA	2016
KY765325	NICARAGUA	2016
KY693677	HONDURAS	2016
KY785418	HONDURAS	2016
KY014315	HONDURAS	2016
KY559021	BRAZIL	2016
MH063261	CUBA	2017
LC191864	JAPAN	2016
KY785435	DOMINICAN-REPUBLIC	2016
KY785415	DOMINICAN-REPUBLIC	2016
MF438286	CUBA	2017
KY785475	DOMINICAN-REPUBLIC	2016
MH063264	CUBA	2017
LC331561	JAPAN	2016
KY785422	USA	2016
KY785476	DOMINICAN-REPUBLIC	2016
KY014314	DOMINICAN-REPUBLIC	2016
MF098769	DOMINICAN-REPUBLIC	2016
MF098768	DOMINICAN-REPUBLIC	2016
MF664436	DOMINICAN-REPUBLIC	2016
KY014299	USA	2016

KX922708	USA	2016
KX922705	USA	2016
KY014322	USA	2016
KX838906	USA	2016
KY014302	DOMINICAN-REPUBLIC	2016
KY785441	DOMINICAN-REPUBLIC	2016
KY785468	USA	2016
KY014324	USA	2016
KX838904	USA	2016
KX832731	USA	2016
KY014325	USA	2016
KX922704	USA	2016
KX922706	USA	2016
KX922703	USA	2016
KY014323	USA	2016
KX838905	USA	2016
KY014295	USA	2016
KX842449	USA	2016
KY014316	USA	2016
KX922707	USA	2016
KX673530	UK	2016
MF384325	HAITI	2016
KY415991	HAITI	2014
KY415989	HAITI	2014
KY415988	HAITI	2014
KY415990	HAITI	2014
KY415987	HAITI	2014
KY415986	HAITI	2014
KY014321	DOMINICAN-REPUBLIC	2016
KY014305	DOMINICAN-REPUBLIC	2016
KY014318	DOMINICAN-REPUBLIC	2016
KY014300	DOMINICAN-REPUBLIC	2016
MF098765	DOMINICAN-REPUBLIC	2016
MF098766	DOMINICAN-REPUBLIC	2016
LC190723	JAPAN	2016
MF098764	DOMINICAN-REPUBLIC	2016
KY785420	DOMINICAN-REPUBLIC	2016
KY014304	DOMINICAN-REPUBLIC	2016
KX269878	ITALY	2016
KU853013	ITALY	2016

KU853012	ITALY	2016
KX520666	BRAZIL	2015
KY003154	ITALY	2016
KY003153	ITALY	2016
KY441401	BRAZIL	2016
MF098767	RUSSIA	2016
KY693680	VENEZUELA	2016
KY558996	BRAZIL	2015
KU926310	BRAZIL	2016
KY785455	BRAZIL	2016
KY559013	BRAZIL	2016
KY559007	BRAZIL	2016
KY559005	BRAZIL	2016
KY120352	BRAZIL	2016
KU729217	BRAZIL	2015
MF073359	BRAZIL	2015
MF073358	BRAZIL	2015
KY631492	BRAZIL	2016
KX056898	CHINA	2016
KY379148	CHINA	2016
MF167360	CHINA	2016
KU955590	CHINA	2016
KX766028	DOMINICAN-REPUBLIC	2016
KU740184	CHINA	2016
KU761564	CHINA	2016
KU820898	CHINA	2016
KY559027	BRAZIL	2016
KU497555	BRAZIL	2015
KY693679	PERU	2016
KY693678	PERU	2016
KY785466	COLOMBIA	2016
KX548902	COLOMBIA	2015
MH544701	COLOMBIA	2016
MF574561	COLOMBIA	2015
MF574555	COLOMBIA	2015
KX893855	VENEZUELA	2016
KX702400	VENEZUELA	2016
MF574575	COLOMBIA	2015
MF574567	COLOMBIA	2015
MF574576	COLOMBIA	2015

MF574560	COLOMBIA	2015
MF574568	COLOMBIA	2015
MF574572	COLOMBIA	2015
MF574559	COLOMBIA	2015
MF574562	COLOMBIA	2015
MF574556	COLOMBIA	2015
MF574558	COLOMBIA	2015
MF574566	COLOMBIA	2015
MF574565	COLOMBIA	2015
MF574571	COLOMBIA	2015
MF574573	COLOMBIA	2015
MF574569	COLOMBIA	2015
MF574563	COLOMBIA	2015
MF574564	COLOMBIA	2015
MF574570	COLOMBIA	2015
MF574557	COLOMBIA	2015
KX247646	COLOMBIA	2016
MF574554	COLOMBIA	2015
KX087102	COLOMBIA	2015
KU820897	COLOMBIA	2015
MF574553	COLOMBIA	2015
MF574583	COLOMBIA	2015
MF574588	COLOMBIA	2015
MF574586	COLOMBIA	2015
MF574584	COLOMBIA	2015
MF574582	COLOMBIA	2015
MF574581	COLOMBIA	2015
MF574580	COLOMBIA	2015
MF574577	COLOMBIA	2015
MF574574	COLOMBIA	2015
MF574552	COLOMBIA	2015
KY317937	COLOMBIA	2016
KU647676	MARTINIQUE	2015
KU922960	MEXICO	2016
KU922923	MEXICO	2016
KY785469	COLOMBIA	2016
KY989971	COLOMBIA	2015
KX156775	PANAMA	2015
KX156776	PANAMA	2015
MF574585	COLOMBIA	2015

MF574587	COLOMBIA	2015
KX156774	PANAMA	2015
KY317940	COLOMBIA	2016
KY317939	COLOMBIA	2016
KY317936	COLOMBIA	2016
KY014303	DOMINICAN-REPUBLIC	2016
KY317938	COLOMBIA	2016
KX198135	PANAMA	2016
NC035889	BRAZIL	2015
KU527068	BRAZIL	2015
MF099651	CHINA	2016
MF964216	CHINA	2016
KU761560	CHINA	2016
KU761561	CHINA	2016
MF036115	CHINA	2016
KY967711	CHINA	2016
MH055376	CHINA	2016
MG674719	CHINA	2016
MG674718	CHINA	2016
KX266255	CHINA	2016
KX253996	CHINA	2016
KU955589	CHINA	2016
KU820899	CHINA	2016
KX185891	CHINA	2016
KU963796	CHINA	2016
KU866423	CHINA	2016
KU997667	CHINA	2016
KX013000	CHINA	2016
KX117076	CHINA	2016
KY014320	BRAZIL	2016
KY014296	BRAZIL	2016
KX879603	ECUADOR	2016
KX879604	ECUADOR	2016
MF794971	ECUADOR	2016
KU926309	BRAZIL	2016
KY272991	BRAZIL	2016
KU312312	SURINAME	2015
KY785464	PUERTO-RICO	2016
MH158237	PUERTO-RICO	2015
KX377337	PUERTO-RICO	2015

MF574578	COLOMBIA	2015
KU501215	PUERTO-RICO	2015
KX601168	PUERTO-RICO	2015
MF574579	COLOMBIA	2015
KX087101	PUERTO-RICO	2015
KY348640	SURINAME	2016
KY785450	BRAZIL	2016
KY014297	BRAZIL	2016
MF073357	BRAZIL	2016
KY441403	BRAZIL	2016
KY441402	BRAZIL	2016
KU365778	BRAZIL	2015
KU758877	FRENCH-GUIANA	2015
KU937936	SURINAME	2016
KY014317	BRAZIL	2016
KY559015	BRAZIL	2016
KX806557	AUSTRALIA	2016
KU707826	BRAZIL	2015
KU365779	BRAZIL	2015
KU365780	BRAZIL	2015
KU365777	BRAZIL	2015
KX280026	BRAZIL	2015
KX811222	BRAZIL	2016
KU729218	BRAZIL	2015
KX830930	BRAZIL	2016
KX197205	BRAZIL	2015
KU991811	ITALY	2016
KX447517	FRENCH-POLYNESIA	2014
KX051563	USA	2016
KU509998	HAITI	2014
KU321639	BRAZIL	2015
KX197192	BRAZIL	2015
KR872956	BRAZIL	2016
MF352141	BRAZIL	2015
KY558999	BRAZIL	2016
KX447510	FRENCH-POLYNESIA	2013
KX447516	FRENCH-POLYNESIA	2014
KX447511	FRENCH-POLYNESIA	2014
KX447515	FRENCH-POLYNESIA	2013
KX447513	FRENCH-POLYNESIA	2013

KX447512	FRENCH-POLYNESIA	2013
MG827392	FRENCH-POLYNESIA	2013
KX369547	FRENCH-POLYNESIA	2013
KX447514	FRENCH-POLYNESIA	2014
KJ776791	FRENCH-POLYNESIA	2013
KX447509	FRENCH-POLYNESIA	2013

Appendix VI. Equation used for Shannon diversity calculation.

The proportion (p_i) of each nucleotide at each position is multiplied by the natural logarithm of the same respective proportion and summed over all four nucleotides.

$$SE = -\sum_{i=0}^{4} p_i \ln p_i$$