

**Genetic Characterization of Canadian Group A  
Human Rotavirus Strains Collected in Multiple  
Paediatric Hospitals from 2007-2010**

**by**

**Andrew McDermid**

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

**MASTER OF SCIENCE**

Department of Medical Microbiology and Infectious Diseases

University of Manitoba

Winnipeg, Manitoba, Canada

Copyright © 2012 by Andrew McDermid

## Abstract

Group A rotaviruses are a major cause of acute gastroenteritis in children. Rotavirus most commonly infects infants from the ages of 3 months to 2 years of age, while almost all children are infected, in varying severities, by the age of 5 years old. Rotavirus disease causes around 600,000 deaths per year, mostly in developing countries. Two recently licensed vaccines have been shown to be safe and efficacious in phase III clinical trials. VP4 (P type) and VP7 (G type) rotavirus genotypes and domains were analyzed for prevalence and potential antigenicity, as they are known to be key rotavirus proteins eliciting a neutralizing antibody response during infection. This study is focused on predicting the effectiveness of rotavirus vaccines in a Canadian population by genetically comparing genotypes found in Canadian rotaviruses to those found in the vaccines.

I hypothesize that surveillance will reveal currently circulating rotavirus strains in Canada are well matched to the currently licensed vaccines in Canada.

Diarrhea samples from 8 paediatric hospitals, suspected of rotavirus infection, were sent to the National Microbiology Laboratory for genetic analysis. In all, 348 samples were collected, while 259 were successfully genotyped and confirmed by heminested multiplex PCR and sequencing, respectively. There were an additional 12 samples with multiple infections. A real-time PCR assay, detecting two VP2 genotypes, was developed as an alternative rotavirus detection and confirmation assay.

Canadian rotavirus genotypes were found to be mostly G1P[8] (approximately 67%), followed by G3P[8] (approximately 17.5%), G2P[4] (approximately 8.5%), G9P[8] (approximately 5%), G4P[8] (approximately 1.2 %) and G9P[4] (approximately 0.8%), between 2007 and 2010. Rotavirus genotype prevalence fluctuated greatly among areas and seasons, but matched the multivalent vaccine well.

A subset of rotavirus-positive samples was chosen for full genome analysis, which allowed for reassortment analysis, motif analysis and genotyping assay quality analysis. There were no unusual reassortment events found in Canadian strains. Variations amongst strains were commonly genotype-specific variations. Unusual amino acid variations were rare in Canada. Antigenic domains varied, often in a genotype-specific manner. In conclusion, rotavirus vaccine escape is presently unlikely amongst Canadian strains.

## Acknowledgements

I would like to thank my supervisor, Tim Booth, for allowing me to complete this research. His guidance and support are thoroughly appreciated and I have been privileged to study as a Master's student under his supervision. In addition, my advisory committee and mentors were essential for my success as a graduate student. Thanks to Joanne Embree, Kevin Coombs, Maryanne Crockett, Raj Bhullar, Daniel Beniac and George Zhanel for their support and advice.

I would also like to thank the Enterovirus section of the National Microbiology Laboratory for the training and support I received during my studies. I would especially like to thank Elsie Grudeski, whose co-operation, training and guidance was important for the success of this research. In addition, I would like to thank Kathy Manguiat, Lily MacDonald and Susan Christianson for their patience during my initial training and orientation in the laboratory. I would also like to thank all the members of the Bloodborne Pathogens and Hepatitis laboratory at the National Microbiology Laboratory for their support and friendship over the course of my laboratory work.

I would like to thank John Gentsch, Daniel Payne and Michael Bowen from the CDC for their generous donation of positive controls.

Finally, I would like to thank my friends, family and loved ones for their constant patience and support throughout my graduate program.

## Table of Contents

Abstract	i
Acknowledgements	ii
Table of Contents	iii
List of Tables	v
List of Figures	vi
List of Abbreviations	vii

### **1: Introduction**

1.1 Rotavirus History	1
1.2 Rotavirus Molecular Biology	4
1.3 Rotavirus Life Cycle	5
1.4 Rotavirus Detection	11
1.5 Rotavirus Classification	12
1.6 Rotavirus Genotypes	15
1.6.1 Emerging Rotavirus Strains	19
1.7 Rotavirus Vaccines	22
1.7.1 Vaccine Use in Canada	28
1.7.2 Cost Effectiveness	28
1.8 Rotavirus Immunity	29
1.9 Study Background	31
<b>1.10 Hypothesis</b>	<b>32</b>
1.10.1 Scientific Reasons for Approach Taken to Answer Hypothesis	32
<b>1.11 Objective</b>	<b>33</b>

### **2: Methods**

2.1 Sample Collection	35
2.2 Stool Clarification	35
2.3 Genome Extraction	35
2.4 Sample Screening	36
2.5 Genotyping by Heminested Multiplex PCR	37
2.6 Genotype Confirmation	39
2.7 Positive Control Proliferation	40
2.8 Full Genome Sequencing	42
2.8.1 Primer Development	43
2.8.2 Genome Alignment	45
2.8.3 Analysis of Open Reading Frames	45
2.8.4 Phylogenetic Analysis	46
2.9 Real Time PCR Development and Validation	48
2.9.1 Culturing of Positive Controls	52

<b>3: Results</b>	
3.1 Heminested Multiplex PCR	54
3.2 Full Genome Analysis	57
3.2.1 Analysis of Reassortment	57
3.2.2 Analysis of Phylogenetics and Motifs	62
3.2.3 Analysis of Heminested Multiplex PCR Assays	95
3.3 Real Time PCR Assay	98
<b>4: Discussion</b>	
4.1 Heminested Multiplex PCR	103
4.2 Full Genome Analysis	107
4.2.1 Reassortment	107
4.2.2 Open Reading Frame Phylogenetics and Motifs	109
4.2.3 Summary of Phylogenetic Analysis	129
4.2.4 Assessment of Current Heminested Multiplex PCR Methods	130
4.3. Real Time PCR Assay	134
4.4 Future Directions	137
<b>5: Conclusion</b>	138
<b>6: References</b>	140
<b>7: Appendix</b>	162

## List of Tables

Table 1. Primers used in heminested multiplex PCR	38
Table 2. PCR thermocycling conditions	39
Table 3. Description of motifs analyzed in full genome analysis	47
Table 4. Primers analyzed for quality control of heminested multiplex PCR	48
Table 5. VP2 real-time assay primers and probes	51
Table 6. Full genome analysis results table for Canadian strains	59
Table 7. Sub-genotypic lineages after phylogenetic analysis	61
Appendix:	
Table A1. Samples collected from paediatric hospitals	162
Table A2. Reference strain accession numbers	163
Table A3. Reference strain full genotypes	164
Table A4. Primers for full genotyping assay	164
Table A5. Raw data for heminested multiplex genotyping assay	165
Table A6. Canadian strain accession numbers	166
Table A7. Contingency table for C1 strains	167
Table A8. Contingency table for C2 strains	167

## List of Figures

Figure 1: Rotavirus image (EM)	2
Figure 2: Conceptual rotavirus image (layered)	6
Figure 3: Image of a viroplasm inside a cell	9
Figure 4: Continental rotavirus prevalence pie graph	17
Figure 5: Conceptual diagram of the two licensed oral rotavirus vaccines	24
Figure 6: Bar graph of Canadian rotavirus genotypes from 2007-2010	56
Figure 7: Map of Canadian rotavirus sites and genotype prevalences from 2007-2010	57
Figure 8: Phylogenetic tree of VP1 strains	63
Figure 9: Representative amino acid alignment of VP1 Polymerase Motifs	64
Figure 10: Phylogenetic tree of VP2 strains	66
Figure 11: Phylogenetic tree of VP3 strains	68
Figure 12: Phylogenetic tree of VP4 strains	71
Figure 13: Representative amino acid alignment of VP4 antigenic motifs	74
Figure 14: Phylogenetic tree of VP6 strains	75
Figure 15: Phylogenetic tree of VP7 strains	77
Figure 16: Representative amino acid alignment of VP7 antigenic motifs	80
Figure 17: Phylogenetic tree of NSP1 strains	82
Figure 18: Phylogenetic tree of NSP2 strains	84
Figure 19: Representative amino acid alignment of the putative HIT-like domain	85
Figure 20: Phylogenetic tree of NSP3 strains	87
Figure 21: Phylogenetic tree of NSP4 strains	90
Figure 22: Representative alignment of NSP4 enterotoxin/VP4-binding domain	92
Figure 23: Phylogenetic tree of NSP5 strains	94
Figure 24: Representative real-time assay standard curve graphs for C1 strains	101
Figure 25: Representative real-time assay standard curve graphs for C2 strains	102

## List of Abbreviations

AGE – Acute Gastroenteritis

CDC – Centre for Disease Control

DLP – Double Layered Particle

DNA – Deoxyribonucleic Acid

DPBS – Dulbecco's Phosphate Buffered Saline

dsRNA – Double Stranded Ribonucleic Acid

EIA – Enzyme Immunoassay

EM – Electron Microscopy

ER – Endoplasmic Reticulum

GAVI – Global Alliance for Vaccines and Immunizations

IRF3 – Interferon Regulatory Factor 3

LA – Latex Agglutination

NACI – National Advisory Committee for Immunizations

NML – National Microbiology Laboratory

NSP – Non-Structural Protein

nt – Nucleotide

ORF – Open Reading Frame

PAGE – Polyacrylimide Gel Electrophoresis

PBS – Phosphate Buffered Saline

PCR – Polymerase Chain Reaction

RCWG – Rotavirus Classification Working Group

RNA – Ribonucleic Acid

RoV – Rotavirus

RT-PCR – Reverse Transcriptase Polymerase Chain Reaction

TLP – Triple Layered Particle

VP – Viral (Structural) Protein



## 1. Introduction - Rotavirus Literature Review

### Rotavirus:

Rotavirus is a virus from the Reoviridae family of viruses. It is responsible for causing acute gastroenteritis predominantly in infants between the ages of 3 months and 2 years. Acute gastroenteritis symptoms include diarrhea, vomiting and pyrexia. Presently, Rotavirus is responsible for the deaths of an estimated 600,000 people per year, the majority of which occur in developing parts of the world, such as Sub-Saharan Africa and South Asia [Dennehy. 2008].

### 1.1 Rotavirus History:

Throughout the 1950s and 1960s the cause of most severe gastroenteritis in young children eluded scientists. That changed in 1973, when Bishop and colleagues first identified a 70nm viral particle from a patient's duodenal mucosa using thin-layer electron microscopy (EM) [Estes and Kapikian. 2007]. The rotavirus was subsequently identified in faecal samples from infants and young children, also by EM, depicted in Figure 1 [Riepenhoff-Talty et al. 1983]. Rotavirus was shown to be responsible for 35-50% of hospitalizations due to severe gastroenteritis, for those less than 2 years of age [Estes and Kapikian. 2007], and was shown to be the dominant cause of diarrheal disease worldwide in infants and young children and estimated to cause 870,000 deaths per year during the 1980s [Estes and Kapikian. 2007, Greenberg and Estes. 2009].

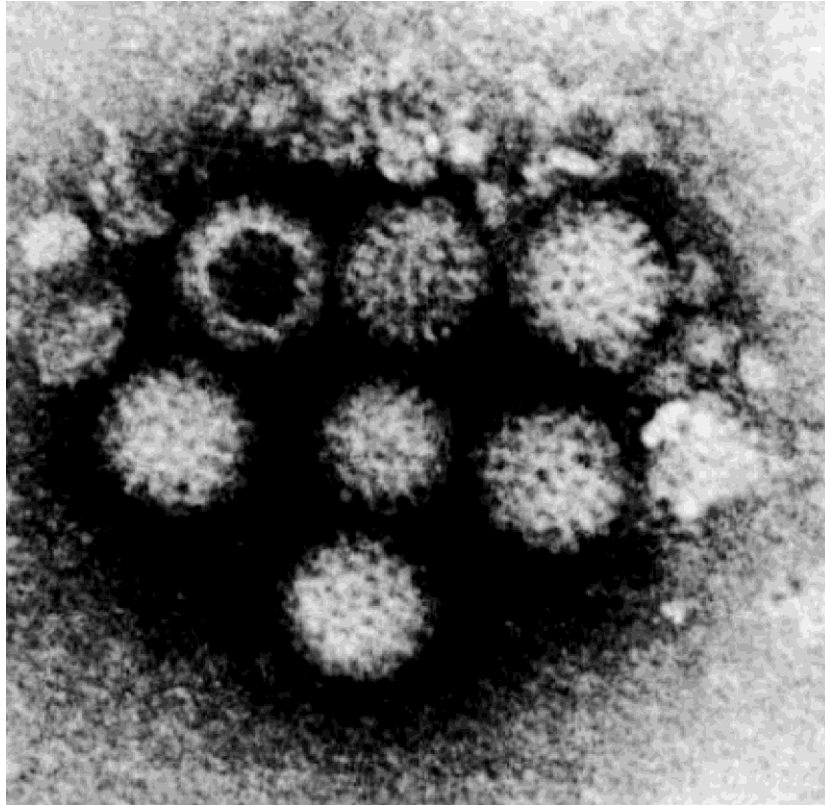


Figure 1: Electron microscopy image of rotavirus particles from a fecal specimen. The image is magnified 165,000 times. Adapted from *Annals of the New York Academy of Sciences* (1983) 420: 391-400 "Negative Staining and Immune Electron Microscopy as Techniques for Rapid Diagnosis of Viral Agents". Riepenhoff-Talty, M. *et. al* Figure 5 with kind permission from John Wiley and Sons.

It is estimated that 95% of people worldwide have been infected with rotavirus by the age of 5 [Dennehy. 2008, Matthijnsens et al. 2008b]. Severity of disease varies from symptom-free to severe gastroenteritis and dehydration. Estimates of the burden of rotavirus disease vary, but it is clear that rotavirus disease causes mortality and substantial morbidity in developing countries. One global estimate of the burden of rotavirus is that 1 in 65 children infected with rotavirus will require hospitalization due to rotavirus, while 1 in 293 will die [Matthijnsens et al. 2008b]. In the developed world estimates differ slightly. A study, involving children with health insurance coverage in

the United States, estimated 1 in 27 will require an emergency department visit, while 1 in 74 will require hospital admission by the age of 5 [Cortes et al. 2009]. Another study, highlighting the burden of rotavirus in 2 United States of America counties, estimates emergency department visits will be required by 1 in 13 to 1 in 17 children by the age of 5, depending on the county [Yee et al. 2008]. Only rough estimates can be made on the burden of rotavirus disease in a North American population, as disease rates change from year to year and area to area. In addition, testing methods, such as rectal swabs, are not always adequate to detect rotavirus and may result in false negatives. Hospital code use for acute gastroenteritis (AGE) may result in underestimation of rotavirus burden as well [Mast et al. 2010, Yee et al. 2008].

Rotavirus seasonality varies with climate. Temperate climates, such as those in Canada and the United States of America, get a rotavirus season (and rotavirus incidence peaks) in the colder months or in the latter part of winter. Tropical climates often have rotavirus epidemics year round, with a less pronounced peak of incidence [Dennehy. 2008].

In North America, the mortality rate of rotavirus is relatively low; nevertheless, the economic burden of rotavirus is substantial. In the U.S. healthcare costs of rotavirus diarrhea are nearly \$300 million [Cortes et al. 2009]. However, that estimate is too low when considering other costs, especially when incorporating societal costs (work days lost, disease-related purchases, etc.). Including societal costs, an estimate for the total cost of rotavirus-related disease is near \$893 million in the U.S.A. [Widdowson et al. 2007]. In Canada, the economic rotavirus-disease burden, including societal costs, has been estimated to be from \$8.9 million to \$18.4 million per year [Le Saux et al. 2011].

## 1.2 Rotavirus Molecular Biology:

Reoviridae are non-enveloped, icosahedral viruses with segmented double stranded RNA genomes. In the case of rotavirus, there are 11 genome segments, each of which encode a single protein. Occasionally, there is a second protein encoded by the NSP5 genome segment. Generally, the rotavirus genome encodes six structural proteins, VP1-4, VP6, VP7, and six non-structural proteins, NSP1-6 [Greenberg and Estes. 2009].

Viral particles consist of three concentric layers, the inner capsid, middle capsid and outer capsid, as shown in Figure 2. The inner capsid is composed of the VP1 (RNA dependant RNA polymerase), VP2 (core scaffold/shell protein) and VP3 (capping enzyme) proteins. The VP2 protein forms the shell of the inner capsid in an arrangement of 120 quasi-equivalent proteins, arranged into decamers. The middle capsid is composed of 260 trimers of VP6 (middle capsid shell) protein. The outer capsid is composed of the 60 trimers of VP4 (spike) protein and 260 trimers of VP7 (outer capsid shell) protein [Greenberg and Estes. 2009, Guglielmi et al. 2010]. The segmented rotavirus genome is attached to the core at 11 of the 12 five-fold vertices formed by joining VP2 proteins. Each of the 11 genome segments is bound to one VP1 polymerase complex and VP3 capping enzyme at one of the 12 five-fold vertices, shown in Figure 2 [Guglielmi et al. 2010, Jayaram et al. 2004].

The rotavirus non structural proteins (NSPs) play a variety of roles in rotavirus pathogenesis. The NSP1 protein is thought to antagonize the cellular innate immune response by causing the degradation of factors such as interferon regulatory factor (IRF) proteins [Heiman et al. 2008]. Both NSP2 and NSP5 play important roles in the

development of viroplasms. NSP2 and NSP5 have RNA-binding and protein-binding capabilities to recruit proteins to viroplasms facilitating replication. NSP3 is able to bind cellular mRNA, inhibiting cellular translation. Also, NSP3 may recruit cellular translational components to viral genome segments in the initial stages of replication, to facilitate viral translation [Arnoldi and Burrone. 2009]. Functions of the NSP4 include facilitating the viral replication cycle, via chaperoning core proteins and functioning as enterotoxin, inducing diarrhea [Arnoldi and Burrone. 2009, Ball et al. 2005, Greenberg and Estes. 2009].

### 1.3 Rotavirus Life Cycle:

Rotaviruses are transmitted via the fecal-oral route. The rotavirus triple layered particle (TLP) is a very stable particle and able to survive the harsh environment of the digestive system [Greenberg and Estes. 2009]. The rotavirus particles pass through the stomach into the small intestine, where non-dividing differentiated enterocytes are infected by rotavirus. The targeted mature enterocytes are located at the tips of intestinal villi [Greenberg and Estes. 2009, Isa et al. 2008].

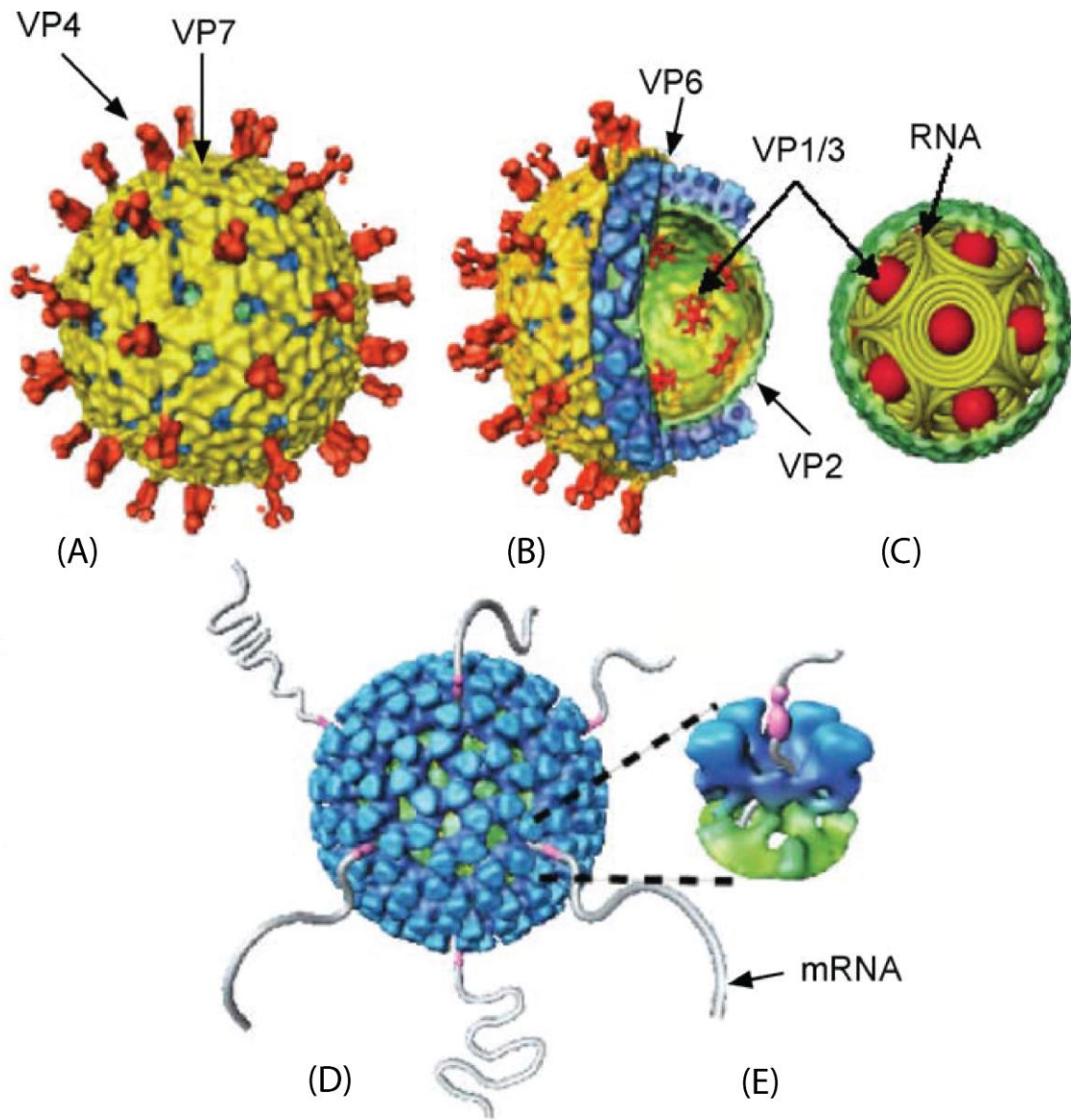


Figure 2: Cryo-EM reconstruction of the rotavirus triple-layered particle. (A) Full TLP with VP4 proteins in red and VP7 outer capsid proteins in yellow. (B) Cutaway of the TLP exposing the green core and blue middle capsid layer. (C) Genomic RNA depicted as conical spirals surrounding the red RNA-dependant RNA-polymerase particles (VP1) within the core. (D) Depiction of the middle capsid layer with mRNA transcripts protruding from five-fold vertices of the DLP. (E) Blown up representation of the transcription of viral mRNA from a five-fold vertex. Adapted from *Virus Research* (2004) 101: 67-81 "Emerging themes in rotavirus cell entry, genome organization, transcription and replication". Jayaram, H. et. al Figure 1 with kind permission from Elsevier.

Entry into the target cell is still not clearly understood. However, it is evident that trypsin-activated TLPs enter the target cell quickly by direct penetration, while viruses not activated by trypsin can be absorbed by endocytosis at much slower rates [Baker and

Prasad. 2010]. However, non trypsin-cleaved rotaviruses are transcriptionally inactive. This indicates that trypsin-based protein cleavage facilitates viral activation. Currently, the proposed mechanism of cell entry involves the VP4 proteins being cleaved by trypsin, into two proteins, VP5\* and VP8\* (\* indicates cleaved product of VP4 protein). The newly activated spike proteins are held in place by calcium stabilized VP7 outer capsid shell proteins. Initial binding occurs through the VP8 portion of the spike protein with either sialic acid receptors or integrins, thought to be found on lipid rafts of the cell. Next, newly exposed spike protein domains can interact with integrin proteins such as  $\alpha 2\beta 1$ , then heat shock protein 70 and other integrins [Baker and Prasad. 2010, Greenberg and Estes. 2009, Lopez and Arias. 2006].

During cell entry the viral particle sheds its outer capsid layer, resulting in the internalization of only the double-layered particle (DLP). The DLP acts as a transcription centre for viral RNA in the cytoplasm [Greenberg and Estes. 2009]. Using the templates contained within the core the VP1 protein transcribes plus-sense (messenger) RNA, which is capped at the 5' end by the VP3 protein, then released into the cytoplasm for translation. Genomic RNA is conserved inside the core during this process. Only positive sense RNA transcripts are released to the cytoplasm through the type 1 channels. Type 1 channels are created by the alignment of the 5-fold vertices of VP2 and VP6 protein shells. The channels are located above each VP1 polymerase complex [Guglielmi et al. 2010].

Core formation of rotavirus progeny occurs via a unique process involving a cytoplasmic gathering of messenger RNA, proteins and genomic RNA. These molecular gatherings, called viroplasms, can be visualized in rotavirus infected cells as membrane-

free inclusion bodies surrounded by host ribosomes, as shown in Figure 3 [Patton et al. 2006]. The non-structural proteins NSP2 and NSP5 are critical for the recruitment of proteins to viroplasms. Although not completely understood, the non-structural proteins organize the structural proteins and complete genome complements of positive sense RNA, in such a way that cores can assemble with all the necessary components. It has been suggested that this assembly process occurs simultaneously and that the correct genome complement is integrated due to RNA-RNA interactions between positive-sense RNA segments [Pesavento et al. 2006]. This process is likely to be partly mediated by the natural affinity of the components (core proteins) for each other, but also by the affinity of NSP2 to VP1 and NSP5 to VP2 [Patton et al. 2006]. While the centre of the viroplasm is the primary assembly site for the core particles, there are few particles with the middle capsid VP6 layer. However, on the outer edges of the viroplasm VP6 is prevalent, possibly recruited to that position by NSP4 protein. Hence, as the core particles exit the viroplasm, they are changed into DLPs as the VP6 proteins integrate onto core proteins [Patton et al. 2006].

Genome replication also occurs in the viroplasm. Short interfering RNA evidence has indicated that, unlike RNA destined for protein production, genomic RNA found in viral progeny does not originate outside of the viroplasm [Silvestri et al. 2004]. Evidence suggests genome replication occurs within the viroplasm, which also means that there are some transcriptionally active double-layered particles within the viroplasm, responsible for genome replication. This was supported in a study by labelled nucleotide analysis [Patton et al. 1997, Patton et al. 2006]. This observation led to the suggestion that viroplasms were seeded by DLPs dawned from infecting rotavirus particles [Silvestri et



al. 2004]. However, more recent evidence, from a study using Cy5-labelled infecting rotavirus particles, indicates that DLPs in viroplasm do not originate from infecting particles [Carreno-Torres et al. 2010].

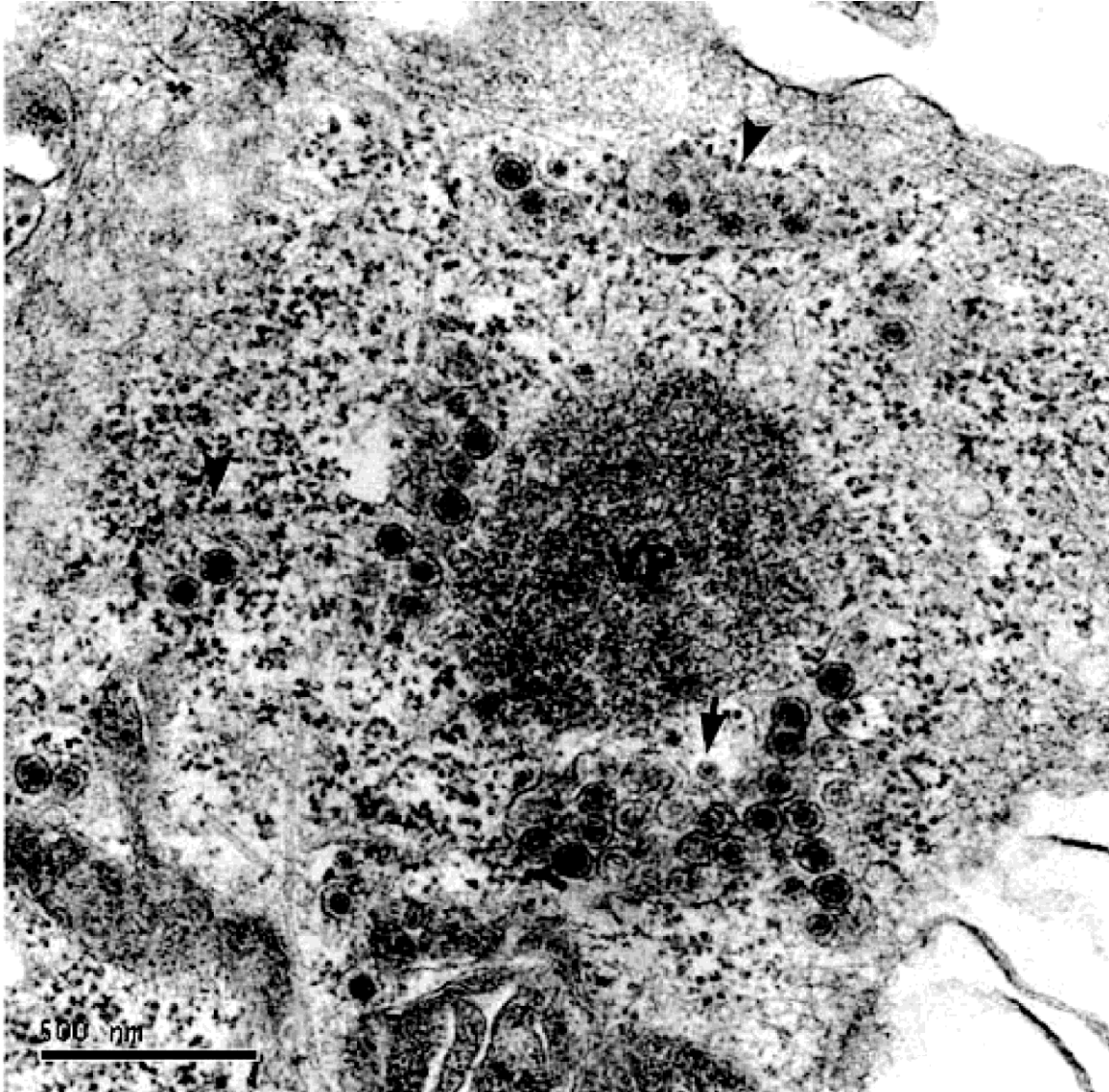


Figure 3: A viroplasm (VP) within a cell surrounded by ER vesicles, where TLPs form and collect during rotavirus infection. The arrowhead points at ER vesicles covered with ribosomes, while the arrow points to a DLP bound to the ER membrane exterior. Adapted from *CTMI* (2006) 309: 169-187 "Rotavirus Genome Replication and Morphogenesis: Role of the Viroplasm". Patton J.T. *et. al* Figure 1 with kind permission from Springer Science+Business Media.

After assembly in the viroplasm, DLPs are recruited to the endoplasmic reticulum (ER) by NSP4. The DLPs proceed to bud into the ER, where VP7 proteins are integrated

onto the particle, transiently, by a mechanism that is not completely understood.

However, this process involves displacing ER membrane with VP7 protein as the process continues. The VP4 spike proteins are likely incorporated on the particle as it exits the cellular membrane [Pesavento et al. 2006].

The NSP4 enterotoxin protein is a key factor in inducing viral shedding by diarrhea. NSP4 is a versatile protein, existing as both a membrane associated protein as well as a widely distributed intercellular protein. The C-terminal portion of the domain is capable of migration to other cells causing diarrhea inducing conditions to spread to uninfected cells neighboring infected cells [Ball et al. 2005]. The enterotoxin domain of the NSP4 is capable of making intestinal crypt cells permeable to  $\text{Cl}^-$ , which in turn, mobilizes calcium ions ( $\text{Ca}^{2+}$ ).  $\text{Ca}^{2+}$  is very important to rotavirus infection.  $\text{Ca}^{2+}$  plays a role in outer capsid stabilization and shedding during entry. It also acts as a messenger, indicating cellular damage, inducing cell death. As cells making up the upper intestinal microvilli die, the absorption properties of the intestine are hampered, resulting in a high amount of fluid remaining in the intestine. The influx of intercellular calcium ions into the cytoplasm causes cell death at high concentrations. NSP4 can also induce plasma membrane permeability to calcium ions in healthy cells, causing fluid loss to the intestine [Ruiz et al. 2000]. Research has also shown that NSP4 inhibits a sodium channel symporter in rabbits. The symporter SGLT1 is specifically designed for the reabsorption of water in the intestine [Halaihel et al. 2000]. All mechanisms contributing to the increase in fluid in the intestine contribute to diarrhea. NSP4 can also bind microtubules on the cellular membrane. Binding microtubules may result in lack of activity in intestinal brush border disaccharidases, which can result in the sustenance of diarrhea.

Finally, NSP4 interacts with several cellular calcium ion chaperone and regulatory proteins, which effects calcium distribution and possibly the enteric nervous system [Ball et al. 2005, Ruiz et al. 2000].

#### 1.4 Rotavirus Detection:

The diagnosis of rotavirus is a key factor in rotavirus gastroenteritis prevention and control. As gastroenteritis can be caused by a number of enteric pathogens in children, it is important to identify the cause in order to contain and control outbreaks, especially for viruses with a low infectious dose such as norovirus and rotavirus. Enteric virus outbreaks present with similar symptoms, so identification of the pathogen involved is dependant on laboratory testing [Feeney et al. 2011]. Some other viruses known to cause AGE include norovirus, adenovirus 40, adenovirus 41, astrovirus and sapoviruses [Cunliffe et al. 2010].

Several enzyme immunoassay (EIA) kits and latex agglutination (LA) assays are available and are typically used in a clinical setting for fast detection of rotavirus. However, EIA and LA methods are unreliable due to cross-reactivity, antigenic drift, low specificity and low sensitivity when compared to polymerase chain reaction (PCR) based methods [Gomara et al. 2001, Jothikumar et al. 2009, Nordgren et al. 2010]. Electron microscopy (EM) has been used for detection of rotavirus since the discovery of the virus in 1973. However, this method is limited by low sensitivity, technical expertise and cost of equipment in health care facilities [Jothikumar et al. 2009, Pang et al. 2004]. Recently, developments in real time PCR technology have opened a new door for fast and sensitive detection of rotavirus and other enteric pathogens. Several assays have been published.

Rotavirus real-time PCR assays have been based on NSP3, VP6 and VP7 gene segments and, thus far, are the most sensitive rotavirus detection method [Feeney et al. 2011, Jothikumar et al. 2009, Nordgren et al. 2010, Pang et al. 2004].

### 1.5 Rotavirus Classification:

In addition to detection of rotavirus, methods to classify rotavirus have been drastically improved during the past 2 decades. In the past, rotaviruses were classified by their electropherotypes and serogroup types. Electropherotyping involved the separation of the 11 genome segments in polyacrylimide gel electrophoresis (PAGE) gels. In addition, two monoclonal antibodies, specific to the VP6 protein, further classified the rotavirus into one of four different subgroups (SGI, SGII, SGI+SGII or non SGI + non SGII). Depending on antigenicity and the separation pattern on the PAGE gel, rotaviruses fell into one of seven serogroups (A-G) [Matthijnssens et al. 2008a]. As more knowledge of rotavirus was gained, serotyping of the two outer capsid proteins (VP4 and VP7) became the prominent method for classification. Rotavirus strains were classified based on several specific neutralizing monoclonal antibodies. However, use of monoclonal serotype-specific antibodies is limited by availability of the antibody for every serotype, including new ones, as well as sample state (DLP vs. TLP) [Fischer and Gentsch. 2004, Gouvea et al. 1990]. If the virus in the sample contains too few triple layered particles the antibodies will fail [Fischer and Gentsch. 2004].

With improvements in technology came improvements in rotavirus classification. The next form of classification developed was genotype based, which was not dependent on TLPs or creating new antibodies. As the VP4 and VP7 proteins are key antigenic

proteins, they became the basis of a new system of genotyping, which correlated highly with the serotyping assays of the past [Gentsch et al. 1992, Gouvea et al. 1990]. The dual nomenclature system presently used for classification of rotaviruses includes a G type (VP7) and a P type (VP4). The P type is written with its genotype number in square brackets, which is occasionally preceded by its serotype number (P1B[4]). For the remainder of this genotyping study the serotype will not be included as part of the P type. The methods used to genotype rotaviruses involved an initial reverse transcriptase PCR (RT-PCR) using two primers complimentary to consensus regions of the VP4 or VP7 regions. Subsequent PCR reactions involving multiplexes of unidirectional genotype-specific primers were used to identify the specific VP4 or VP7 genotype [Gentsch et al. 1992, Gouvea et al. 1990]. The amplicon sizes are determined by agarose gel electrophoresis and used to determine genotype.

As genomic sequencing methods become cheaper and more effective, the standard for classification is beginning to change again. Recently, it has been proposed that classification be done as a comprehensive comparison of rotavirus genomes. A rotavirus classification working group (RCWG) was established in 2008 and recommended the completed open reading frame (ORF) sequencing of the rotavirus genome for classification. As there are 11 genome segments to consider, the new genome nomenclature involves giving each segment a genome number based on nucleotide identity cut-off percentages. The letter to designate the segment is dependent on an aspect of the function of its protein. In the segment order of VP7-VP4-VP6-VP1-VP2-VP3-NSP1-NSP2-NSP3-NSP4-NSP5 the genotype nomenclature is G<sub>x</sub>-P<sub>[x]</sub>-I<sub>x</sub>-R<sub>x</sub>-C<sub>x</sub>-M<sub>x</sub>-A<sub>x</sub>-N<sub>x</sub>-T<sub>x</sub>-E<sub>x</sub>-H<sub>x</sub>, where x is the genotype number for a given segment and capital letters

refer to attributes of their respective proteins [Matthijnsens et al. 2008a]. The cut-off percentages were created with nucleotide identity rather than amino acid identity for several reasons. Firstly, there is less overlap in nucleotide identity between genotypes, which means that the vast majority of strains fall into only a single genotype. Secondly, many gene segments are too highly conserved at the amino acid level to distinguish between genotypes. Thirdly, nucleotide identity based trees are more highly supported by bootstrapping than were amino acid identity based trees. Finally, diversity due to point mutations is more evenly distributed across the genome than amino acid variations, making cut-off values more reliable than those from amino acids [Matthijnsens et al. 2008a].

Study of the full genomes of rotaviruses has revealed two common genetic pools and one rarer genetic pool. The three pools are named after their genogroup defining strains. The Wa-like genogroup is the most common genogroup. Phylogenetic evidence supports Wa-like strains sharing a common origin with many porcine rotaviruses. The DS-1-like genogroup is the second most common to infect humans. The DS-1-like strains seem to have a common origin with bovine rotaviruses [Matthijnsens et al. 2008b]. The third genogroup, named AU-1-like rotaviruses, is rare in comparison to the other two. AU-1-like viruses seem to have a common origin with feline rotaviruses and possibly canine and bovine rotaviruses amongst others [De Grazia et al. 2010].

PCR methods are only reliable if the primer matches the template. Due to the high rate of genetic drift in the rotavirus genome, PCR assays are subject to failure due to nucleotide changes in primer binding regions. In order to ensure the assays do not fail due to genetic drift, primers for these assays must be checked and either re-designed or

modified regularly [Iturriza-Gomara et al. 2000, Iturriza-Gomara et al. 2004, Simmonds et al. 2008]. As novel strains are discovered, new primers can be added to the heminested multiplex assays to make the assays more robust [Aladin et al. 2010, Das et al. 1994, Iturriza-Gomara et al. 2004]. For genotyping assays, new primers should be created based on regions that are approximately consensus, but also genotype-specific. Although the 5' and 3' ends of most segments are relatively stable genetically, there are regions in each of the ORFs that may be subject to higher rates of mutation [Matthijnssens et al. 2008b]. When creating new primers for sequencing, care should be taken to make primers complementary to regions that are stable and non-genotype specific.

#### 1.6 Rotavirus Genotypes:

Genotyping of rotavirus has become an important surveillance technique worldwide. For several decades now, the incidence of different rotavirus genotypes in many countries has been monitored to identify the most prevalent disease-causing rotaviruses in various regions of the world. Genotypes can be determined using percentage identity cut-off values. For the G and P genotypes the identity cut-off value has been defined as 80% [Matthijnssens et al. 2008a]. A total of 27 different G genotypes and 35 different P genotypes have been detected to date [Matthijnssens et al. 2009, Matthijnssens et al. 2011, Santos and Hoshino. 2005]. However, only 11 different G and 12 different P genotypes have been found in humans [Matthijnssens et al. 2008a]. Rotavirus genotyping and surveillance is necessary to monitor new and emerging strains, as well as for identifying potential recombinants that may become health threats in the future. Furthermore, establishment of which circulating strains are predominant can be

used to evaluate the potential effectiveness of vaccines. Rotavirus surveillance also allows the quantification of diarrheal causes that are not due to rotaviruses [Desai et al. 2010].

Early rotavirus surveillance was compiled in a comprehensive study of rotavirus surveillance from the years 1973 to 2003 [Santos and Hoshino. 2005]. The study included genotype data from 124 studies published from 1989-2004, which included more than 45,000 human rotavirus samples from over 50 countries. The study indicated that, worldwide, the most prevalent genotype that caused rotavirus infection was G1P[8], which is also frequently the most prevalent rotavirus genotype in the developed world. Other commonly found genotypes were G2P[4], G3P[8], G4P[8] and G9P[8]. A global presentation of all samples, which include both P and G types in the study (n=16474), indicated the highest prevalence rates were 64.7% for G1P[8], 12.0% for G2P[4], 3.3% for G3P[8], 8.5% for G4P[8] and 2.7% for G9P[8], as seen in Figure 4. However, these values were biased toward samples from the developed world, as most of the data were from Europe, Australia and North America [Santos and Hoshino. 2005].

National rotavirus genotyping studies have continued to emerge in the literature. Recent studies have indicated various genotype frequency shifts in rotavirus prevalence and emergences of novel genotypes around the world. One such study in the United States showed a shift of rotavirus genotype prevalence in Rochester, New York from 87% G1P[8] in 2006 to 69% G12P[8] in 2007 [Payne et al. 2009].

Recent European studies from Ireland, France and Spain have shown that, despite year-to-year and region-to-region fluctuations in prevalence, on average G1P[8], continues to account for around 60% of rotavirus strains [Cilla et al. 2010, de Rougemont



et al. 2010, Lennon et al. 2008]. Similarly, studies from ongoing surveillance in Australia have indicated G1P[8] to be the most prevalent strain for the majority of years surveillance had taken place [Kirkwood et al. 2006a, Kirkwood et al. 2006b, Kirkwood et al. 2007, Kirkwood et al. 2008, Kirkwood et al. 2009, Kirkwood et al. 2010]. In addition, studies from the African surveillance network, Japan and New Zealand have also shown G1P[8] to be the most prevalent strain [Chandrasekaran et al. 2010, Esona et al. 2010, Yoshinaga et al. 2006].

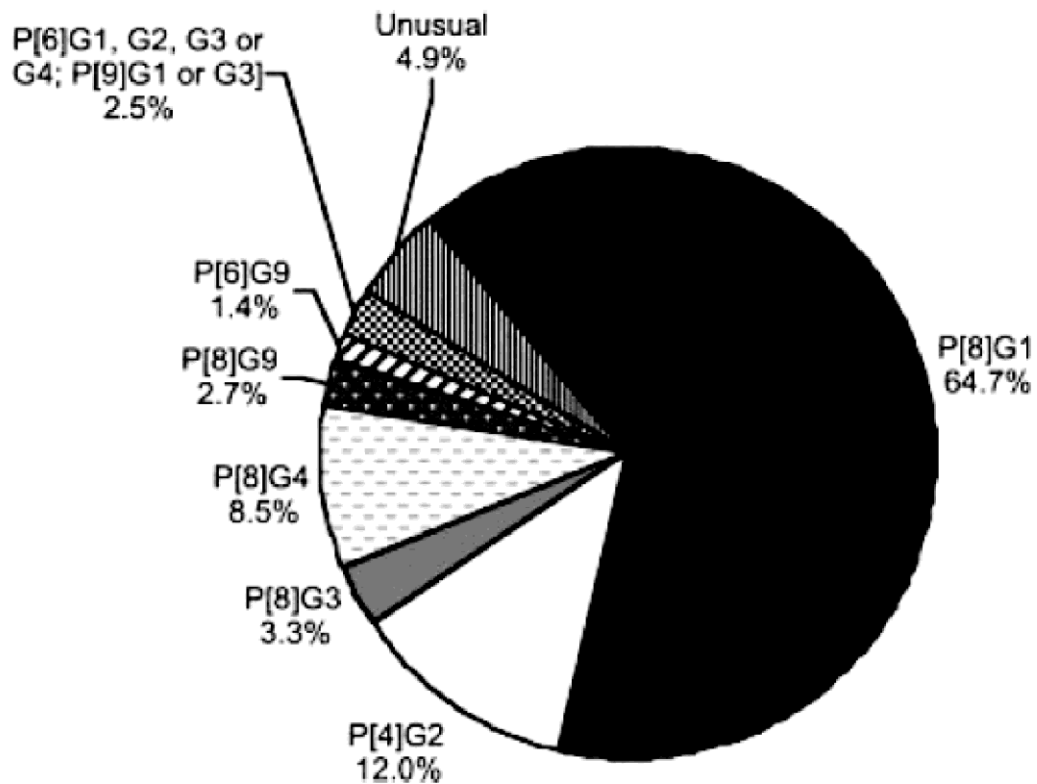


Figure 4: Global prevalence of rotavirus genotypes. This pie chart represents 16,474 samples from 124 studies located on 5 continents published between 1989 and 2004. Adapted from *Reviews of Medical Virology* (2005) 15: 29-56 "Global distribution of rotavirus serotypes/genotypes and its implication for the development and implementation of an effective rotavirus vaccine". Santos, N and Hoshino, Y. Figure 1 with kind permission from John Wiley and Sons.

Some studies from the past decade, such as those from Brazil, Malaysia and South Korea, have shown strains other than G1P[8] to be the most prevalent strains. However, these studies showed that the most prevalent strain was one of the other four most common strains. The Korean study also showed a drastic decrease in G9P[8] prevalence from about 69% (n=75) to about 5% (n=64) in consecutive rotavirus seasons [Esteban et al. 2010, Han et al. 2010, Zuridah et al. 2010].

There have been only two studies of rotavirus genotype incidence in Canada; both collected data from single sites during single seasons. The first Canadian rotavirus genotyping study took place in Toronto, Ontario, and was based on phylogenetic genotyping of the VP7 segment only, in samples from the 1997-1998 rotavirus season. The study found G1 to be the most prevalent genotype, followed by G2, G3, G4 and G9 which were detected at very low frequency [Kostouros et al. 2003]. Similar results were found in the other study from Edmonton, Alberta during the 2001-2002 rotavirus season: no G3 or G9 rotaviruses were identified [Pang et al. 2004].

The data from past rotavirus surveillance studies indicate several aspects of rotavirus genotype frequency. Firstly, the most prevalent strains of rotavirus circulating worldwide are G1P[8], G2P[4], G3P[8], G4P[8] and G9P[8]. Secondly, fluctuations in prevalence are frequent. The nature of these fluctuations is not understood and is unpredictable [Han et al. 2010, Kirkwood et al. 2009, Kirkwood et al. 2010, Santos and Hoshino. 2005]. It is possible that rotavirus strains in adjacent regions can have completely different prevalence profiles, indicating that the prevalent strain in one area is not necessarily indicative of the prevalence of rotavirus strains in another [de Rougemont et al. 2010].

There have only been a handful of studies that analyze full open reading frames of rotavirus [Matthijssens et al. 2008b, Pietsch and Liebert. 2009, Than et al. 2011]. Full genome classification will prove useful in the future for many reasons. Firstly, reassortment events amongst different genome segments are not well understood. Some segments reassort more frequently than others. Full genome analysis is a good way to analyze how segment reassortment is driven in emerging strains, and to identify genetic drift due to selection by vaccines or other unknown factors [Kirkwood et al. 1993, Matthijssens et al. 2010a, McDonald et al. 2009]. In addition, genetic drift can be monitored closely by full genome sequencing, including tracking of antigenic sites. Moreover, genetic information collected can be used further, to analyze different motifs related to genotypes and infectivity. As more information is collected genetic markers of successful strains may become more apparent.

#### 1.6.1 Emerging Rotavirus Strains:

The rotavirus genome is segmented, so it is possible that reassortment can occur during a mixed infection. Reassortment events occur frequently in G and P genome segments [Matthijssens et al. 2008b, Matthijssens et al. 2011, Santos and Hoshino. 2005]. Furthermore, reassortment is an important, but not well understood mechanism of rotavirus genetic variation, in all genome segments [Laird et al. 2003, McDonald et al. 2009]. Studies analyzing reassortment events in circulating and emerging strains are important to ensure current vaccines are appropriate to create maximal immunity in the population. They are also important to elucidate the incidence of naturally occurring reassortment events in general; for instance, the preferences of reassortment in genome

segments, the effects of evolutionary pressures and the temporal and regional differences in reassortment viruses can be identified [McDonald et al. 2009].

All of the most common genotypes have at least one G or P genotype in common with the RotaTeq (RV5) and Rotarix vaccines. However, the P[4] genotype and G9 genotype are part of two of the five most prevalent rotavirus strains and are not represented in either vaccine. Notably, the vaccines have been shown to provide a degree of cross protection, but it is not as high as the homotypic protection provided by the vaccines. [Dennehy. 2008, Matthijnssens et al. 2009, Santos and Hoshino. 2005].

Emerging G9 strains have likely been circulating and spreading in many areas of the world since the discovery of G9, in Philadelphia, U.S.A. in 1983. Subsequently, the G9 genotype was found in Japan in 1985-1986, and Yugoslavia, Thailand and India, often in combination with uncommon P types [Laird et al. 2003]. Since its recognition G9 has become one of the five most prevalent rotavirus strains detected globally [Laird et al. 2003]. In addition, G9 has emerged as the most dominant genotype in certain rotavirus seasons in certain regions. In the late 1990s G9 was the most prevalent strain in Sapporo, Japan and Alice Springs, Australia [Santos and Hoshino. 2005]. Since the year 2000, G9 has been found to be the most prevalent in many regions during annual surveillance, including ones in Argentina, France, Portugal, Thailand, South Korea and the United States of America [Abdel-Haq et al. 2011, de Rougemont et al. 2010, Esteban et al. 2010, Han et al. 2010, Matthijnssens et al. 2009]. Based on nucleotide identity there are three lineages of G9 rotaviruses; of these, only lineage III of G9 strains has emerged globally [Santos and Hoshino. 2005]. Of the five most prevalent genotypes, only G9 is not represented in either vaccine [Dennehy. 2008]. Also, the G9 strain is considered to have a

higher frequency of reassortment events that occur as compared to the other four most common circulating G genotypes [Kirkwood et al. 2003, Laird et al. 2003].

Other emerging rotavirus G types that show outbreak potential include G12, G8, G5 and G10. The G12 rotavirus strain was first discovered in the Phillipines in the late 1980s. It was not identified again until 11 years later, when it was found in the United States. Since 1999 the G12 rotaviruses have been recognized worldwide and, like G9, have been recorded as the most dominant strain during annual surveillance in at least one site [Matthijnssens et al. 2009, Payne et al. 2009, Santos and Hoshino. 2005]. During the 2007 rotavirus season in Rochester, New York, G12 was detected in 69% of the 111 rotavirus positive samples collected [Payne et al. 2009]. Unlike other emerging strains, G12 rotaviruses have, to date, only been found in humans. Both G8 and G10 rotavirus are most commonly associated with bovine species. However, G10 has been identified in many human rotavirus outbreaks in India, while G8 has been found in children in Africa and South Asia [Matthijnssens et al. 2009]. G8 rotaviruses have a high frequency of reassortment with both animal and human rotaviruses. G5 rotaviruses are most commonly associated with pig and horse species. However, there have been a few incidences of human infection by G5 rotaviruses beginning in 1982. G5 rotaviruses have regularly circulated in South America, and have been shown to be the dominant genotype for at least one rotavirus season. Human G5 rotaviruses have also been found in Cameroon [Santos and Hoshino. 2005].

Potentially emerging rotavirus P types, not including P[8] and P[4], include P[6], P[9] and P[11]. The P[6] genotype has often been found in asymptomatic rotavirus incidents of neonatal and infantile rotavirus, thereby nicknamed the “nursery” genotype.

However, P[6] has also been found in many symptomatic occurrences of rotavirus globally. The predominantly feline rotavirus strain, P[9], is a genotype being considered for inclusion in vaccine development. In 1977, the P[9] genotype was discovered in rotavirus-related diarrhea of a 14 year old boy, in Japan. Since then, it has been found in many countries worldwide and associated with several of the prevalent and emerging G genotypes. The P[11] rotavirus strain was found in an outbreak of rotavirus-related diarrhea in Pune, India. However, the strain is primarily of concern due to the outbreak and has not emerged worldwide [Santos and Hoshino. 2005].

#### 1.7 Rotavirus Vaccines:

The expansion of our knowledge of rotavirus disease in the 1980s and 1990s prompted the development of a rotavirus vaccine. The first rotavirus vaccine developed was a bovine-rotavirus based vaccine, designated as RIT4327. Despite successful vaccination of calves and fairly successful vaccination of Finnish children, the bovine RIT4327 vaccine was not successful enough in vaccinating children in Africa and Latin America. The RIT4327 vaccine was not further developed, but it did provide a starting point for future vaccine development. The RIT4327 vaccine also showed that a vaccine with antigenic mismatches could elicit some degree of immunity in many children [Dennehy. 2008, Greenberg and Estes. 2009].

The RIT4327 vaccine was followed by the RRV a vaccine, that was developed using a similar strategy. The RRV vaccine was based on the Rhesus Monkey rotavirus RRV. However, trials for the RRV vaccine showed that it was not highly efficacious [Dennehy. 2008]. The failure was thought to be caused by antigenic mismatch. The RRV

rotavirus was a G3 type virus, while the majority of circulating viruses at the time were G1, G2 or G4 viruses. The vaccine developer, Wyeth Pharmaceuticals, decided to modify the vaccine to make it more efficacious. Their new strategy involved generating recombinant viruses using the RRV strain as the vaccine backbone with human G1, G2 and G4 VP7 genome segments. The final version of the vaccine, RRV-TV or RotaShield, was a quadrivalent rhesus monkey recombinant vaccine. The vaccine underwent trials in the United States of America, Finland and Venezuela, all indicating that the vaccine was 80-100% efficacious in the prevention of severe diarrheal disease. In August of 1998 the vaccine was approved for use in the United States. In 1999 it was clear that, for unknown reasons, RotaShield caused a 25-fold increase in rates of intussusception. The vaccine was administered to an estimated 600,000 children from the United States in the time period following its release until its withdrawal from the market. Although deemed unsafe to market the vaccine for infants in the United States of America market, there was debate as to whether benefits of the vaccine would have outweighed the risks in developing countries, where mortality is far greater than in the United States [Greenberg and Estes. 2009].

After the failure of RotaShield, no rotavirus vaccines were available for several years. Finally, in 2006, two new vaccines (shown conceptually in Figure 5), were released in dozens of markets worldwide. Merck developed RotaTeq, a pentavalent bovine recombinant vaccine, using the bovine rotavirus strain WC3 as the backbone. The G6P[5] WC3 virus was recombined with the human VP7 genes for G1, G2, G3 and G4, as well as the VP4 gene for P[8]. Concurrently, GlaxoSmithKline (GSK) created a

monovalent, live attenuated vaccine using the human G1P[8] rotavirus strain 80-12 [Dennehy. 2008, Greenberg and Estes. 2009].

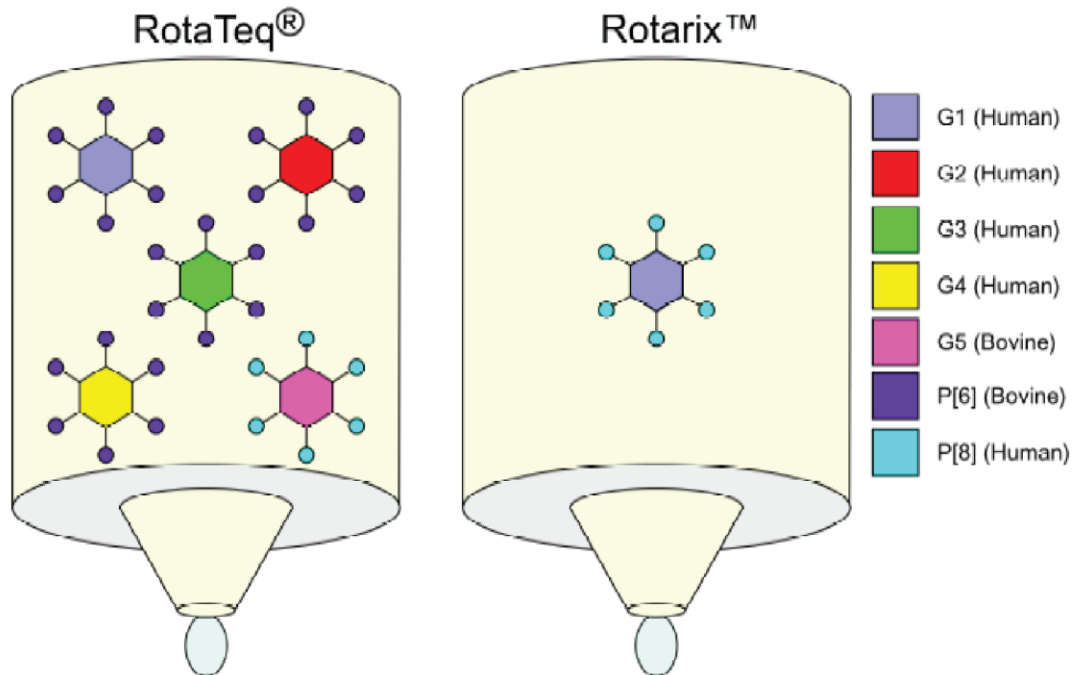


Figure 5: Conceptual diagram of the two currently licensed rotavirus vaccines. The oral dropper on the left represents RotaTeq, the pentavalent bovine-reassortant vaccine. The oral dropper on the right represents the monovalent live-attenuated vaccine, Rotarix. Hexagonal depictions of viruses are shown within the droppers, coloured with respective VP7 genotypes, with radiating circle symbols representing VP4 genotypes.

Due to the risks associated with RotaShield, both new vaccines underwent large clinical trials before licensure. RotaTeq had clinical trials in the United States of America and other developed countries, enlisting about 70,000 infants. Rotarix had clinical trials in Central and South America involving more than 60,000 infants. Both trials showed no significant increase in the occurrence of intussusception. Both trials also showed that the vaccines were highly efficacious in preventing rotavirus disease. RotaTeq showed 98% protection against severe rotavirus disease and a 63% reduction in hospitalizations when compared to the control group. One clinical trial of Rotarix showed 85% protection against severe rotavirus disease and a 42% reduction in hospitalizations when compared



to the control group [Dennehy. 2008]. In addition to protection against rotavirus, the areas of the trial experienced herd immunity. The CDC estimated a decrease of 50% in rotavirus disease. In addition there was a delay in the rotavirus season, for the seasons following the trials [Greenberg and Estes. 2009].

The vaccine trials also confirmed that strain specific antibodies did not provide the only source of protection against rotavirus disease. Both RotaTeq and Rotarix were fairly efficacious in protecting the trial group against all predominant circulating genotypes of virus, against severe rotavirus gastroenteritis, despite Rotarix only containing one serotype. The vaccines even showed protection against G9 strains, which are not part of the VP7 components of either vaccine [Greenberg and Estes. 2009]. An alternative European trial of Rotarix showed 79% efficacy against any severity of rotavirus disease and 90% efficacy against severe rotavirus disease. The vaccine also protected against severe rotavirus disease caused by G1P[8] with 96% efficacy and non-G1P[8] disease with an efficacy of 88%. Another trial of 20,000 infants showed that Rotarix was only 41% efficacious against G2 strains. This measure of G2 efficacy, however, was not significant [Dennehy. 2008].

Although seemingly very different, both of the currently licensed vaccines have unique benefits. RotaTeq is a 3 dose vaccine with a broad antigenicity and a high efficacy. Rotarix was developed as a 2 dose live-attenuated human strain vaccine. As Rotarix is based on a human strain, it replicates well in vaccine recipients resulting in possible shedding, which results in greater herd immunity. Rotarix can also be delivered at lower concentrations due to its ability to replicate [Dennehy. 2008, Greenberg and

Estes. 2009]. However, Rotarix should not be used in immunocompromised children with HIV or SCID, as it has been shown to cause rotavirus-related disease [Patel et al. 2010].

Both vaccines have proven to be effective in post-marketing surveillance. Post-vaccine introduction surveillance studies in countries with high rates of vaccine coverage have shown large decreases in incidences of rotavirus disease [Boom et al. 2010, Braeckman et al. 2011, Dennehy. 2008, Desai et al. 2010, Munos et al. 2010, Patel et al. 2011, Paulke-Korinek et al. 2011, Richardson et al. 2010]. In the United States of America the rotavirus vaccines were released in 2006 and recommended for regular use. It was estimated that in the first year of vaccine availability, 58% of 3 month old infants received 1 dose, while only 31% of those eligible received more than one dose of RotaTeq. Despite coverage rates, the decline in rotavirus incidents in the year following vaccination, 2007-2008, compared to the average of the incidents in the previous 6 rotavirus seasons, was 64%. The decline in rotavirus positive incidents indicated both vaccine success and herd immunity in the U.S.A. [Centers for Disease Control and Prevention (CDC). 2009].

In Austria, RotaTeq vaccine was released in 2007. Austria released the vaccine as a part of a universal mass vaccine program and found rotavirus-related hospitalizations decreased by 74% in eligible vaccine recipients [Tate et al. 2010]. Two years after the universal vaccination program was introduced, Austria continued to provide high coverage with the vaccines (74% in 2009). They realized a sharp decline in rotavirus incidents as well as some herd immunity [Paulke-Korinek et al. 2011]. Similarly, in Brazil, universal vaccination with Rotarix, starting in 2006, resulted in a reduction in rotavirus-related hospitalizations. In one city, which had more than 50% coverage by the

monovalent vaccine, all 21 rotavirus positive samples were G2P[4]. Although the emergence of a non-vaccine genotype provides evidence that the G1P[8] (Rotarix) vaccine caused an increase in G2P[4] prevalence, this may not be the case. Surveillance of surrounding unvaccinated countries indicated that, they too, experienced a spike in G2P[4] prevalence the year following vaccination in Brazil. It is possible that the G2P[4] emergence is related to natural fluctuation [Gurgel et al. 2007, Patel et al. 2008, Tate et al. 2010]. Finally, a country-wide study of vaccine effectiveness in Mexico saw a dramatic decrease in death rate due to usage of the Rotarix (RV1) vaccine. Although roughly half of the eligible infants were vaccinated, the country saw a reduction of 35% in the total annual death rate caused by rotavirus disease. This change was most dramatically seen in the youngest age category of less than 11 months of age, which realized a reduction in death rate of 41%. Some age categories were not vaccinated due to ineligibility, but realized a reduction in annual deaths, indicating that some herd immunity may have occurred in this study [Richardson et al. 2010].

In contrast to the success of vaccines in Europe and the Americas, Africa and South Asia have had less success with rotavirus vaccines. A study of vaccine efficacy in two African countries showed that Rotarix had a far lower efficacy in African nations. Rotarix was shown to have only 25.1% efficacy in Malawi and 44.1% in South Africa, against any rotavirus disease [Madhi et al. 2010]. In a presentation at the 11<sup>th</sup> Annual Conference on Vaccine Research, Roger Glass from the Fogarty Institute in California, stated that only 44% of children in South Africa and 55% of children in Bangladesh showed an immune response after given Rotarix [O'Brien. 2008]. Relatively low efficacy against all rotavirus disease was also shown in 5 developing nations using RotaTeq

[Breiman et al. 2012]. The low efficacy in developing African and South Asian countries is not well understood (Pentavalent Rotavirus vaccine in 5 sites). It has been suggested that poor nutrition and other interference by antibodies in breast milk may be involved in reduced efficacy. Since oral vaccines have also shown low success in Africa, Roger Glass suggested that there is something unique about the GI tracts of African children that may be the cause of the lack of efficacy of the Rotarix [Madhi et al. 2010, O'Brien. 2008].

#### 1.7.1 Vaccine Use in Canada:

The first of the two presently approved vaccines in Canada was RotaTeq in 2006. The release of RotaTeq was closely followed by that of Rotarix, which was approved for Canadians in 2008 [Salvadori and Le Saux. 2010]. Both licensures were approved based on safety and efficacy of each vaccine in the large clinical trials. Clinical trial data and previous Canadian rotavirus surveillance data led the National Advisory Committee on Immunization to recommend the usage of RotaTeq in 2008 [National Advisory Committee on Immunization (NACI). 2008b]. Assuming that a very large proportion (94%) of rotavirus gastroenteritis is caused by strains covered by the vaccine, vaccination has the potential to prevent approximately 56,000 incidents annually. NACI estimates the reduction in incidents would include around 15,000 emergency department visits and 5,000 hospitalizations [National Advisory Committee on Immunization (NACI). 2008b].

#### 1.7.2 Cost Effectiveness:

The cost effectiveness of rotavirus vaccine is difficult to estimate. Firstly, estimates of the burden of rotavirus disease vary. Second, the price of the vaccine is

inconsistent. In the United States of America the cost of the complete three-dose complement of the RotaTeq vaccine is around \$217.50, including administration [Widdowson et al. 2007]. In another scenario, reduced price of vaccines can be negotiated if a long-term vaccination program is established, such as in Australia [Newall et al. 2007]. Other countries, depending on economic status, may realize a substantial discount on vaccines from organizations such as GAVI (Global Alliance for Vaccines and Immunization) [Danchin and Bines. 2009, Tate et al. 2010]. A recent study in the U.S. estimates that, given the current price for the rotavirus vaccine and the current rotavirus-burden cost (\$893 million per year), the vaccine will not be cost effective [Widdowson et al. 2007], although the vaccine is paid for by insurers and not the government. An estimate of the rotavirus vaccine cost-effectiveness has not been undertaken in Canada. However, estimates of hospital and societal costs associated with rotavirus have been made [Le Saux et al. 2011].

### 1.8 Rotavirus Immunity:

Immunity to rotavirus is not completely understood. The initial RIT4327 vaccine showed some protection despite the lack of antigenic match to human strains. However, the lack of antigenic matching was thought to be the cause of the low efficacy of the initial RRV vaccine. Although some vaccines seem to protect against rotaviruses, including those with different VP7 genotypes than those in the vaccine, protection is thought to be highly dependant on the VP4 and VP7 proteins. Evidence for this relationship is provided by the RRV-TV vaccine: once human VP7 recombinants were

added to the RRV vaccine the efficacy of the vaccine was improved [Greenberg and Estes. 2009].

In 1983, a study assessed immunity induced in adults by VP4, VP7 and NSP4 proteins expressed using a baculovirus vector. The adults were challenged with a G1P[8] virus and an assessment of protective immunity was made. The group found that protective immunity was mostly established by homotypic antibody created prior to the challenge. Lower levels of protective immunity were induced by heterotypic antibodies [Yuan et al. 2009]. The evidence provided in the study by Yuan and colleagues emphasizes the need for genotype specificity in vaccine development and the need to survey for genotype reassortment events.

Other studies have assessed antigenic regions specifically. For instance, there have been several studies assessing particular VP7 antigenic sites with antibodies. Six epitopes for rotavirus VP7 protein were described by Kirkwood and colleagues and Chen and colleagues. Some protective antibodies are heterotypic while others are specific to the infecting strain's G genotype. It is important to consider the variability of the epitopic sites when assessing the potential effectiveness of the vaccines [Chen et al. 2009, Kirkwood et al. 1993]. Mutations in particular sites in the VP7 genome segment, for instance, could result in antibody escape of that strain, if the mutations cause an amino acid change in the antigenic site. Also, a particular antigenic site may affect the antigenicity of other antigenic sites. For instance, a previously described mutation in antigenic site C, on the VP7 protein, can affect the ability for antibody to bind site A, as they orient themselves near one another [Kirkwood et al. 1993, Yoshinaga et al. 2006]. If a mutation were to occur in a broadly reactive antigenic neutralization epitope, protective

antibody elicited by a vaccine could have reduced or no neutralizing capabilities [Yoshinaga et al. 2006].

Evidence of antigenic specificity effecting protective immunity means it is important to follow circulating virus genotypes and vaccine effectiveness to ensure that emerging strains do not escape the strain specific immunity created by the vaccine due to genetic shift (reassortment) or genetic drift (mutation).

### 1.9 Study Background:

Rotavirus is a major cause of morbidity in Canada. There is a substantial burden on the Canadian healthcare system and economy by rotavirus-infected children across the nation. Parents, clinics, hospitals and tertiary care facilities are burdened by the effects of rotavirus disease on Canadian children. Recently, two vaccines, RotaTeq and Rotarix, were licensed in Canada. The vaccines have been shown elsewhere to be highly efficacious against severe rotavirus disease [Dennehy. 2008].

Prior to the current study, there were no up-to-date Canadian rotavirus genotyping data. In fact, previous Canadian rotavirus studies took place several years prior to the licensure of the new vaccines, and only sampled one site over a single rotavirus season [Kostouros et al. 2003, Pang et al. 2004]. Rotavirus genotypes are known to fluctuate and different rotavirus genotypes are known to vary in different areas of the world [Santos and Hoshino. 2005]. The aim of the current study was to obtain a picture of rotaviruses in Canada, which would allow more clarity in vaccine policy development.

Estimated coverage of the vaccines in Canada is quite low [Dube et al. 2010]. Use of the highly efficacious vaccines may grow with increased funding and confidence,

provided by expert advice and evidence that the vaccine will be useful in the Canadian population. Recently, the Canadian body of experts on vaccines, the National Advisory Committee on Immunizations (NACI), recommended the pentavalent reassortant vaccine, RotaTeq, to be used by Canadian infants [National Advisory Committee on Immunization (NACI). 2008a]. Before implementation of the recommendation, it was important to establish that the recommended vaccine was well matched to the current population of rotaviruses in Canada. Provided there is a match between the vaccine strains and genotypes in circulating Canadian rotaviruses, there would be additional evidence that the vaccine would be effective in the Canadian population.

The purpose of the current study was to provide up-to-date genotypic, and characteristic information of currently circulating rotaviruses in Canada.

#### 1.10 Hypothesis:

I hypothesize that surveillance will reveal that currently circulating rotavirus strains in Canada are well matched to the currently licensed vaccines in Canada by analysis of VP4 and VP7 genotypes, genetic drift within genotypes and analysis of all circulating genotypes for abnormalities and reassortment events.

##### 1.10.1 Scientific Reasons for Approach Taken to Answer Hypothesis:

Analysis of VP7 and VP4 genotypes of all available Canadian rotavirus samples was important because the outer capsid proteins, VP7 and VP4, are targets for neutralizing antibodies in the host [de Rougemont et al. 2010]. This study related currently circulating Canadian VP7 and VP4 genotypes to those known to be protected



by currently available vaccines [Dennehy. 2008]. Further analysis, on a sub-group of Canadian samples of all genome segments and proteins via full open reading frame analysis, elucidated any abnormalities in Canadian rotaviruses. Abnormalities included those affecting normal pathogenesis, antigenicity or structural composition.

Abnormalities make Canadian rotaviruses unique compared to other circulating viruses. Unique viral composition may be implicated in future strain fitness or vaccine failure, as vaccines are based on old rotaviruses [Dennehy. 2008]. This sub-group of fully genotyped rotaviruses also served to analyze reassortment events in Canada.

Reassortment events have not been well studied, due to a lack of full genome sequencing studies [McDonald et al. 2009]. Creating constellations in phylogenetic analysis allowed in-depth analysis into the nature of reassortment in different, but related areas over a relatively short time-span. Analysis of Canadian VP4 and VP7 sequences was used as a quality control mechanism for commonly used genotyping methods. This ensured surveillance data from the current study and the future is reliable and assays are not in need of updating [Iturriza-Gomara et al. 2000, Iturriza-Gomara et al. 2004]. Finally, a real-time assay was developed as a faster, more sensitive alternative method to confirm rotavirus detection and genotyping results when compared to sequencing. A faster, more efficient confirmation method will be useful in future surveillance when determining and confirming genotypes.

#### 1.11 Objectives:

In the following study the objectives include establishing updated genotype prevalence rates of Canadian rotaviruses in order to assess the potential effectiveness of

the new rotavirus vaccines within the Canadian population. In addition to full rotavirus ORF genotyping, I intended to identify any unusual reassortment events that have occurred in circulating Canadian strains. This was used to analyze the nature of reassortment events in Canadian rotaviruses. Full ORF genotyping was also used to assess the validity of the currently used heminested multiplex genotyping assays. In addition, full ORF genotyping was used to monitor any genetic drift causing changes in important amino acid motifs, especially antigenic binding regions of the VP7 protein and functional abnormalities. Finally, a real-time PCR assay was established and validated for fast detection and establishment of lineage of Canadian rotaviruses.

## 2. Methods

### 2.1 Sample Collection:

Stool samples were collected from eight paediatric hospitals from July 24, 2007 until June 2, 2011. Three hundred and forty-eight samples were collected in total. Samples were shipped on dry ice from paediatric hospitals then stored at -20°C at the National Microbiology Laboratory (NML) until RNA extraction took place. Sample collection from the paediatric hospitals is summarized in Table A1.

### 2.2 Stool Clarification:

A pea sized amount (Approximately 0.1g) of stool was added to 0.9mL of Gibco 1X PBS pH 7.4 (Invitrogen, Carlsbad, Ca). The stool was vortexed vigorously for 1 minute or until the 10% stool suspension was fairly homogenous. If the stool solution could not be homogenized after vigorous vortexing, glass beads were added to break down more fibrous material. The suspended stool solution was centrifuged at 9000 RPM (roughly 8600 x g) for 10 minutes using an Eppendorf 5417R centrifuge (Eppendorf, Hamburg, Germany). The upper clarified stool solution was collected for extraction and storage for future use.

### 2.3 Genome Extraction:

RNA was extracted using the EasyMag extractor (bioMérieux, Marcy l'Etoile, France). To begin the extraction, 200µL of clarified stool was added to extraction buffer (bioMérieux, Marcy l'Etoile, France) in an extraction tray and incubated at room temperature for 10 minutes. Next, 100µL of magnetic silica beads were added to the

mixture and mixed thoroughly with the sample solution by pipetting up and down. The extraction was run using the Generic 2.0.1 method. The extraction resulted in the elution of 60 $\mu$ L of RNA solution.

#### 2.4 Rotavirus PCR screening:

Due to the double-stranded RNA nature of the rotavirus genome, a preliminary denaturing step was required for all PCR reactions. Sample tubes containing only 5 $\mu$ L of appropriate primers and 5 $\mu$ L rotavirus extracted RNA were placed into an Applied Biosystems Veriti Thermocycler (Applied Biosystems, Carlsbad, Ca) and denatured at 95°C for 5minutes before cooling to 4°C. This step was necessary to ensure that genomes denature and re-anneal to primers, while the reverse transcriptase enzyme is not deactivated at high temperature.

Following denaturation, the sample and primer solution was added to 40 $\mu$ L of One-step RT-PCR solution (Qiagen, Hilden, Germany). Final concentrations of the kit's reagents were as recommended by the manufacturer without using Q solution. The final primer concentration in every reaction was 600ng/ $\mu$ L for each primer. PCR products were analyzed with agarose gel electrophoresis on 1.0% agarose gels (Invitrogen) or with the Qiaxcel using the screening gel cartridge using the AM420 method (Qiagen). PCR cycling conditions and agarose gel band lengths for screening can be found in Table 2.

Samples were screened for rotavirus using the Qiagen One-step RT-PCR (Qiagen, Hilden, Germany) and previously described primer pairs Beg9/End9 and Con3/Con2 or VP7F/VP7R and VP4F/VP4R [Gentsch et al. 1992, Gomara et al. 2001, Gouvea et al. 1990, Simmonds et al. 2008].

## 2.5 Genotyping by Heminested Multiplex PCR:

Parent PCR amplicons, created with the Con3/Con2 or VP4F/VP4R for VP4 genotyping and VP7F/VP7R for VP7 genotyping, were further used for genotyping by heminested multiplex PCR. Five microlitres of screening reaction amplicon solution were added to forty-five microlitres of Platinum Taq PCR solution with the reagent concentrations recommended by the manufacturer. All appropriate nested primers and their complementary parent primers were also in the solution at a final concentration of 600ng/ $\mu$ L each. Heminested primers and cycling conditions can be found in Tables 1 and 2, respectively. Genotypes for samples were assigned based on amplicon size, as found in 1% agarose gel electrophoresis. Gels with faint bands or additional non-specific bands were retested for confirmation of contaminant or multiple genotypes.

Table 1: Heminested multiplex PCR primer sets. Con3/Con2 primers were used as the first VP4 testing set. VP4F/VP4R were used to type and genotype VP4 segments when Con3/Con2 failed. The G12 primer was added to the VP7 multiplex for 2010 samples.

Primer	Genotype	Sequence	Amplicon Size (bp)		Reference
Con3	Parent VP4 Primers	TGGCTTCGCCATTTTATAGACA	876		[Gentsch et al. 1992]
Con2		ATTCGGACCATTATAACC			
VP4F	Parent VP4 Primers	TATGCTCCAGTNAATTGG	663		[Simmonds et al. 2008]
VP4R		ATTGCATTTCTTCCATAATG			
1T-1	P[8]	TCTACTTGGATAACGTGC	345	224	[Gentsch et al. 1992]
1T-1DCDN	P[8]	TCTACTGGRTYRACRTG	345	224	This Paper
2T-1	P[4]	CTATTGTTAGAGGTTAGAGTC	483	362	[Gentsch et al. 1992]
3T-1	P[6]	TGTTGATTAGTTGGATTCAA	267	146	[Gentsch et al. 1992]
4T-1	P[9]	TGAGACATGCAATTGGAC	391	276	[Gentsch et al. 1992]
5T-1	P[10]	ATCATAGTTAGTAGTCGG	583	462	[Gentsch et al. 1992]
ND2	P[11]	AGCGAACTCACCAATCTG	122	NA	[Das et al. 1994]
VP7-F	Parent VP7 Primers	ATGTATGGTATTGAATATACCAC	881		[Iturriza-Gomara et al. 2004]
VP7-R		AACTTGCCACCATTTTTTCC			
aBT1	G1	CAAGTACTCAAATCAATGATGG	618		[Gouvea et al. 1990]
aCT2	G2	CAATGATATTAACACATTTTCTGTG	521		[Gouvea et al. 1990]
G3	G3	ACGAACTCAACACGAGAGG	682		[Iturriza-Gomara et al. 2004]
aDT4	G4	CGTTTCTGGTGAGGAGTTG	452		[Gouvea et al. 1990]
aAT8	G8	GTCACACCATTTGTAAATTCG	754		[Gouvea et al. 1990]
G9	G9	CTTGATGTGACTAYAAATTCG	179		[Iturriza-Gomara et al. 2004]
G10	G10	ATGTCAGACTACARATACTGG	266		[Iturriza-Gomara et al. 2004]
G12	G12	CCGATGGACGTAACGTTGTA	387		[Banerjee et al. 2007]

Table 2: Below are thermocycling conditions for all RT-PCR reactions in this study. Parts 1 and 2 are carried out for both multiplex reactions and full genome sequencing amplification, while Part 3 is only relevant for heminested multiplex PCR.

Stage	Number of Cycles	Time	Temperature (°C)
<b>Part 1: Initial Stage for Template Denaturation and Annealing to “Parent” Primers</b>			
Denaturation (Primers and Template only)	1	5 minutes	95
Cool Stage	1	-	4
<b>Part 2: Add Template/Primer Mixture to Qiagen One-Step Enzyme Mix</b>			
Reverse Transcriptase	1	30 minutes	50
Denaturation/ Taq Activation	1	5 minutes	95
Denaturation	40	40 seconds	94
Primer Annealing		50 seconds	Ta (See Appropriate Primer Table)
Extension		60 seconds	68
Final Extension	1	10 minutes	68
Cool Stage	1	-	4
<b>Part 3: Add “Parent” Amplicons from Part 2 to the Invitrogen Platinum Taq mixture and the Corresponding Multiplex of Primers/End Primer</b>			
Taq Activation	1	5 minutes	95
Denaturation	40	40 seconds	94
Primer Annealing		50 seconds	42
Extension		60 seconds	68
Final Extension	1	10 minutes	68
Cool Stage	1	-	4

## 2.6 Genotype Confirmation:

All genotypes were confirmed by sequencing the screening amplicons. Amplified DNA was first purified using the Millipore Montage DNA purification spin filter system (Millipore, Billerica, Ma) according to the manufacturer’s instructions. Samples with both positive screening bands and non-specific bands were purified using the Minelute Gel Extraction Kit (Qiagen) according to the manufacturer’s instructions. Concentrations of the samples were then measured with a NanoDrop 1000 Spectrophotometer (ThermoFisher Scientific, Waltham, Ma). If necessary, samples were diluted to 50ng/μL

and sequenced at the Genomics Core Facility at the National Microbiology Laboratory. The genomics core utilizes an ABI Prism 3100 genetic analyzer (Applied Biosystems) and BigDye v3.1 Terminator chemistry.

Sequences were assembled with the SeqMan software (Lasergene Version 7.2.1(1), 410, DNASTar Inc.). Sequences were exported as .fas files and aligned using MEGA 5.0 software [Tamura et al. 2011]. Confirmatory phylogenetic trees were created using ORFs of VP7 and partial VP4 genome segments. High identity to the type strains confirmed the genotype.

#### 2.7 Positive Control Proliferation:

For the genotyping assays the positive controls used were Wa, DS-1, P, ST-3 and US1205. These positive controls were obtained from Jon Gentsch of the Centers for Disease Control, Atlanta, Georgia. The controls were received as MA104 cell lysate. Propagation of virus control was done by infecting MA104 cells with virus.

MA104 cells were grown and maintained in Gibco Medium 199 (Gibco, Carlsbad, Ca) containing 5% Hyclone Fetal Bovine Serum (Thermo Fisher Scientific, Waltham, Ma), 1X Gibco Glutamax (Gibco) and Gibco penicillin (100 units/mL) streptomycin (100uL/mL) (PenStrep) solution (Gibco). The cells were incubated at 37°C with 5% CO<sub>2</sub>.

Cells were maintained and propagated for infection by regular splitting of cells every 4-7 days. Next, cell monolayers were washed with 5mL Gibco 1X Dulbecco's Phosphate Buffered Saline (DPBS) solution without calcium chloride or magnesium chloride (Gibco). Cells were detached from the flask by adding 1.5mL of Gibco 0.25%



Trypsin-EDTA solution to the flask and incubating at 37°C with 5% CO<sub>2</sub> for 2-4 minutes, or until cells were visibly detached. Cells were resuspended with 8.5mL of maintenance solution and split in small aliquots to flasks containing fresh maintenance media.

For inoculations, 100% confluent MA104 cell layers, in Corning T75 (75cm<sup>2</sup>) polystyrene canted neck flasks, (Corning Inc, Corning, NY) were first washed with 5mL Gibco 1X DPBS solution without calcium chloride or magnesium chloride (Gibco). Next, 15mL serum-free Gibco Medium 199 with 1X Gibco Glutamax (Gibco) and Gibco PenStrep solution (Gibco) were added to the flask. The flasks were then incubated at 37°C with 5% CO<sub>2</sub>. The following day 200µL of rotavirus positive control-infected cell lysate were activated by adding 2mg/mL trypsin solution (Type IX-S, lyophilized powder, 13,000-20,000 BAEE units/mg protein (Sigma, Steinheim, Germany)) dissolved in PBS pH 7.4, (Invitrogen, Carlsbad, Ca)) to a final concentration of 10µg/mL. The trypsin/virus solution was mixed at 37°C at 300rpm in an Eppendorf thermomixer (Eppendorf, Hamburg, Germany) for 1 hour. Next, the uninfected cell monolayers were washed with 5mL serum-free medium 199. Subsequently, the activated virus solution was diluted with serum-free Medium 199, so that 1.5mL of solution was available for inoculation. Finally, the activated virus solution was added to the washed T75 flask containing uninfected MA104 cells and swirled in the flask to coat the monolayer. The infected flasks were incubated at 37°C for 1 hour, with swirling of the flask every 5-10 minutes. Twelve millilitres of serum-free Medium 199 with 1X Gibco Glutamax (Gibco), Gibco PenStrep solution (Gibco) and 1.0 µg/mL trypsin were added to increase medium in the flask.

Cells were monitored for cytopathic effects (CPE) every 1-3 days. A mock-infected control was used as a comparison point to determine CPE effects due to rotavirus infection in other flasks. Collection of rotavirus infected cellular lysate was accomplished after three freeze/thaw cycles in a -82°C freezer. The lysate in the flask was collected in a 50mL Fisherbrand polypropylene centrifuge tube (Thermo Fisher Scientific) and centrifuged at 400 X g for 10 minutes in a Thermo Scientific Heraeus 3SR+ centrifuge (Thermo Fisher). Lysates were collected, frozen, then extracted for genotyping.

## 2.8 Full Genome Sequencing:

Samples for full genome sequencing were selected based on several criteria. Firstly, the chosen samples should have represented the various genotype combinations found in Canada in every study site during all rotavirus seasons screened by the spring of 2011. Secondly, the samples were chosen based on amount of sample available, as many samples lacked enough sample or were exhausted before full genotyping could be completed. Thirdly, some samples were chosen to examine potential diversity within a genotype from a particular site in a specific rotavirus season. This would allow an analysis of diversity within a genotype. Finally, sequencing analysis was undertaken on samples with limited identity to other strains chosen. Co-infected samples should be excluded as genotyping may be skewed by a primer set's affinity to one genotype in a sample over another, then vice-versa with subsequent primer sets. Based on these criteria, 84 samples were chosen for full genome sequencing.

### 2.8.1 Primer Development:

Samples were fully genotyped using specific primer sets to amplify the open reading frames of all 11 segments. The ClustalW alignment algorithm was used in Megalign software (Lasergene Version 7.2.1(1), 410, DNASTar Inc.) to create consensus sequences based on the alignment of the type strains Wa, Ku, DS-1, TB-Chen, AU-1, L26, ST-3, RMC321, 69M, IAL28, B1711, US1205, RV176-00, T152, Matlab13-03, P and 116E [Both et al. 1984, Chen et al. 2008, Cook and McCrae. 2004, Cunliffe et al. 1997, Das et al. 2004, Ernst and Duhl. 1989, Gentsch et al. 1993, Gorziglia et al. 1988, Heiman et al. 2008, Horie et al. 1997, Hua et al. 1993, Isegawa et al. 1992, Ito et al. 1995, Kirkwood and Palombo. 1997, Kojima et al. 1996, Mason et al. 1985, Matthijnssens et al. 2008a, Matthijnssens et al. 2008b, Mohan and Atreya. 2001, Mulherin et al. 2008, Nakagomi and Kaga. 1995, Okada et al. 1984, Padilla-Noriega et al. 1995, Pongsuwanna et al. 2002]. As the majority of the RotaTeq vaccine contains bovine sequences, but the VP4 and VP7 components are human, both vaccines were included in VP4 and VP7 alignments only [Matthijnssens et al. 2010b, Zeller et al. 2011]. Accession numbers and genotypes for these segments can be found in Tables A2 and A3, respectively. Next, using the program PrimerSelect (Lasergene Version 7.2.1(1), 410, DNASTar Inc.), primers were developed based on the requirements of the genome segment. Primers were chosen based on their segment size, their location, their PrimerSelect (Lasergene Version 7.2.1(1), 410, DNASTar Inc.) quality score and the genetic stability of the target sequence throughout genotypes. Although many primers in this study were developed using new alignments, some were previously developed in other studies [Gouvea et al. 1990, Kudo et al. 2001, Matthijnssens et al. 2008b, Tsugawa and Hoshino. 2008]. A complete list of

primers and their references can be found in Table A4. As the end portions of many of the rotavirus genome segments are highly conserved, primers used in other studies were often best suited to genotype ends beyond ORFS. As degenerate primers are less specific and have a lower single primer concentration, degeneracy was avoided if possible. To avoid primer bias, targets for amplification overlapped.

After initial primer development, primers were tested on a subset of samples for validation. The subset of samples included rotavirus positive samples of varying P and G genotypes. Some primers were removed, replaced or adjusted as required to better suit the sample set. Cycling conditions were the same as used previously for the screening RT-PCR, however the annealing temperature step was adjusted to suit the primer set, based on the PrimerSelect (Lasergene Version 7.2.1(1), 410, DNASTar Inc.) software's recommendations. Annealing temperatures were adjusted after the subset was tested to optimize the PCR. Annealing temperature was increased slightly if there were non-specific PCR products, and reduced if the segment did not produce a strong band during gel electrophoresis.

After sequencing, contigs were assembled using the SeqMan software (Lasergene Version 7.2.1(1), 410, DNASTar Inc.). Sequences were corrected and trimmed so that only the appropriate ORF sequence was left for each segment. Any sequences indicating degeneracy due to quasispecies in a sequence were double checked by repeat sequencing or overlapping sequences of the same region. If other reads indicated a lack of degeneracy, they were used to assemble the sequence in the given region.

### 2.8.2 Genome Alignment:

Completed and trimmed segments were exported to the Mega 5.0 program [Tamura et al. 2011], where an alignment was made. The ORFs of Wa, Ku, DS-1, AU-1, B1711, Matlab 13-03, L26, RMC321, 116E, 69M, IAL28, TB-Chen, T152, RV176-00, P, ST-3 and PO-13 were used as type strains and outgroups for the data for all genome segment alignments. The alignment parameters for the ClustalW alignment algorithm in MEGA 5.0 [Tamura et al. 2011] were as recommended by *Phylogenetic Trees Made Easy* [Hall. 2007]. First, ORFs were translated into amino acid sequences to ensure any gaps created did not simulate a frameshift mutation. Next, the Clustal W algorithm was used with a gap penalty of 10, a gap extension penalty of 0.1, a gap opening penalty of 3.0 and a gap extension penalty of 1.8. After the algorithm aligned the sequences, the data was converted back to nucleotide form and exported as a .meg file for further analysis.

### 2.8.3 Analysis of Open Reading Frames:

Canadian rotavirus alignments were analyzed for unusual motifs or anomalies in amino acid sequences based upon literature. Motifs were analyzed using the MEGA 5.0 alignments described previously. Motifs analyzed included those necessary for protein function, those for protein-protein interaction, those enabling proper protein secondary structure and those responsible for antigenicity. Motifs analyzed are shown in Table 3. In addition, primers for the heminested multiplex PCR assay were assessed to ensure they would be suitable for future genotyping studies. Primers assessed included those in Table 4.

#### 2.8.4 Phylogenetic Analysis:

Phylogenetic trees were created using the Neighbour-Joining method with 2000 bootstraps. Statistically speaking, a tree made using 2000 bootstraps produces a tree with 95% confidence. At this point segments must be looked at individually. A tree made using isolate segments composed of two or more concatenated segments will be invalid. An assumption, made at this point, is that few if any recombination events occur within a given segment [Hall and Barlow. 2006]. Trees were rooted with the PO-13 avian rotavirus strain as an outgroup. PO-13 differs greatly from human strains and contains avian genome segments with no human recombinants [Matthijssens et al. 2008a].

Table 3: Motifs analysed in full genome analysis

Gene	Region	Approximate Amino Acid Position	Ref.
VP1	RNA Dependant RNA Polymerase Domain	<b>Motif A</b> (517-531), <b>Motif B</b> (584-611), <b>Motif C</b> (624-640), <b>Motif D</b> (662-680), <b>Motif F</b> (454-469)	1
VP2	RNA Binding/ Viral Protein Interacting Domain	1-156	2, 3
VP3:	GTP/S-adenosyl-L-methionine-binding region	447-472	2
	Structural Proline Residues	P120, P211, P223, P289, P309, P313, P392, P401, P426, P502, P652, P697, P709, P717, P796	4
	Structural Cysteine Residues	C144, C380, C381, C659	4
	Casein Kinase II	383-388	4
	Putative Guanyltransferase Domain	541-545	4
VP4	Structural Cysteine Residues	C203, C317, C380, C768	2, 5
	Structural Proline Residues	P68, P71, P225, P226	2
	Fusion Domain	384-402	2
	Protein Cleavage Site (Activation Domain)	R231, R241, R247	6
	Antigenic Binding Sites	<b>Motif 8-1</b> (108, 146, 148, 150, 188, 190, 192, 194, 195, 196), <b>Motif 8-2</b> (180, 183), <b>Motif 8-3</b> (113-116, 125, 131-133, 135), <b>Motif 8-4</b> (87-89), <b>Motif 5-1</b> (384, 386, 398, 393, 394, 398, 440, 441) <b>Motif 5-2</b> (434), <b>Motif 5-3</b> (459), <b>Motif 5-4</b> (429), <b>Motif 5-5</b> (306)	7
	Integrin Binding Domain	306-309	8
VP6	Trimerization Domain	246-314	9
	Protein-Protein Interacting Domains	<b>Domain B</b> (1-150, 326-397) <b>Domain H</b> (151-325)	3, 9
VP7	Antigenic Binding Sites	<b>Motif 7-1a</b> (87, 91, 94, 96-100, 104, 123, 125, 129, 130, 291) <b>Motif 7-1b</b> (201, 211-213, 238, 242) <b>Motif 7-2</b> (143, 145-148, 190, 217, 221, 264)	7
	Integrin Binding Domains	253-255, 238-240	8
NSP1	RNA Binding Motif	43-76	10
NSP2	RNA Binding Motif	205-241	5
	Structural Cysteine Residues	C6, C8, C85, C285	5
	Putative HIT-like Motif (NTPase)	221-227	11
	Mg <sup>2+</sup> Binding Sites	153, 171	11
NSP3	RNA Binding Motif	104-112	12
	Structural Cysteine Residues	C123, C139, C306, C314	12
	Oligomerization Domain	181-236	12
NSP4	Potential Glycosylation Sites	63-71, 93-101	13
	Hydrophobic Domains	<b>H1</b> (7-21), <b>H2</b> (28-47), <b>H3</b> (67-85)	14
	Oligomerization Domain	86-106	14
	VP4 Binding Domain	112-148	14
	Enterotoxic peptide and membrane-destabilization domain	114-135	14
	Inner-capsid particle binding domain	156-175	14
NSP5	Multimerization Core Motif	187-197	15
	Cysteine Structural Residues	170, 173	15

1. Vasquez et al 2006 2. Chen et al. 2009 3. Heiman et al. 2008 4. Cook and McCrae et al. 2004 5. Patton et al 1993 6. Gilbert et al. 1998 7. Zeller et al 2011 8. Coulson et al 1997 9. Affranchino et al 1997 10. Hua et al 1993 11. Vasquel-Del Carpio et al 2006(2) 12. Rao et al 1995 13. Okada et al 1984 14. Horie et al 1997 15. Torres-Vega et al 2000.

Table 4: Primers were assessed by full genome segment sequence analysis for matching target sites, as well as, those differing from other genotypes. References can be found in Table 1.

<b>Primer Name</b>	<b>Segment/Sense (Position)</b>	<b>Sequence (5'-3')</b>
Con-3	VP4/+ (11-32)	TGG CTT CGC CAT TTT ATA GAC A
Con-2	VP4/- (868-887)	ATT TCG GAC CAT TTA TAA CC
VP4-F	VP4/+ (132-149)	TAT GCT CCA GTN AAT TGG
VP4-R	VP4/- (775-795)	ATT GCA TTT CTT TCC ATA ATG
1T1DCDN (P[8])	VP4/- (339-356)	TCT ACT GGR TYR ACR TG
2T1 (P[4])	VP4/- (474-494)	CTA TTG TTA GAG GTT AGA GTC
VP7-F	VP7/+ (51-71)	ATG TAT GGT ATT GAA TAT ACC AC
VP7-R	VP7/- (914-932)	AAC TTG CCA CCA TTT TTT CC
aBT1 (G1)	VP7/+ (314-335)	CAA GTA CTC AAA TCA ATG ATG G
aCT2 (G2)	VP7/+ (411-435)	CAA TGA TAT TAA CAC ATT TTC TGT G
G3	VP7/+ (250-269)	ACG AAC TCA ACA CGA GAG G
aDT4 (G4)	VP7/+ (480-499)	CGT TTC TGG TGA GGA GTT G
G9	VP7/+ (757-776)	CTT GAT GTG ACT AYA AAT AC

## 2.9 Real-Time Assay Development and Validation:

A real-time assay was developed for detection of rotavirus. Real-time PCR is a quantitative technique, measuring amplification after each cycle; therefore, this technique is good for fast typing. The assay was developed based on results of the full-genotyping assay. Analysis of DS-1 like strains indicates that VP2 genotype does not significantly differ from all other genome segments (including VP7). There is a need for a real-time assay based on a genetically stable structural protein coding genome segment. The basis for choosing a longer segment is that it would better represent viral load as it would replicate and transcribe more slowly and possibly less efficiently than shorter genome segments, assuming RNA-dependant RNA-polymerase replicates all segments at equal rates. Also, VP1, VP3, VP4, VP6 and VP7 are fairly variable in comparison to VP2; therefore, VP2 was chosen as the gene to develop the real-time PCR assay. Also, the new real-time assay was developed with current Canadian strain detection in mind. Therefore,



the sequence data, used to develop the assay, was based on currently circulating Canadian rotaviruses from this study. The 3' end of the coding strain was preferred as that would be the portion of the segment transcribed last.

Development of the VP2 real-time assay began with a multiple alignment of the VP2 genome segments in all fully sequenced isolates. The alignment was created using Mega 5.0 [Tamura et al. 2011] with the Clustal W algorithm. A consensus sequence was exported to PrimerSelect (Lasergene Version 7.2.1(1), 410, DNASTar Inc.) to develop several sets of primer pairs targeting sequences 80-200 base pairs long. Three primer sets were analyzed on the 3' end of the VP2 genome segment. A section of consensus and fairly stable genome was sought out by analysing strains for variation at the potential primer target sites. Primers were developed in genetically stable regions. Also, single probes for both Wa-like and DS-1-like lineages were developed in genotype specific regions. Next, the developed TaqMan probes for every primer set were manually entered into PrimerSelect. Probes and primers were checked for compatibility. The primer or probe sequences were analyzed for hairpin loops, self-dimerization, primer-primer dimerization, primer-probe dimerization and probe-probe dimerization. One suitable assay of primers and probes was found. Primer pairs were ordered and tested before probes were ordered. Sequencing of the primer pair target region showed the correct region was amplified and agarose gel data showed that the primers were highly specific. Primers were ordered from Genomics Core (National Microbiology Laboratory, Winnipeg, MB). Probes were ordered from Integrated DNA Technologies (IDT, Coralville, IA) with the sequences found in Table 5.

Standard curves were developed for each probe. To develop standard curve RNA, two Canadian samples were amplified with the VP2F4T7 primer (Integrated DNA Technologies, Coralville, IA) and the VP2 R primer. Samples from the two lineages, Wa-like and DS-1 like (synonymous with C1 and C2, respectively), were used so that a standard curve could be made for each. The T7 region of the VP2F4T7 primer added a T7 promoter region to DNA amplicons of the previously sequenced samples RT006-07 (G1P[8]) and RT036-07 (G2P[4]). The amplified sequences were confirmed by sequencing by genomics core (National Microbiology Laboratory, Winnipeg, MB). Amplicon DNA was used to transcribe RNA with the MEGAscript kit (Ambion, Austin, TX), using the T7 promoter, according to manufacturer's instructions. Excess DNA was removed from the transcribed RNA by adding 1µL TurboDNase (Ambion) and incubating for 10 minutes at 37°C. RNA for the standard curves was measured for RNA concentration using the Qubit (Invitrogen, Carlsbad, CA), according to manufacturer's instructions. Standard template copies per microlitre were determined with the following equation:

$$\text{Templates}/\mu\text{L} = \{([\text{sample (ng}/\mu\text{L})] \cdot \text{Qubit dilution factor}) / (\text{no. bases} \cdot \text{avg. MW of a nucleotide})\} \cdot N_A$$

, where the dilution factor is 20,  $N_A$  is Avagadro's number ( $6.02 \times 10^{23}$ ) and the average molecular weight of a nucleotide is 330.0 g/mol. The final size of the amplicon was 586 base pairs for both 06-07 and 36-07. Amplicon sizes were confirmed with agarose gel electrophoresis in the Qiaxcel (Qiagen, Hilden, Germany).

Table 5: VP2 Real-time assay primers and probes. The position is based on the Wa type strain for consistency. The bottom two primers were used to create RNA amplicons to be used in standard curves for C1 and C2 assays.

Primer/Probe Name	Position (Wa)	Sense	Sequence (5'-3')
VP2RE-F3	2183-2204	+	GAAAGAGATGAGATGTAYGGAT
VP2RE-R3	2348-2367	-	AGCGATATAACTGATGARTC
VP2Probe C1	2258-2282	+	6-FAM/TTAATGAGA/ZEN/ ACTGGAGACTATGGTC/IABkFQ
VP2Probe C2	2223-2246	-	HEX/TTATCTGTT/ZEN/ GAAATCCCTCTAAAT/IABkFQ
VP2F4AMT7	2131-2150	+	TAATACGACTCACTATAGGTTTCAGATAA AATTGCTCAAG
VP2-R-04 <sup>1</sup>	2699-2717	-	GGTCATATCTCCACA(G)TGG

<sup>1</sup> Reference: [Tsugawa and Hoshino. 2008]

Prior to running the real-time assay extract RNA was prepared in a denaturing step, as described earlier in the conventional PCR setup. The 10µL RNA extract and primer solution was added to 15µL One-Step RT-PCR reaction mixture (Qiagen, Hilden, Germany). The One-Step RT-PCR kit was mixed according to manufacturers' recommendations. The final concentrations of probes and primers in the mixture were 250 nM and 900 nM, respectively. Samples were mixed well with the RT-PCR solution using an Eppendorf Centrifuge 5430 (Eppendorf, Hamburg, Germany) to spin the fast real-time plate for 10 minutes at 1500rpm. The plate was then added to the ABI 7500 fast real-time machine (Applied Biosystems, Carlsbad, CA). A 30-minute 50°C reverse-transcriptase step and a 15-minute 95°C polymerase activation step preceded the cycling stage. Real-time data was collected in 45 cycles of 94°C for 15 seconds followed by 51°C for 30 seconds then 72°C for 30 seconds. Ramp speeds were set to standard default settings.

Ambion Ag-Path master mix (Ambion, Austin, TX) was also used, according to manufacturer's instructions, to ensure the assay was robust enough to use different

enzymes and a fast one-step enzyme. Ramp speeds were set to fast real-time default settings.

The assay was validated with 195 previously characterised samples. Two Canadian rotavirus strains RT018-07 (G1P[8]) and RT068-07 (G2P[4]) were used as positive controls at approximate concentrations of  $5 \times 10^5$  copies per  $\mu\text{L}$  and  $5 \times 10^7$  copies per  $\mu\text{L}$ , respectively. Positive controls used obtained consistent signal values through many runs, further validated the real-time assay. Enzymes with different lot numbers were also used to ensure consistency in the assay in different batches of the same enzyme.

#### 2.9.1 Cell Culture Propagation for Canadian Rotavirus Controls

Two Canadian rotavirus cultures were developed for use in the real-time assay as positive controls. Samples were chosen based on genogroup, sample availability and whether they were fully sequenced. The samples chosen were RT018-07 and RT068-07. Both samples had previously been fully genotyped, thereby ensuring a good positive control, and standards of monitoring changes.

Infecting MA104 cell monolayers was done as previously described, except clarified stool of selected samples was used in place of infected MA104 cell lysate. Multiple passages were required for viruses to adapt to cell culture. Passages were undertaken by collecting cell lysates after freeze thaw cycles, as previously described, and reapplying 2mL of the lysate to a fresh MA104 cell monolayer in a T75 flask. The infection of the fresh monolayer was incubated at 37°C for 1 hour with swirling every 5-10 minutes, then topped up with 8mL of fresh serum free medium 199 with antibiotics

and 2 ng/ $\mu$ L trypsin. Upon completion of 3-5 passages, lysate from all passages was extracted using the EasyMag extraction method, as previously described. Positive control passage lysate extracted RNA was tested for rotavirus using conventional rotavirus PCR and then quantified using the real-time PCR assay.

### 3. Results

#### 3.1 Heminested Multiplex PCR Genotyping

A total of eight different paediatric centres submitted samples, shown in Table A1. Of the 348 samples tested from the 2007-2010 rotavirus seasons, 259 were successfully genotyped for rotavirus. All 259 samples were confirmed by sequencing.

Samples submitted from the 2007-2009 seasons were pre-screened for rotavirus by enzyme immunoassay (EIA) or EM before submission to the NML. Of the 189 samples that pre-screened positive for rotavirus, 187 were genotyped successfully by heminested multiplex PCR. Two samples were pre-screened positive for rotavirus but could not be genotyped due to exhaustion of the sample. There were also 34 samples that were pre-screened negative for rotavirus, but were successfully genotyped by heminested multiplex PCR and confirmed by partial genome sequencing of the VP7 and VP4 genome segments. Fifty of the fifty-seven 2010 samples submitted were genotyped by heminested multiplex PCR and confirmed by genotyping.

In total, there were 174 G1P[8] strains detected, 22 G2P[4] strains detected, 45 G3P[8] strains detected, 3 G4P[8] strains detected, 13 G9P[8] strains detected and 2 other genotype strains detected (Table A5). The other genotype category contained two G9P[4] strains from 2010, but no other genotypes. Raw sample data is shown in Table A5. Multiple genotype samples were not included in the genotyping data as the number of rotavirus strains could not be determined with certainty, due to the possibility of unusual reassortants occurring within the sample.

There were 12 samples with multiple genotypes detected by heminested multiplex PCR from 2007-2010. Samples with multiple genotypes were most often found with both

P[8] and P[4] genotypes. Samples with multiple genotypes also included those with both G1 and G2, G1 and G3, and G1 and G9.

Temporally, the prevalence of rotavirus changed slightly from year to year, but G1P[8] strains were always the most prevalent strain. Often, there were drastic fluctuations in prevalence, such as between the 2008 and 2009 rotavirus seasons, when the frequency of G1P[8] rotaviruses dropped from 86.7% to 41.4%. The 2009 drop in G1P[8] corresponded with an increase in G3P[8] prevalence and an emergence in G9P[8] strains, which were detected in 31.4% and 17.1% of samples, respectively. Fluctuations in genotype prevalence are graphically represented in Figure 6.

G2P[4] samples were detected in this study steadily at between 6.7% and 12.0% prevalence annually. There were no drastic changes in G2P[4] prevalence while all other G types experienced noticeable year-to-year fluctuations, when considering some strains were not detected in some years but detected in others.

Only 3 G4P[8] strains were detected in the course of the study. The samples were detected in 2007 and 2009 in Halifax and Ottawa, respectively. G9P[8] strains emerged substantially in 2009. However, the emergence occurred in Ottawa samples only and was limited to 2009. G9 strains were also found in 2007 and 2010 in Halifax and Saskatoon, but only one sample was found in 2007 and two in 2010.

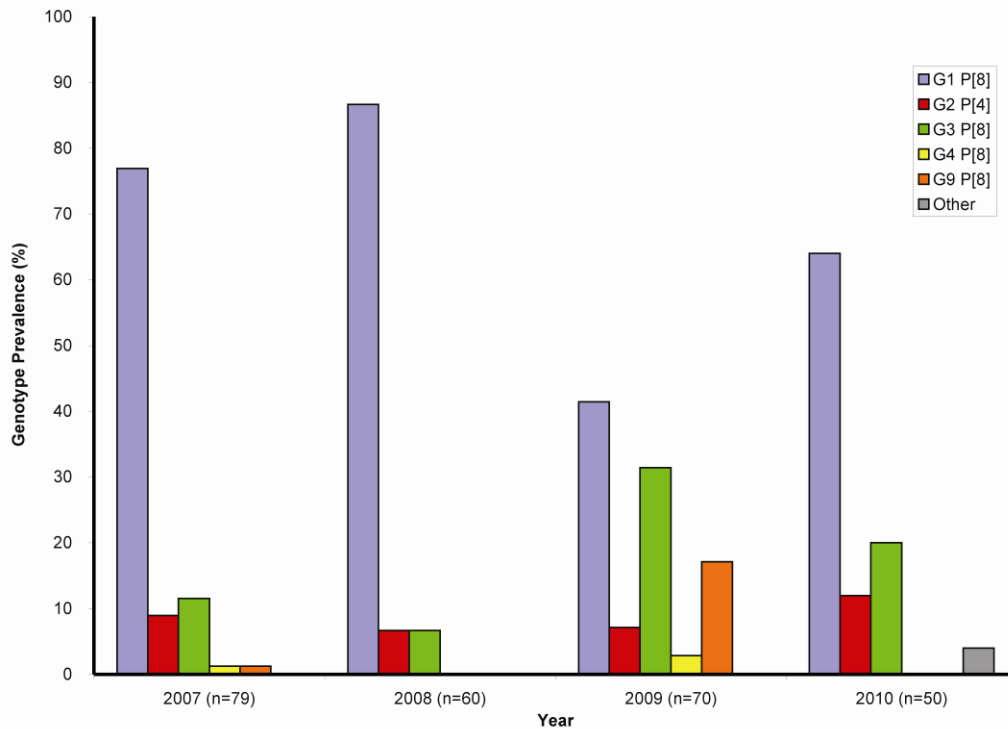


Figure 6: Bar graph showing genotype prevalence over the four rotavirus seasons of Canadian surveillance in this study.

Geographically, the most prevalent samples from all cities G1P[8] the most (Figure 7). Prevalence of other strains varied but throughout all years of the study G1P[8] was the predominant strain in all surveillance sites in all years except for Halifax in 2009. Only 3 G1P[8] samples were detected in Halifax in 2009 out of a total of 22. Most specimens from Halifax in 2009, of which 17 were detected, were G3P[8]. All sites, except Toronto, detected G2P[4] strains. G3 strains were found in all sites except for Toronto and Edmonton. G4 strains were only detected in two sites, while G9 strains were only detected in three.



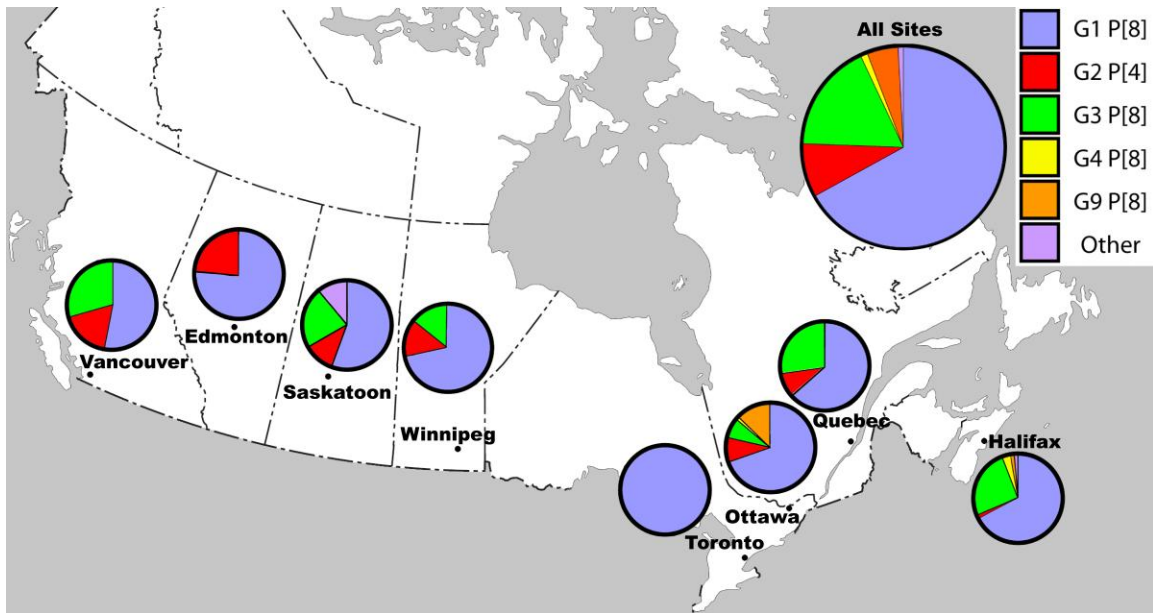


Figure 7: Map showing the eight Canadian sites in this study. Raw data for the table can be found in Table A5. A cumulative pie chart of all genotype data collected in the surveillance study is shown in the top right corner of the map.

### 3.2 Full Genome Analysis

#### 3.2.1 Analysis of Reassortment:

Full ORF sequencing of 84 rotavirus strains revealed that rotavirus reassortant strains only exist for the VP7 genome segment. Canadian rotavirus isolates with five different VP7 genotypes can be placed into one of two genogroups, which highly correlate with genotypes in the other 10 genome segments. Other genome segments associated with G1, G3, G4 and G9 rotavirus strains are almost always genogroup II, or Wa-like genogroup (includes the VP4 genotype P[8]) [Matthijssens et al. 2008a, Matthijssens et al. 2008b]. Genotypes and accession numbers for the 11 genome segments of all 84 fully sequenced rotaviruses are found in Table 6 and A6, respectively. All G2 strains are associated with segments that can be placed into genogroup I or DS-1-like (includes the VP4 genotype P[4]) [Matthijssens et al. 2008a, Matthijssens et al. 2008b]. The lack of reassortment amongst alternative genotypes was confirmed by

statistical analysis comparing VP2 genome segments with all other non-VP7 genome segments. McNemar's test statistical analyses (contingency Table shown in Tables A7 and A8) indicate that there are two genogroups of rotavirus circulating in Canada, DS-1 like and Wa-like. In all Canadian strains the VP2 genotype does not differ from the genotypes of all other non-VP7 genotypes for genogroup II and I.

All non-VP7 segments were confirmed to be in the Wa-like or DS-1-like genogroup using previously described cutoff values of nucleotide identity. VP7 genotypes were assessed in the same way but there were more genotypes to consider [Matthijssens et al. 2008a].

The two main genogroups were further classified into subgenotypic groups when distinct subgenotypic clusters were formed by Canadian strains. Appropriate cut-off values for sub-genotypic clusters were evaluated on a segment to segment basis (described below). As there are only eight genogroup I strains in every genome segment, only one cluster was identified for most genome segments. However, phylogenetic analyses of a few segments show a degree of genogroup I divergence meriting subgroups.

Table 6: Genotypes as defined by open reading frame identities of the 84 fully genotyped Canadian samples. Genotypes are colour coded by genotype. Genogroup I (DS-1) strains are coloured in red, the rest are genogroup II (Wa) strains.

Sample	Testing Site	VP7	VP4	VP6	VP1	VP2	VP3	NSP1	NSP2	NSP3	NSP4	NSP5	Year
002-07	Quebec City	G3	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
004-07	Quebec City	G3	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
005-07	Quebec City	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
006-07	Quebec City	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
008-07	Quebec City	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2	2007
012-07	Quebec City	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
013-07	Quebec City	G3	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
015-07	Quebec City	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
016-07	Quebec City	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
018-07	Quebec City	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
019-07	Quebec City	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
025-07	Quebec City	G3	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
028-07	Halifax	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
031-07	Halifax	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
032-07	Halifax	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
033-07	Halifax	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
034-07	Halifax	G9	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
036-07	Halifax	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2	2007
037-07	Halifax	G3	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
048-07	Ottawa	G4	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
049-07	Ottawa	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
054-07	Ottawa	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
055-07	Ottawa	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
059-07	Ottawa	G3	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
061-07	Ottawa	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
064-07	Ottawa	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
065-07	Ottawa	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
066-07	Ottawa	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
068-07	Edmonton	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2	2007
072-07	Edmonton	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
074-07	Edmonton	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
077-07	Edmonton	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
092-07	Winnipeg	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
094-07	Winnipeg	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2007
095-07	Winnipeg	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2008
096-07	Quebec City	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2	2008
097-07	Quebec City	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2008
098-07	Quebec City	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2008
099-07	Quebec City	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2008
101-07	Quebec City	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2008
102-07	Quebec City	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2008
103-07	Quebec City	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2008
118-07	Ottawa	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2008
120-07	Ottawa	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2008
121-07	Ottawa	G3	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2008
122-07	Ottawa	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2008
124-07	Ottawa	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2008
125-07	Ottawa	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2	2008
128-07	Ottawa	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2	2008
131-07	Ottawa	G3	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2008
133-07	Ottawa	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2008
134-07	Ottawa	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2008
135-07	Ottawa	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2008
172-07	Halifax	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2008
177-07	Halifax	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2008
178-07	Halifax	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2008
179-07	Halifax	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2008
185-07	Halifax	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2008
186-07	Halifax	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2008
187-07	Halifax	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2008
188-07	Halifax	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2008
193-07	Halifax	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2008
004-09	Quebec City	G3	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2009
005-09	Quebec City	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2009
008-09	Quebec City	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2	2009
010-09	Quebec City	G3	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2009
017-09	Halifax	G3	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2009
018-09	Halifax	G3	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2009
019-09	Halifax	G3	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2009
020-09	Halifax	G3	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2009
025-09	Halifax	G3	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2009
026-09	Halifax	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2009
031-09	Halifax	G3	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2009
038-09	Halifax	G4	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2009
039-09	Halifax	G4	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2009
053-09	Ottawa	G9	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2009
060-09	Ottawa	G9	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2009
063-09	Ottawa	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2009
070-09	Ottawa	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2009
072-09	Ottawa	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2009
073-09	Ottawa	G9	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2009
074-09	Ottawa	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2	2009
082-09	Ottawa	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2009
085-09	Ottawa	G9	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2009

Depending on the genome segment there were as many as 5 subclusters within the genogroup II cluster. Canadian strains in a single cluster for one genome segment did not necessarily indicate they would be in the same subcluster throughout all genome segments. Evolutionary divergence between clusters varied from segment to segment. Subclusters for NSP1 strains, for instance, diverged by as much as 0.20 base substitutions per site when pairwise identity was calculated using the Maximum Likelihood model in the MEGA 5.0 program [Tamura et al. 2011]. Strains were considered in the same NSP1 subcluster if they differed by less than 0.06 base substitutions per site. Other segments' genogroup II clusters had varying cutoff values based on suitable divergence amongst strains in different clusters with little overlap of strains in different clusters. Minimum cut-off values between Canadian genogroup II and I clusters are as follows: VP1 0.025, VP2 0.02, VP3 0.05, VP4 0.04, VP6 0.035, VP7 0.05, NSP1 0.06, NSP2 0.05, NSP3 0.04, NSP4 0.08 and NSP5 0.02 base substitutions per site. Clonal viruses were considered clonal when all of the sub-clusters in the 11 segment phylogenetic analyses were identical. Although several clonal strains were found, there were many with differences. Of the 76 genogroup II strains genotyped, 28 different clonal groups were found. Likewise, the 8 G2 strains differed enough in some segments to create 4 different clonal groups, but only differed in sub-genotypic clusters in the VP1, VP3 and NSP4 genome segments, as shown in Table 7.

Table 7: Subgenotypic lineages for all 84 fully genotyped strains in this study. Colours in boxes correspond to the sub-genotype lineages defined in 10 segments. Sub-genotypes are also shown after the decimal in genotype boxes. As VP7 types are not sub-genotype differences but full genotype differences, VP7 was not colour coded. Clonal numbers refer to identical genome constellations (all genotypes and subgenotypes are the same)

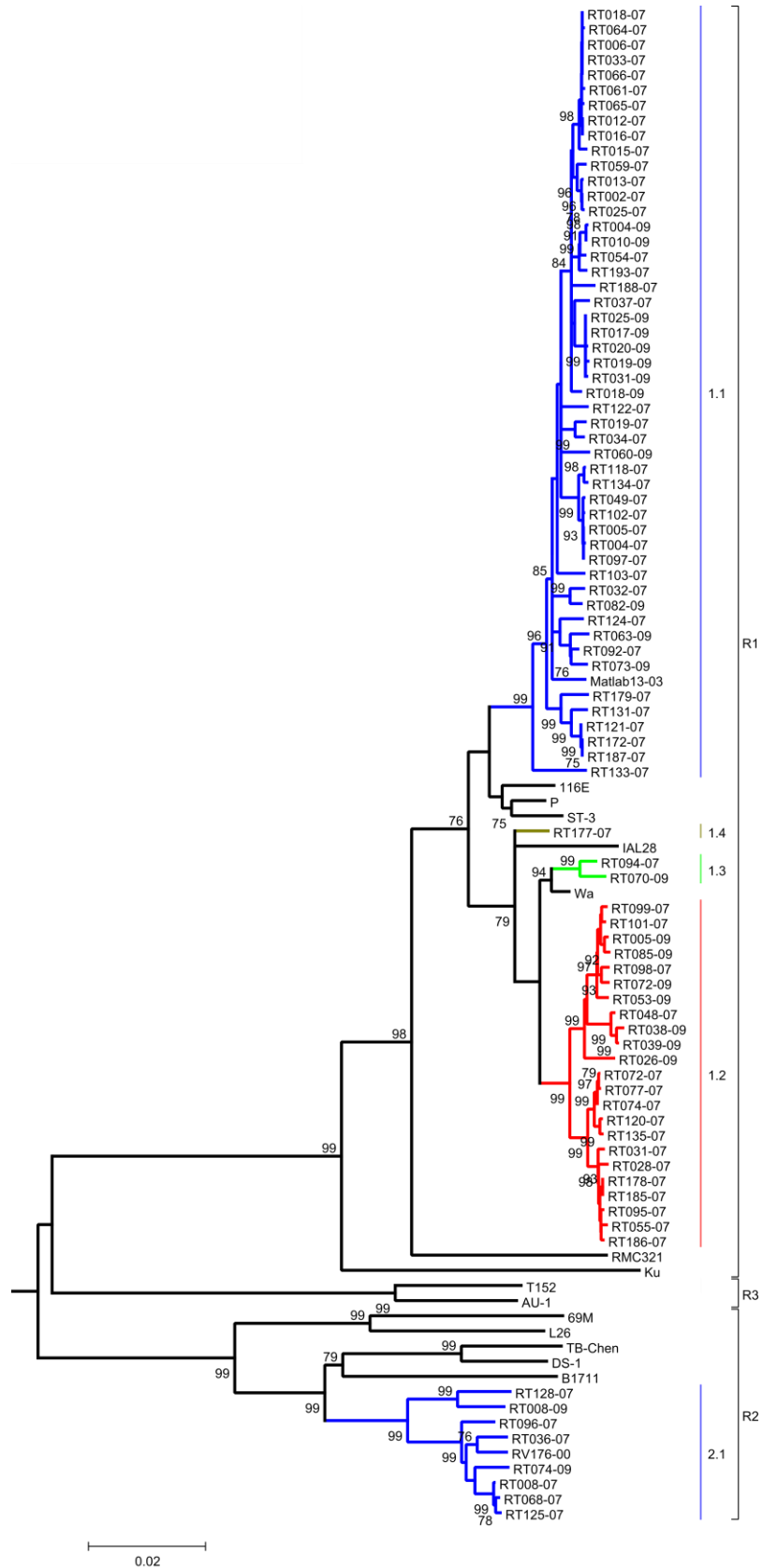
Sample	Testing Site	VP7 (G)	VP4 (P)	VP6 (I)	VP1 (R)	VP2 (C)	VP3 (M)	NSP1 (A)	NSP2 (N)	NSP3 (T)	NSP4 (E)	NSP5 (H)	Clonal Number	Year
006-07	Quebec City	1.1	P[8]	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1	2007
012-07	Quebec City	1.1	P[8]	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1	2007
016-07	Quebec City	1.1	P[8]	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1	2007
018-07	Quebec City	1.1	P[8]	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1	2007
033-07	Halifax	1.1	P[8]	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1	2007
061-07	Ottawa	1.1	P[8]	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1	2007
064-07	Ottawa	1.1	P[8]	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1	2007
065-07	Ottawa	1.1	P[8]	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1	2007
066-07	Ottawa	1.1	P[8]	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1	2007
134-07	Ottawa	1.1	P[8]	1.1	1.1	1.1	1.1	1.2	1.1	1.3	1.1	1.1	2	2008
118-07	Ottawa	1.1	F[8]	1.1	1.1	1.1	1.1	1.1	1.1	1.3	1.1	1.1	4	2008
005-07	Quebec City	1.1	P[8]	1.1	1.1	1.3	1.1	1.2	1.1	1.3	1.1	1.1	5	2007
049-07	Ottawa	1.1	P[8]	1.1	1.1	1.3	1.1	1.2	1.1	1.3	1.1	1.1	5	2007
097-07	Quebec City	1.1	P[8]	1.1	1.1	1.3	1.1	1.2	1.1	1.3	1.1	1.1	5	2008
102-07	Quebec City	1.1	P[8]	1.1	1.1	1.3	1.1	1.2	1.1	1.3	1.1	1.1	5	2008
193-07	Halifax	1.1	P[8]	1.1	1.1	1.3	1.1	1.2	1.1	1.1	1.1	1.1	5	2008
054-07	Ottawa	1.1	P[8]	1.1	1.1	1.3	1.1	1.1	1.1	1.1	1.1	1.1	8	2007
179-07	Halifax	1.1	P[8]	1.1	1.1	1.3	1.1	1.1	1.1	1.1	1.1	1.1	8	2008
187-07	Halifax	1.1	P[8]	1.1	1.1	1.3	1.1	1.1	1.1	1.1	1.1	1.1	8	2008
188-07	Halifax	1.1	P[8]	1.1	1.1	1.3	1.1	1.1	1.1	1.4	1.1	1.1	9	2008
172-07	Halifax	1.1	P[8]	1.1	1.1	1.3	1.1	1.1	1.1	1.1	1.1	1.2	10	2008
133-07	Ottawa	1.1	P[8]	1.1	1.1	1.3	1.3	1.2	1.5	1.1	1.1	1.1	12	2008
098-07	Quebec City	1.1	P[8]	1.1	1.2	1.1	1.1	1.2	1.1	1.3	1.1	1.1	14	2008
095-07	Winnipeg	1.1	P[8]	1.1	1.2	1.1	1.1	1.2	1.1	1.1	1.1	1.1	15	2008
026-09	Halifax	1.1	P[8]	1.2	1.2	1.1	1.2	1.2	1.1	1.1	1.2	1.1	16	2009
099-07	Quebec City	1.1	P[8]	1.2	1.2	1.1	1.2	1.2	1.4	1.2	1.2	1.1	17	2008
101-07	Quebec City	1.1	P[8]	1.2	1.2	1.1	1.2	1.2	1.4	1.2	1.2	1.1	17	2008
005-09	Quebec City	1.1	P[8]	1.2	1.2	1.1	1.2	1.2	1.4	1.2	1.2	1.1	17	2009
072-09	Ottawa	1.1	P[8]	1.2	1.2	1.1	1.2	1.2	1.4	1.2	1.1	1.1	18	2009
185-07	Halifax	1.1	P[8]	2.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.1	23	2008
028-07	Halifax	1.1	P[8]	2.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	24	2007
031-07	Halifax	1.1	P[8]	2.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	24	2007
055-07	Ottawa	1.1	P[8]	2.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	24	2007
178-07	Halifax	1.1	P[8]	2.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	24	2008
186-07	Halifax	1.1	P[8]	2.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	24	2008
015-07	Quebec City	1.1	P[8]	1.1	1.1	1.1	1.1	1.2	1.1	1.1	1.1	1.1	27	2007
032-07	Halifax	1.1	P[8]	1.1	1.1	1.1	1.1	1.2	1.1	1.1	1.1	1.1	27	2007
092-07	Winnipeg	1.1	P[8]	1.1	1.1	1.1	1.1	1.2	1.1	1.1	1.1	1.1	27	2007
103-07	Quebec City	1.1	P[8]	1.1	1.1	1.1	1.1	1.2	1.1	1.1	1.1	1.1	27	2008
122-07	Ottawa	1.1	P[8]	1.1	1.1	1.1	1.1	1.2	1.1	1.1	1.1	1.1	27	2008
124-07	Ottawa	1.1	P[8]	1.1	1.1	1.1	1.1	1.2	1.1	1.1	1.1	1.1	27	2008
063-09	Ottawa	1.1	P[8]	1.1	1.1	1.1	1.1	1.2	1.1	1.1	1.1	1.1	27	2009
082-09	Ottawa	1.1	P[8]	1.1	1.1	1.1	1.1	1.2	1.1	1.1	1.1	1.1	27	2009
019-07	Quebec City	1.2	P[8]	1.1	1.2	1.1	1.2	1.1	1.2	1.1	1.3	1.1	13	2007
177-07	Halifax	1.2	F[8]	1.3	1.2	1.4	1.2	1.2	1.1	1.2	1.2	1.2	22	2008
072-07	Edmonton	1.2	P[8]	2.2	1.2	1.2	1.2	1.2	1.1	1.2	1.2	1.2	25	2007
074-07	Edmonton	1.2	P[8]	2.2	1.2	1.2	1.2	1.2	1.1	1.2	1.2	1.2	25	2007
077-07	Edmonton	1.2	P[8]	2.2	1.2	1.2	1.2	1.2	1.1	1.2	1.2	1.2	25	2007
120-07	Ottawa	1.2	P[8]	2.2	1.2	1.2	1.2	1.2	1.1	1.2	1.2	1.2	25	2008
135-07	Ottawa	1.2	P[8]	2.2	1.2	1.2	1.2	1.2	1.1	1.2	1.2	1.2	25	2008
094-07	Winnipeg	1.2	P[8]	1.1	1.2	1.3	1.2	1.2	1.3	1.2	1.1	1.1	26	2007
070-09	Ottawa	1.2	P[8]	1.1	1.2	1.3	1.2	1.1	1.2	1.3	1.2	1.1	28	2009
002-07	Quebec City	3	P[8]	1.1	1.1	1.1	1.3	1.1	1.2	1.1	1.1	1.1	6	2007
013-07	Quebec City	3	P[8]	1.1	1.1	1.1	1.3	1.1	1.2	1.1	1.1	1.1	6	2007
025-07	Quebec City	3	P[8]	1.1	1.1	1.1	1.3	1.1	1.2	1.1	1.1	1.1	6	2007
131-07	Ottawa	3	P[8]	1.1	1.1	1.1	1.3	1.1	1.2	1.1	1.1	1.1	6	2008
004-07	Quebec City	3	P[8]	1.1	1.1	1.1	1.3	1.1	1.2	1.1	1.3	1.1	7	2007
017-09	Halifax	3	P[8]	1.1	1.1	1.1	1.3	1.1	1.2	1.1	1.3	1.1	7	2009
019-09	Halifax	3	P[8]	1.1	1.1	1.1	1.3	1.1	1.2	1.1	1.3	1.1	7	2009
020-09	Halifax	3	P[8]	1.1	1.1	1.1	1.3	1.1	1.2	1.1	1.3	1.1	7	2009
025-09	Halifax	3	P[8]	1.1	1.1	1.1	1.3	1.1	1.2	1.1	1.3	1.1	7	2009
031-09	Halifax	3	P[8]	1.1	1.1	1.1	1.3	1.1	1.2	1.1	1.3	1.1	7	2009
037-07	Halifax	3	P[8]	1.1	1.1	1.1	1.3	1.1	1.1	1.1	1.1	1.1	11	2007
059-07	Ottawa	3	P[8]	1.1	1.1	1.1	1.3	1.1	1.1	1.1	1.1	1.1	11	2007
121-07	Ottawa	3	P[8]	1.1	1.1	1.1	1.3	1.1	1.1	1.1	1.1	1.1	11	2008
004-09	Quebec City	3	P[8]	1.1	1.1	1.1	1.3	1.1	1.1	1.1	1.1	1.1	11	2009
010-09	Quebec City	3	P[8]	1.1	1.1	1.1	1.3	1.1	1.1	1.1	1.1	1.1	11	2009
018-09	Halifax	3	P[8]	1.1	1.1	1.1	1.3	1.1	1.1	1.1	1.1	1.1	11	2009
048-07	Ottawa	4	P[8]	1.2	1.2	1.2	1.1	1.2	1.1	1.2	1.2	1.1	20	2007
038-09	Halifax	4	P[8]	1.2	1.2	1.2	1.1	1.2	1.3	1.2	1.2	1.1	21	2009
039-09	Halifax	4	P[8]	1.2	1.2	1.2	1.1	1.2	1.3	1.2	1.2	1.1	21	2009
034-07	Halifax	9	P[8]	1.1	1.1	1.1	1.1	1.2	1.1	1.1	1.1	1.1	3	2007
060-09	Ottawa	9	P[8]	1.1	1.1	1.1	1.1	1.2	1.1	1.1	1.1	1.1	3	2009
073-09	Ottawa	9	P[8]	1.1	1.1	1.1	1.1	1.2	1.1	1.1	1.1	1.1	3	2009
053-09	Ottawa	9	P[8]	1.2	1.2	1.1	1.2	1.2	1.4	1.2	1.2	1.1	19	2009
065-09	Ottawa	9	P[8]	1.2	1.2	1.1	1.2	1.2	1.4	1.2	1.2	1.1	19	2009
008-07	Quebec City	2	F[4]	1.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	1	2007
068-07	Edmonton	2	F[4]	1.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	1	2007
096-07	Quebec City	2	F[4]	1.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	1	2008
125-07	Ottawa	2	F[4]	1.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	1	2008
074-09	Ottawa	2	F[4]	1.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	1	2009
036-07	Halifax	2	F[4]	1.1	2.1	2.1	2.1	2.1	2.1	2.1	2.3	2.1	2	2007
128-07	Ottawa	2	F[4]	1.1	2.1	2.1	2.1	2.1	2.1	2.1	2.2	2.1	3	2008
008-09	Quebec City	2	F[4]	1.1	2.1	2.2	2.1	2.2	2.1	2.1	2.3	2.1	4	2009

### 3.2.2 Analysis of Phylogenetics and Motifs:

Many amino acid motifs involved in rotavirus protein structure, function and antigenicity were analyzed. Motifs were analyzed and compared to those previously found in the rotavirus literature, as described in Table 3.

#### VP1:

The VP1 segment encodes the RNA-dependant RNA polymerase. Phylogenetic analysis of the VP1 segment of Canadian rotaviruses revealed high relatedness within the Canadian strains shown in Figure 8. All Canadian strains in the same genotype differed by less than 0.05 nucleotide substitutions per site. Of the 84 strains sequenced, the 76 R1 strains clustered into four sub-genotypic lineages defined by identities of less than 0.025 nucleotide substitutions per site. The largest of the three lineages contained 50 Canadian strains and was closely related to the type strain Matlab 13-03. Lineage 2 contained 23 strains and was highly conserved. No type strains fell within sub genotypic lineages 2 or 4, but they were most closely related to the Wa type strain. Lineage 3 was highly identical to Wa, but only contained two strains, RT094-07 and RT070-09. Strain RT177-07 fell into lineage 2, but evidently differs from most Canadian strains including many within its lineage. RT177-07 is most closely related to Wa but is also fairly closely related to the type strain IAL28. R2 strains clustered into two single sub-genotypic clusters, based on the previously stated cut-off value of 0.025 nucleotide substitutions per site. Two Canadian strains, RT128-07 and RT008-09, were in cluster 1 and the other six Canadian R2 strains were part of sub-genotypic lineage 2. All R2 strains were most identical to RV176-00. The phylogenetic tree for VP1 is shown in Figure 8.



Five VP1 motifs were analyzed. They included those predicted to function in the RNA-dependant RNA polymerase (RdRp) as bivalent cation coordinating motifs, and sugar selection motifs [Vasquez-del Carpio et al. 2006]. All of the VP1 motifs are part of the right-hand resembling finger and palm region of the RdRp. All of the described motifs (A-D and F) were completely conserved, except for motif D, which had two variable sites. Two of the 84 strains tested had a mutation causing variation in the alanine at position 665. The alanine changed to valine or threonine. At position 666, 1 of the 84 samples analyzed contained a mutation resulting in the change of the charged amino acid lysine to asparagine. All mutations found were in samples that were part of the R1 genotype group. A representative alignment of several samples in the study and type strains can be found in Figure 9.

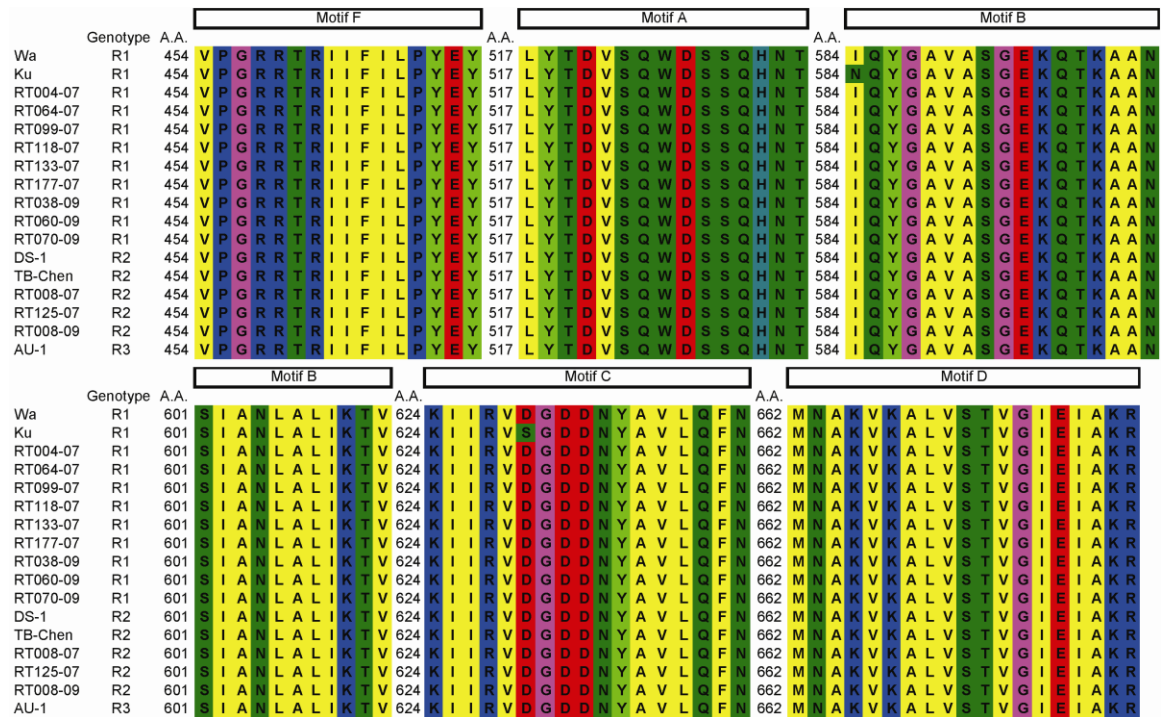


Figure 9: Representative amino acid alignment of the five RNA-dependant RNA-polymerase motifs found in Canadian strains. Motifs are organized from lowest amino acid residue to highest.



VP2:

The VP2 genome segment encodes the core scaffold protein. Full genome analysis of the Canadian rotavirus strains revealed two VP2 genotypes, C1 and C2. Canadian strains in the C1 genotype clustered into three sub-genotypic lineages. The largest of the three lineages, sub-lineage 1, included 31 Canadian strains. Sub-genotype 3, the second largest sub lineage, was comprised of 27 strains. The largest two genogroup II clusters are most closely identical to the Matlab 13-03 type-strain, while sub-lineage 2, comprised of 18 strains, is most identical to the P type strain. All G2 strains fall into the same cluster and are most closely related to the type strain RV176-00. A phylogenetic tree for the VP2 genome segment is shown in Figure 10.

Analysis of the region from amino acid 1 to 156, describing VP1, VP3 and NSP2 interacting domains, showed substantial variation [Heiman et al. 2008]. There were 26 variable sites, not including exclusively genotype-specific sites. In addition, amino acid analysis revealed 33 genotype-specific variations among all C1 and C2 strains.

A comparison of G4 strains (which were C1 strains) with other C1 strains showed that G genotype-specific alterations in the VP2 genome segment at sites 29, 30, 54 and 416 existed between G4 strains and other C1 strains. Several strains also contained inserts, composed of poly-lysine and poly-asparagine, at lengths up to four amino acids, beginning at amino acid 38. These strains' inserts were not found in genogroup I strains.

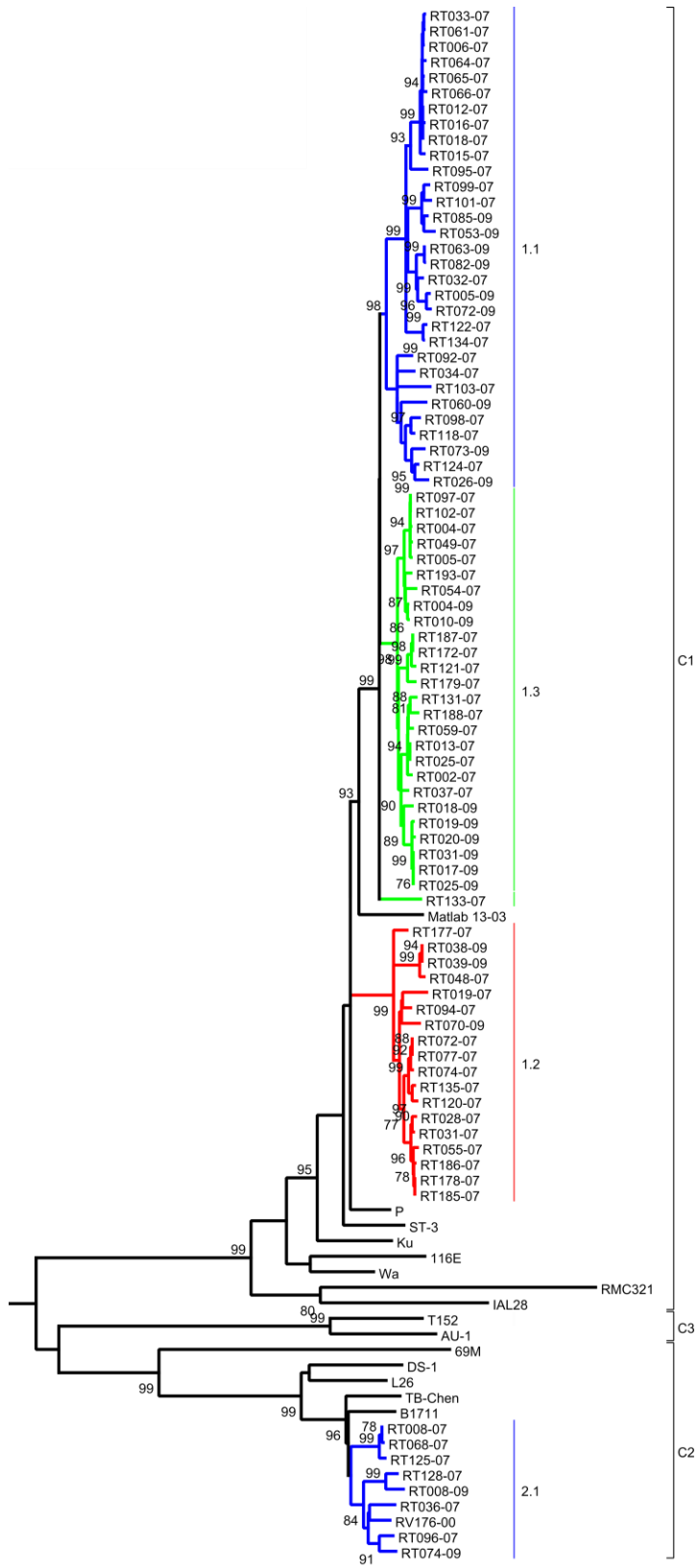


Figure 10: Neighbor Joining phylogenetic tree with 2000 bootstraps of the VP2 genome segment. Square brackets represent genotypes, while coloured lines represent sub-genotypic lineages. Measured in units of base substitutions per nucleotide.

## VP3:

Phylogentic analysis of the VP3 genome segment of Canadian rotaviruses showed a relatively high amount of divergence within VP3 genotypes and between VP3 genotypes. The VP3 protein is part of the core structural protein group of rotaviruses, thought to cap the viral mRNA strains during transcription, and a degree of conservation might be expected in this important functional gene. The largest of the three M1 sub-clusters contained 55 of the 76 Canadian M1 strains. Sub-genotype cluster 1 is most identical to the P type strain. Sub cluster 3 contains only the Canadian strain RT133-07. RT133-07 is most identical to Matlab 13-03; a type strain that is often highly identical to the majority of Canadian genogroup II strains, when considering other segments. The final Canadian M1 cluster (sub cluster 2) contains 20 Canadian strains and is most identical to the Wa type strain. Seven of the eight M2 genotype strains fall into the first cluster, most closely related to the RV176-00 type strain. RT008-09 falls into the M2 sub-genotypic lineage 2 most closely related to DS-1, but has lower identity with DS-1 of the other Canadian M2 strains. The VP3 phylogenetic tree is shown in Figure 11.

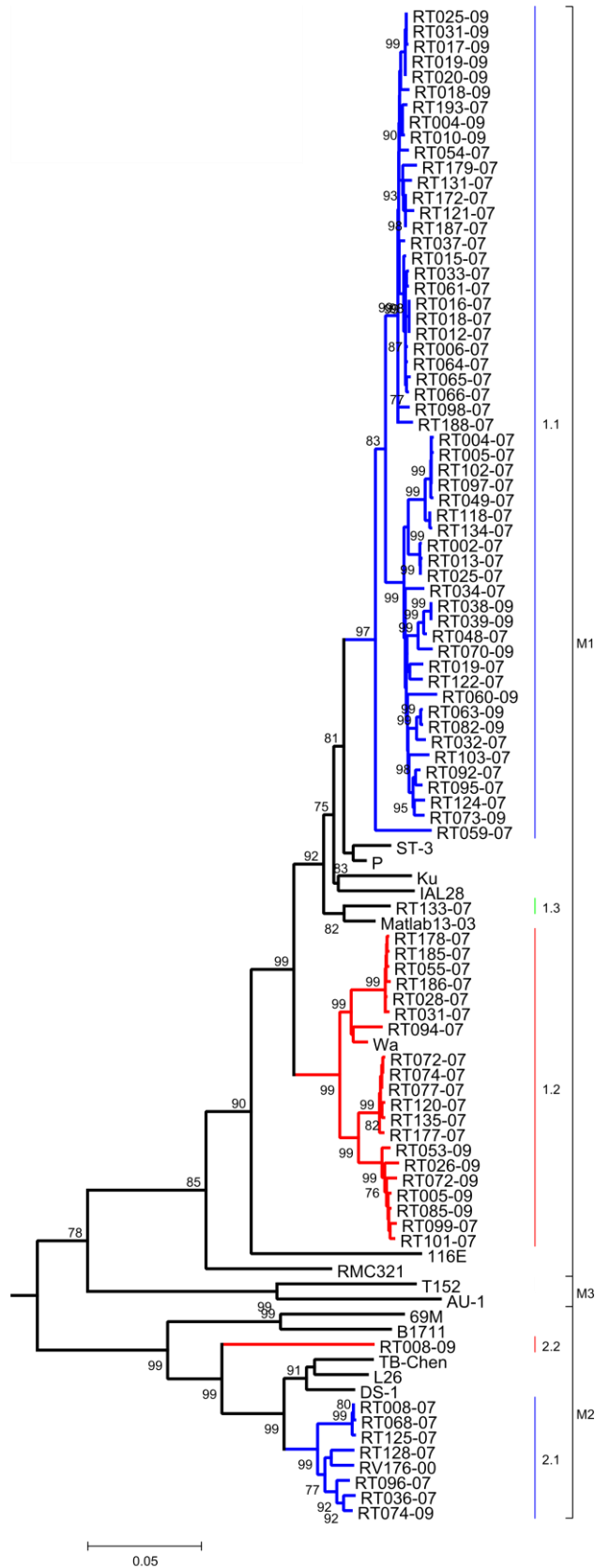


Figure 11: Neighbor Joining phylogenetic tree with 2000 bootstraps of the VP3 genome segment. Square brackets represent genotypes, while coloured lines represent sub-genotypic lineages. Measured in units of base substitutions per nucleotide.

Analysis of the VP3 ORF region motifs showed many conserved structural and functional motifs. Amino acids 447-474, consisting of the GTP/S-adenosyl-L-methionine-binding region, are not completely conserved. There were several variable sites in this region in the 76 M1 strains, as well as four genotype-specific variations in M2 strains. Variations mostly occurred in the first 12 amino acids of the domain. Roughly half of the sites in M1 strains varied in amino acids 447-459, while only 3 sites showed any variation in the final 12 amino acids. M2 genotype strains were highly conserved in this motif, except for three sites. The three variable M2 sites varied in only one strain each. There was less variation when comparing the M2 strains with the M1 strains within the GTP/S-adenosyl-L-methionine binding region. The casein kinase II functional region described as K followed by TAMD region was found as KXTAMD, where X was leucine in M1 strains and methionine in M2 strains. The casein kinase II region was completely conserved in residues 383-388. Cook and McCrae described a potential guanylyltransferase activity site beginning at position 541 [Cook and McCrae. 2004]. Residues 541-545 are completely conserved with the sequence KPTGN, as previously described [Cook and McCrae. 2004].

Structural motifs were also analyzed. All except one of the fifteen previously described proline residues were found in the previously described position with complete conservation within Canadian strains. However, the proline at position 502 was not found in any of the Canadian sequences. Instead, proline sites at positions 182, 280 and 498 were completely conserved. The four cysteine residues described by Cook and McCrae were all 100% conserved in the investigated strains. In addition, conserved cysteine residue sites were found at position 78 and 528 in the Canadian strains.

VP4:

Phylogenetic analysis of the VP4 protein, depicted in Figure 12, indicates that there are four sub-genotypic lineages within the P[8] genotype of Canadian strains. Most strains belong to lineage 1 while only single strains belong to lineages 3 and 4. Sub genotypic lineage 1 is most closely related to the type strain P. Sub genotypic lineage 2 contains 11 strains and is most closely related to the Wa type strain. While most lineages were closely related to the type-strain Wa, the sub-genotypic lineage 2 is distinct from lineage 1. Lineage 4, made up of RT118-07, is most distinct from all other P[8] strains, differing by up to 0.1223 nucleotide substitutions per site from other P[8] strains. RT118-07 is most closely related to the lineage 3 strain, but differs enough not to be included in lineage 3. All Canadian P[4] strains fall into the same sub-genotypic lineage and are most closely related to the type strain TB-Chen.

Rotavirus spike proteins, coded for by the VP4 genome segment, have several regions of structural and functional interest. Several functional cysteine residues at amino acid sites 203, 317, 380 and 786 were previously described, in a lamb rotavirus [Chen et al. 2009]. In the current study, cysteine residues were found at sites 215, 317, 379 and 773. Although the positions have changed, it is likely that functional homology is conserved, since the four residues align in conserved regions throughout all rotavirus genotypes. Previously described functional proline residues were found completely conserved in all genotypes. There was however a deletion in the Canadian human strains when compared to the lamb strain causing the final two prolines to shift to positions 224 and 225 rather than 225 and 226 [Chen et al. 2009].

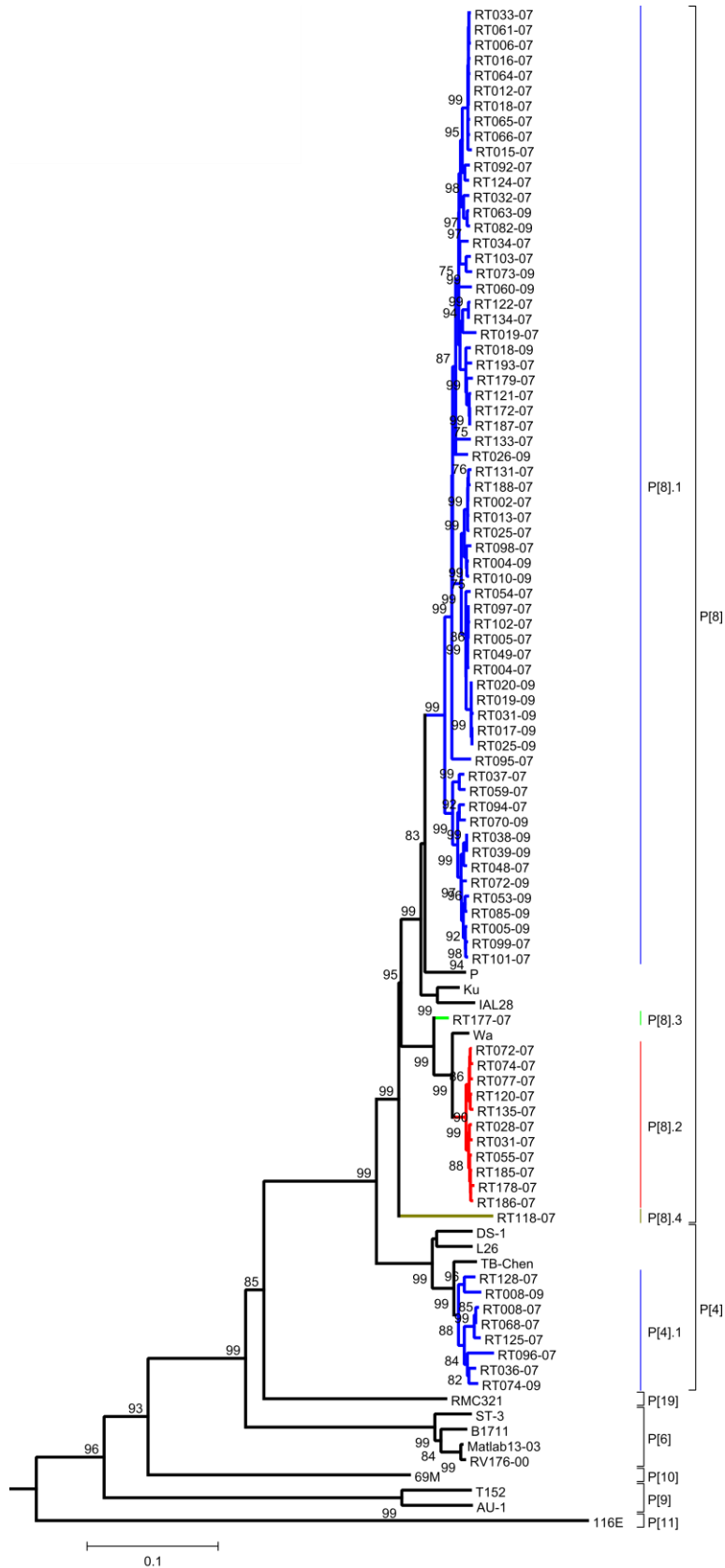


Figure 12: Neighbor Joining phylogenetic tree with 2000 bootstraps of the VP4 genome segment. Square brackets represent genotypes, while coloured lines represent sub-genotypic lineages. Measured in units of base substitutions per nucleotide.

Residues 377 to 395, comprising the internal fusion domain described by Chen and colleagues [Chen et al. 2009], were found to be highly conserved. There was some variability amongst all strains in 2 sites (384, 392). However, this variability was rare and for the most part the site was well conserved. There were also four P[4] genotype-specific variations noted in this region (384, 391, 396 and 399).

In Canadian strains the potential arginine residues marking VP4 cleavage sites are located at the conserved positions 230 and 240, but site 246 varies in one strain, RT120-07 [Gilbert and Greenberg. 1998], having a lysine at position 246 instead of arginine.

Coulson and colleagues described potential integrin ligand binding sites on the VP4 spike protein in a 2003 publication. One important motif was a R-D-G-E motif found in Rotavirus A strains, which is a motif known to bind specific  $\alpha\beta$  heterodimeric transmembrane glycoprotein ligands [Coulson et al. 1997]. In the 84 sequenced Canadian samples the R-D-G-E site can be found at residues 306-309 and is completely conserved.

Antigenic amino acid motifs were also examined in this study. The previously analyzed antigenic amino acids of the P[8] sequence from RotaTeq and Rotarix were used as reference residues [Zeller et al. 2011]. Comparing antigenic sites involved taking into account the position changes caused by the amino acid deletion in Canadian strains, few of the domains, described previously by Zeller and colleagues, were conserved. As previously described, all sites in the 8-1 region were variable except for site 100 and 193. Amino acid 192 varied in a genotype-specific manner. Most variations in region 8-1 were rare, if a site was analyzed individually. However, when looking at the 8-1 motif as a whole, one or more differences between the Canadian strains and vaccine strains were



identified. Analyzing variations temporally, it was shown that site 149 changed from asparagine to serine in both P[8] and P[4] strains in 2007-2008. However, in 2009, the variation only occurred in P[4] strains. Also, variation from RotaTeq amino acids was more frequent in 2009 than in 2007 or 2008 at sites 194 and 195.

Antigenic region 8-2 was nearly completely conserved in Canadian rotavirus strains. However, there was a difference in the amino acid sequence at residue 182 in strain RT026-09.

Domain 8-3 contained a few highly variable sites, such as sites 113, 125, 131 and 135. Interestingly, when comparing sites 125, 131 and 135 temporally, the strains were variable in 2007 and 2008, but all residues were conserved and identical to RotaTeq in 2009. Most sites contain genotype-specific P[4] variability that is divergent from RotaTeq in antigenic region 8-3; the alignment is shown in Figure 13. Antigenic region 8-4 is completely conserved in Canadian strains except for residue 89, which varies from RotaTeq in P[4] strains only.

Analysis of the VP5\* (asterisk identifies products of VP4 cleavage) region of the VP4 protein involves five antigenic epitopes. The first of these epitopes is conserved within genotypes, but not intergenotypically in most sites. In the eight sites of the epitope that were analyzed, all P[4] strains differ from RotaTeq at sites 383, 385, 387, 392, and 439 and are completely conserved. P[8] strains differ from RotaTeq in sites 383 and 385 only, and are otherwise completely conserved. All other VP5\* sites were completely conserved and had the same residues as RotaTeq except for epitope 5-5 at residue 305, in which all P[4] strains contained a serine rather than a lysine residue.

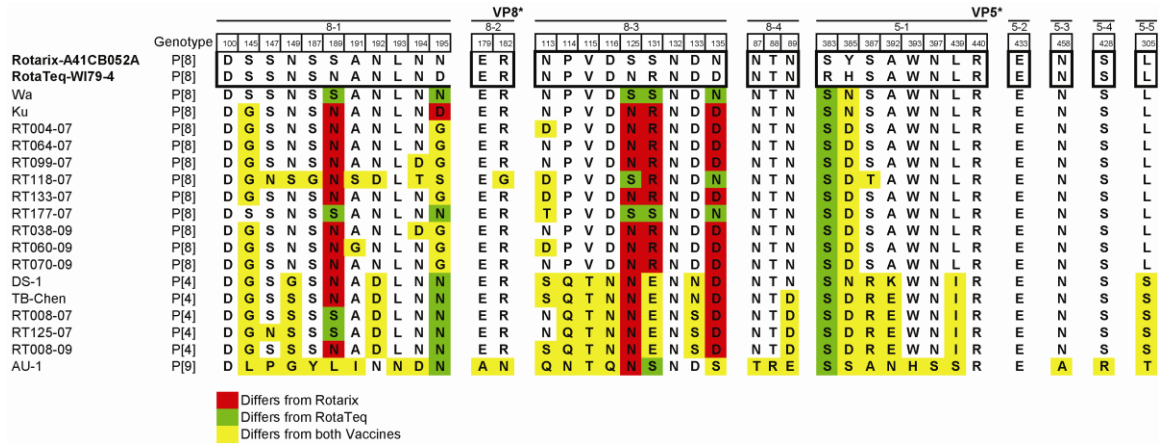


Figure 13: VP4 antigenic motif alignment using the vaccine sequences in these regions as references. Four VP8\* motifs and five VP5\* motifs were analyzed.

VP6:

The positions of Canadian rotavirus strains within the phylogenetic tree revealed high identity within genotypes. Figure 14 shows the phylogenetic tree for Canadian VP6 genome segments. There were two distinct clusters of Canadian strains in the I1 genotype. One strain, RT019-07, was related highly to both sub-genomic lineages, but was more closely related to lineage 2 strains. There were 51 sub-genotype 1 Canadian strains and about half of those were in lineage 2. Lineage 1 strains were highly identical to the type strain Matlab 13-03, while sub-genotypic lineage 2 strains were most identical to the ST-3 type strain. The I2 genotype strains were highly identical to each other. However, the I2 strains differed substantially from the type strains DS-1, TB-Chen and L26. Rather, the I2 strains clustered with RV176-00 and B1711.

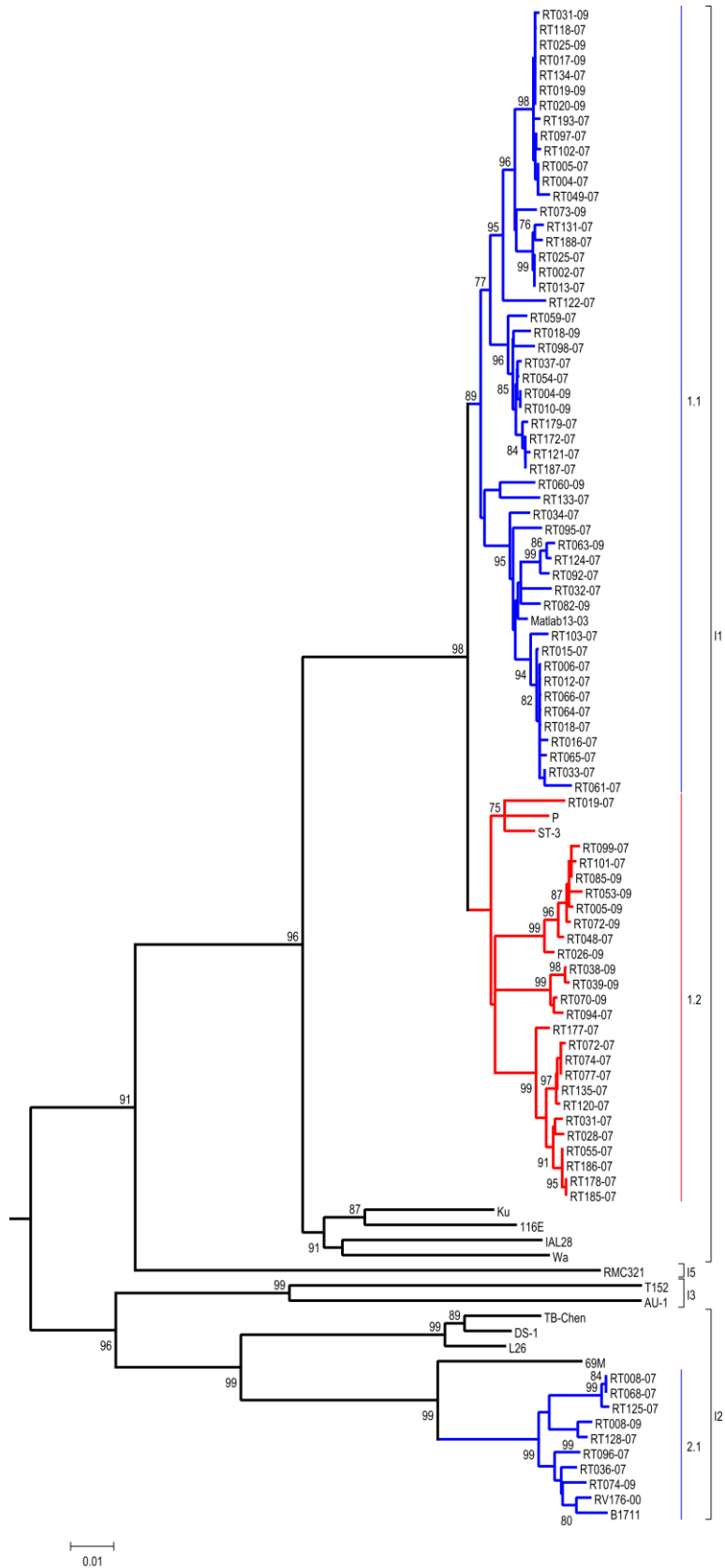


Figure 14: Neighbor Joining phylogenetic tree with 2000 bootstraps of the VP6 genome segment. Square brackets represent genotypes, while coloured lines represent sub-genotypic lineages. Measured in units of base substitutions per nucleotide.

Key differences in the amino acid sequences of the VP6 trimerization domain are mostly genotype-specific; amino acid sequences are highly conserved within genogroups between amino acids 246-314 [Affranchino and Gonzalez. 1997].

The B and H domains of the VP6 protein were named based on their folding pattern. B and H domains contain sub domains interacting with VP2, VP4 and VP7 proteins [Affranchino and Gonzalez. 1997, Heiman et al. 2008]. The regions specifically encoding binding sites to VP2, VP4 and VP7 are relatively large amino acid sequences compared to other binding domains. The sequences seem to be highly conserved within genotypes, with a steady number of genotype-specific variations throughout. However, the variations are well spaced throughout the genome segment and highly conserved within the genome segments.

VP7:

Canadian strains and their respective genotype clusters are shown in Figure 15. VP7 components of Canadian strains fall into five different VP7 genotypes. Of the strains fully genotyped only the G1 Canadian strains were divided into two subgenotypic lineages, the larger of which is less identical than the smaller. Canadian G2 strains were most closely identical to the TB-Chen strain and more divergent from the DS-1 type strain. Canadian G9 strains were more closely identical to RMC321 than to the Indian strain 116E. Canadian G3 strains were more closely related to the P type strain than the AU-1 type strain.

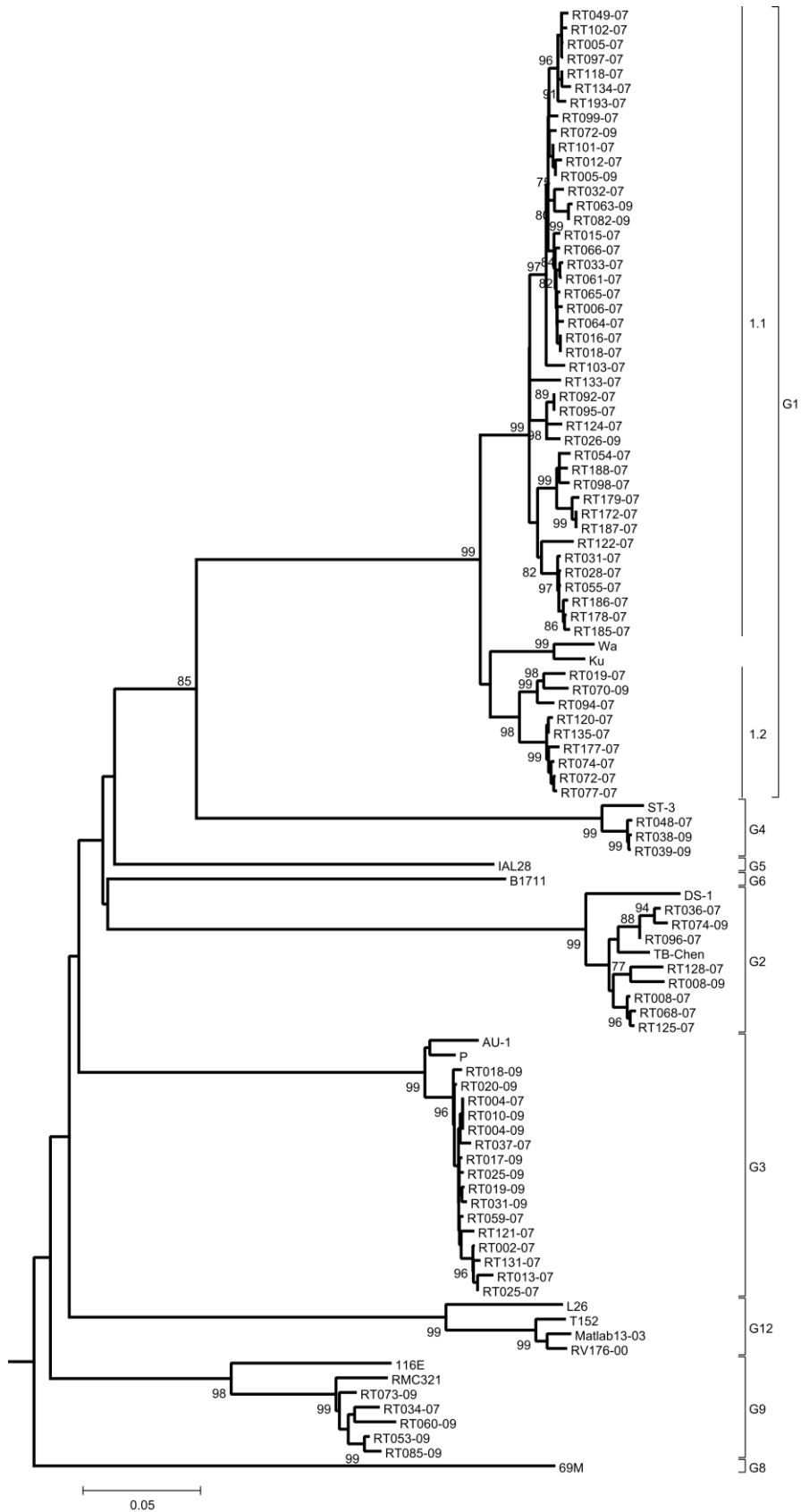


Figure 15: Neighbor Joining phylogenetic tree with 2000 bootstraps of the VP7 genome segment. Square brackets represent genotypes, while unbracketed lines represent sub-genotypic lineages. Measured in units of base substitutions per nucleotide.

The 84 chosen VP7 genome segment motifs are based on the antigenic regions of the VP7 protein [Zeller et al. 2011]. Antigenic amino acid groups were grouped into epitopes found in different regions of the VP7 protein. Major epitopes include 7-1a, 7-1b and 7-2. Amino acids from these epitopes that were analyzed were residues known to be associated with antigenic escape. In this case, these sites were analyzed with vaccines in mind. In region 7-1a, there were several genotype-specific changes. In fact, only two sites were conserved in all Canadian genotypes. The most divergent strains, from the vaccines and other genotypes, were G2 strains, differing from Rotarix in 9 of 14 residues and up to 4 from the RotaTeq vaccine in the G2 region. Canadian G4 strains were also very divergent from Rotarix with variations in seven sites, but only two were divergent from the G4 strain of RotaTeq. All strains contained variations in epitope 7-1a at sites known to cause immune escape except for G3. Canadian G9 strains (which have no corresponding VP7 protein in vaccines) contained only three variable sites when compared to the G3 of RotaTeq. However, G9 differed in several additional sites when compared to the Rotarix VP7 component. G1 strains differed more from RotaTeq than Rotarix. Moreover, G1 had the least conservation within the genotype, when compared to other genotypes.

Analysis of antigenic site 7-1b involved surveying six amino acids, one of which was conserved among all genotypes. Both vaccine strains completely matched all Canadian G1 amino acids for the first five amino acids of this epitope. A single G1 strain, RT072-07, did not match vaccine strains at this epitope, since it had an isoleucine residue rather than a valine residue at site 212. There was also some variation at residue 242 in the Canadian G1 strains. The G2 strains differed from the G2 of RotaTeq in two sites,

and from Rotarix in three sites. One site in the 7-1b epitope was variable within the G2 genotype. The G2 variable site often did not correspond with the G2 component of the RotaTeq vaccine or with the VP7 component of the Rotarix vaccine, as shown in Figure 16. Epitope 7-1b diverged most from both vaccines amongst G3 strains. In this epitope G3 strains differed from Rotarix in four sites and from RotaTeq G3 in three sites. Canadian G4 and G9 strains contained the same sequence for the first five amino acids of this epitope, which varied from the RotaTeq G4 component at a single site and from Rotarix at three sites. The final site of the epitope (not usually associated with escape mutants) was G1-like in G4 strains and G2-like in G9 strains.

Canadian strains differed greatly from Rotarix at site 147 of the 7-2 epitope. In fact, all strains except for some G1 strains differed from the Rotarix vaccine at site 147. RotaTeq vaccines matched all non-G1 strains very well in this epitope. G4 and some G1 strains were the only genotypes in which there were Canadian strains with potential amino acid variations capable of immunological escape from RotaTeq. Despite the single mismatch at site 147, Rotarix was well matched to most G1 strains. However, Rotarix was a poor match to other genotypes in the 7-2 epitope, especially with G2 strains. Seven of nine residues differed from Canadian G2 strains when compared to Rotarix.

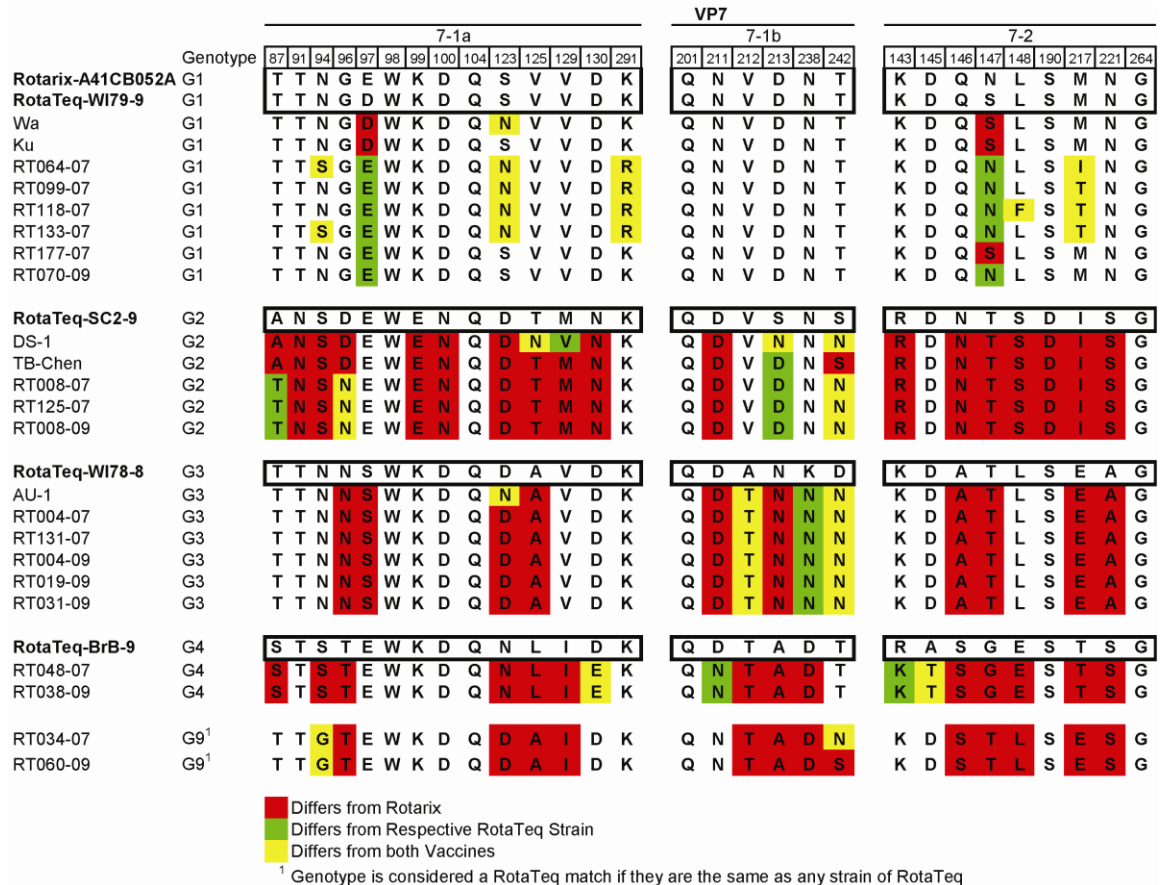


Figure 16: Representative amino acid alignment of VP7 antigenic regions. Strains are organized by VP7 genotype under their respective vaccine genotype component if applicable. Three VP7 antigenic motif regions were analyzed.

NSP1:

Variation of NSP1 is demonstrated by phylogenetic analysis, shown in Figure 17.

Phylogenetic analysis for NSP1 was greatly skewed when using the chosen outgroup, PO-13, to such a high degree that a substitute outgroup, T152, was used instead. Amino acid homology of P0-13 compared with mammalian viruses ranges from 16-18% [Ito et al. 2001].

Analysis of Canadian rotavirus nucleotide sequences showed as many as 0.1942 base substitutions per site within the A1 genotype. The A1 strains were divided into two distinct sub genotypic lineages. The larger of the two sub-genotypic lineages (sub-



genotype 2) contained 49 Canadian strains out of the total 76 A1 strains. The larger group was highly identical to the NSP1 genome segment of Matlab13-03. There were 27 A1 strains in the sub genotypic lineage 1, which was highly identical to the P type-strain. A2 strains fell into one sub-genotypic lineage which was highly identical to RV176-00. A phylogenetic tree for Canadian VP6 genome segments can be found in Figure 17.

The only motif analyzed for the NSP1 genome segment is the cysteine rich RNA-binding motif. The binding motif, previously described by Hua and colleagues in 1993, is cysteine rich but does not follow the sequence C-X<sub>2</sub>-C-X<sub>8</sub>-C-X<sub>3</sub>-H-X-C-X<sub>2</sub>-C-X<sub>5</sub>-C [Hua et al. 1993]. Instead, Canadian strains contain a cysteine rich domain with the conserved sequence C-X<sub>2</sub>-C<sub>2</sub>-X<sub>7</sub>-C-X<sub>2</sub>-C-X<sub>3</sub>-H-X-C-X<sub>2</sub>-C-X<sub>5</sub>-C. The amino acid motif is quite similar, but not identical to, the originally described RNA-binding motif.

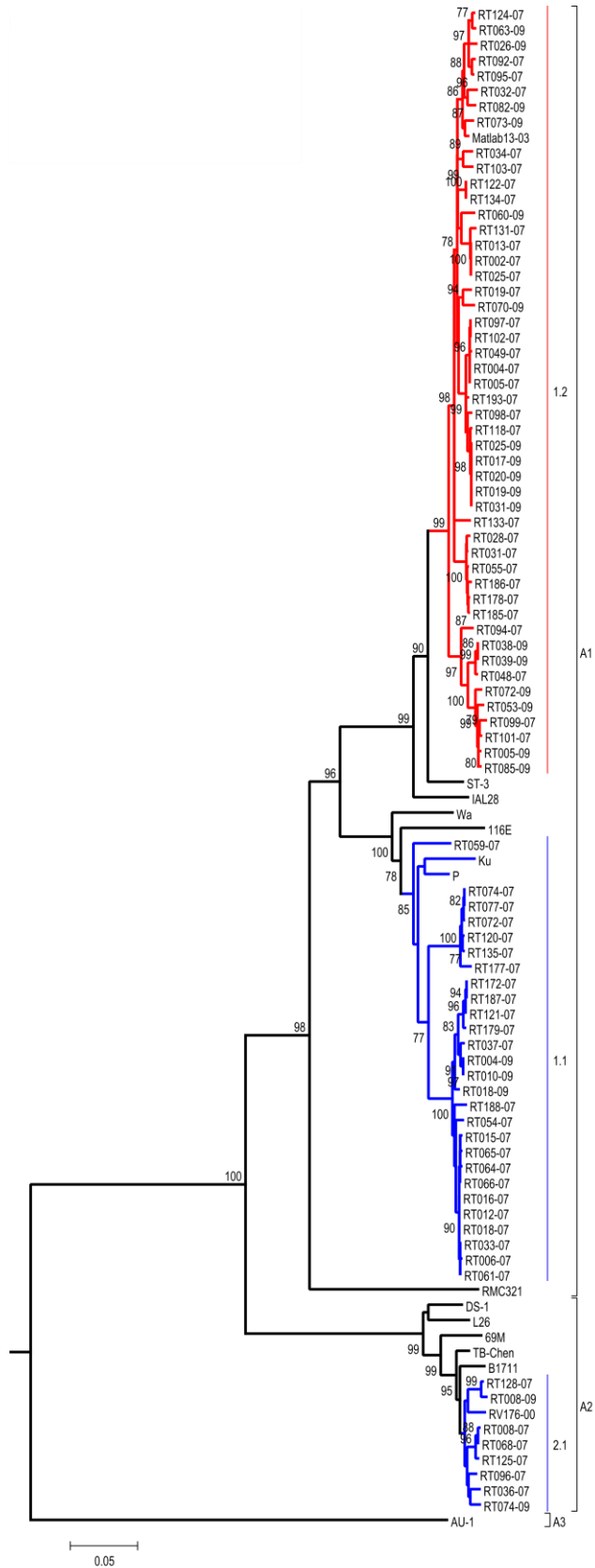


Figure 17: Neighbor Joining phylogenetic tree with 2000 bootstraps of the VP7 genome segment. Square brackets represent genotypes, while coloured lines represent sub-genotypic lineages. Measured in units of base substitutions per nucleotide.

NSP2:

Phylogeny of the NSP2 genome segment showed divergence of roughly 0.05 bases per site within sub-genomic clusters. Given a maximum of 0.05 base substitutions per site cut-off, 5 sub-genotypic lineages were made for the N1 genotype. Having 5 sub-genotypic lineages gives NSP2 the highest number of lineages. Lineage 1 is most closely related to the G12P[6] strain Matlab 13-03. Lineages 2 and 3 are quite distinct and are more closely related to 116E and Wa, respectively. Lineage 4 and 5 are closely related to each other and are most closely related to the P strain of rotavirus. Lineage 4, consisting of only the RT133-07 strain, is related to P but falls outside of the cut-off value to be included in sub-genotypic lineage 5.

All N2 rotaviruses are quite distantly related to the type strains DS-1 and TB-Chen. N2 strains are rather closely related to the G12P[6] DS-1-like strain RV176-00. Total variation within the N1 or N2 genotypes amongst Canadian strains was as high as 0.1223 base substitutions per site. A phylogenetic tree for NSP2 strains can be seen in Figure 18.

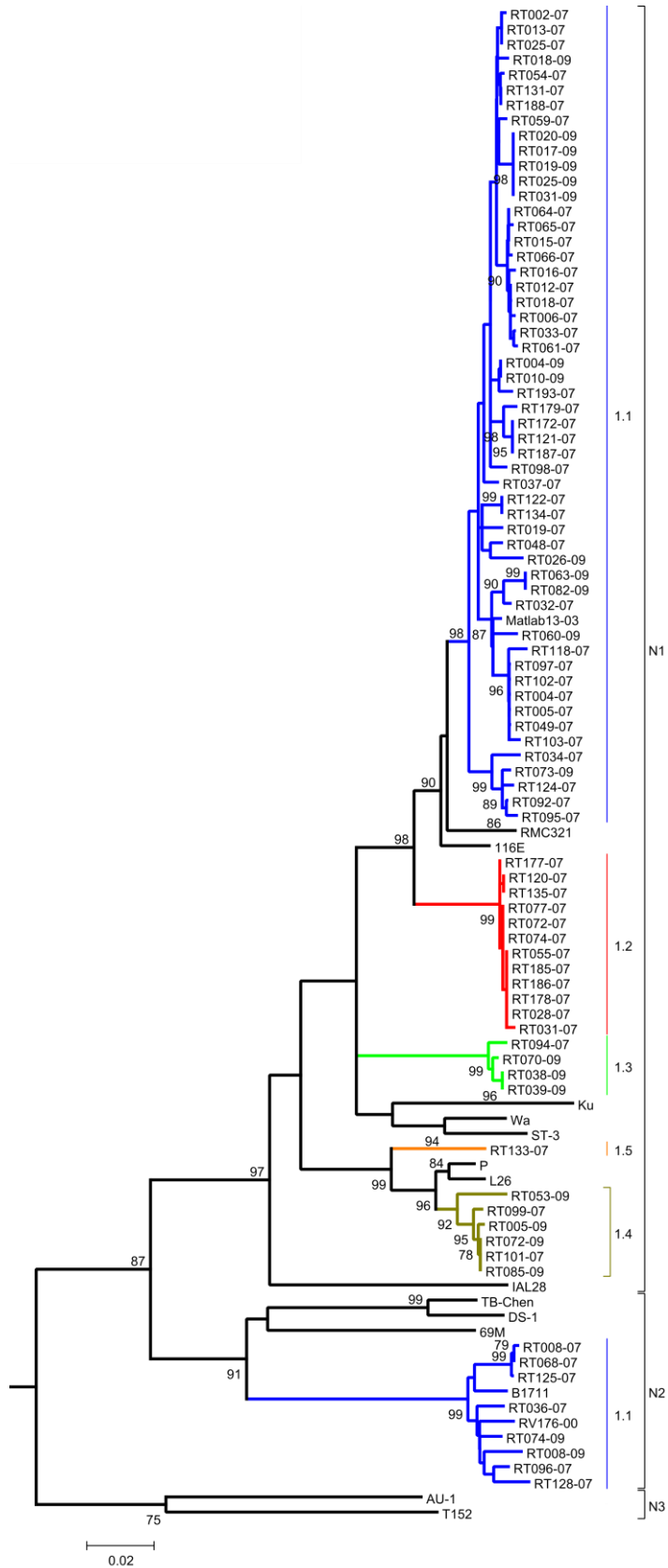


Figure 18: Neighbor Joining phylogenetic tree with 2000 bootstraps of the NSP2 genome segment. Square brackets represent genotypes, while coloured lines represent sub-genotypic lineages. Measured in units of base substitutions per nucleotide.

The NSP2 protein is a viroplasm coordinating protein. There is an affinity for NSP2 to bind RNA with its basic binding region between residues 205 and 241 [Patton et al. 1993]. Canadian rotavirus sequence data show that this region is highly conserved and basic with few variations within the motif. Variable sites include residues 205, 206, 215, 218, 230 and 239. Sites 230 and 239 differed in one sequence each. Sites 205, 206 and 218 were more highly variable but always contained neutral, non-polar amino acids such as valine, leucine and isoleucine. Variability at site 215 was genotype specific, changing from arginine in N1 strains to lysine in N2 strains. Other functional sites include a nucleotide binding motif at sites 221 to 227 and a magnesium ion binding region requiring glutamic acid at site 153 and tyrosine at 171. The motif between 221 and 227 is completely conserved and contains the previously described histidine residue at site 225 surrounded by charged amino acids. The sequence is not the previously described histidine triad (H-X-H-X-H-X-X) motif, as there is a H223K with complete conservation. The magnesium ion binding residues at sites 153 and 171 are completely conserved in Canadian strains, as shown in Figure 19 [Jayaram et al. 2002, Vasquez-Del Carpio et al. 2006].

		RNA-Binding Region																																					
																						HIT-Like Motif																	
Genotype		188	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241
Wa	N1	K	L	V	A	E	L	R	W	Q	Y	N	R	F	A	V	I	T	H	G	K	G	H	Y	R	V	V	K	Y	S	S	V	A	N	H	A	D	R	V
Ku	N1	K	L	I	A	E	L	R	W	Q	Y	N	R	F	A	V	I	T	H	G	K	G	H	Y	R	V	V	K	Y	S	S	V	A	N	H	A	D	R	V
RT004-07	N1	K	L	V	A	E	L	R	W	Q	Y	N	R	F	A	V	I	T	H	G	K	G	H	Y	R	V	V	K	Y	S	S	V	A	N	H	A	D	R	V
RT064-07	N1	K	L	V	A	E	L	R	W	Q	Y	N	R	F	A	V	I	T	H	G	K	G	H	Y	R	V	V	K	Y	S	S	V	A	N	H	A	D	R	V
RT099-07	N1	K	L	V	A	E	L	R	W	Q	Y	N	R	F	A	V	I	T	H	G	K	G	H	Y	R	V	V	K	Y	S	S	V	A	N	H	A	D	R	V
RT118-07	N1	K	L	V	A	E	L	R	W	Q	Y	N	R	F	A	V	I	T	H	G	K	G	H	Y	R	V	V	K	Y	S	S	V	A	N	H	A	D	R	V
RT133-07	N1	K	L	V	A	E	L	R	W	Q	Y	N	R	F	A	V	I	T	H	G	K	G	H	Y	R	V	V	K	Y	S	S	V	A	N	H	A	D	R	V
RT177-07	N1	K	L	I	A	E	L	R	W	Q	Y	N	R	F	A	V	I	T	H	G	K	G	H	Y	R	V	V	K	Y	S	S	V	A	N	H	A	D	R	V
RT060-09	N1	K	L	V	A	E	L	R	W	Q	Y	N	R	F	A	V	I	T	H	G	K	G	H	Y	R	V	V	K	Y	S	S	V	A	N	H	A	D	R	V
RT070-09	N1	K	L	V	A	E	L	R	W	Q	Y	N	R	F	A	V	I	T	H	G	K	G	H	Y	R	V	V	K	Y	S	S	V	A	N	H	A	D	R	V
RT085-09	N1	K	L	V	A	E	L	R	W	Q	Y	N	R	F	A	V	I	T	H	G	K	G	H	Y	R	V	V	K	Y	S	S	V	A	N	H	A	D	R	V
DS-1	N2	K	L	I	A	E	L	R	W	Q	Y	N	K	F	A	V	I	T	H	G	K	G	H	Y	R	V	V	K	Y	S	S	V	A	N	H	A	D	R	V
TB-Chen	N2	K	L	I	A	E	L	R	W	Q	Y	N	K	F	A	V	I	T	H	G	K	G	H	Y	R	V	V	K	Y	S	S	V	A	N	H	A	D	R	V
RT008-07	N2	K	L	I	A	E	L	R	W	Q	Y	N	K	F	A	V	I	T	H	G	K	G	H	Y	R	V	V	K	Y	S	S	V	A	N	H	A	D	R	V
RT125-07	N2	K	L	I	A	E	L	R	W	Q	Y	N	K	F	A	V	I	T	H	G	K	G	H	Y	R	V	V	K	Y	S	S	V	A	N	H	A	D	R	V
RT008-09	N2	K	L	I	A	E	L	R	W	Q	Y	N	K	F	A	V	I	T	H	G	K	G	H	Y	R	V	V	K	Y	S	S	V	A	N	H	A	D	R	V
AU-1	N3	K	L	V	A	E	L	R	W	Q	Y	N	K	F	A	V	I	T	H	G	K	G	H	Y	R	V	V	K	Y	S	S	V	A	N	H	A	D	R	V

Figure 19: NSP2 RNA-binding region alignment. Strains are organized by genotype. There is a reference type-strain for each genotype. The putative HIT-like motif with NTPase capabilities is highlighted in red.

Structural motifs in the NSP2 genome segment include four previously described cysteine residues at amino acid sites 6, 8, 85 and 285 [Patton et al. 1993]. In Canadian samples, the cysteine residue sites are consistent with those described previously and completely conserved.

#### NSP3:

T2 NSP3 segments can differ from T1 segments by as many as 0.3880 base substitutions per site. The vast difference between the two genotypes and their lack of reassortment supports the idea of reassortment between lineages being very rare due to incompatibility of reassortant protein-protein interactions.

Four sub-genomic clusters of T1 strains exist. Every strain within an NSP3 sub-genomic lineage is within 0.04 base substitutions per site from all others within the lineage. One strain, RT188-07, differed substantially from all other Canadian T1 rotaviruses, but not enough to be considered a new genotype. RT188-07 was shown to have 0.1318 base substitutions per site when compared pairwise to the type strain Wa. However, the strain did fall into its own sub-genomic lineage, being the most closely related to strain 116E. The T1 subgenomic lineages 1, 2 and 3 were all distinct but most closely related to the P rotavirus strain. All Canadian T2 strains were part of the same sub-genomic lineage most closely related to RV176-00. The tree for NSP3 can be seen in Figure 20.



Motifs analyzed for the NSP3 protein include a putative RNA binding motif, which is potentially a site used to downregulate cellular mRNA [Arnoldi and Burrone. 2009]. In addition, structural motifs and oligomerization motifs were examined. The motif described as the RNA binding motif was found with the methionine amino acid in a slightly different position in Canadian strains. In Canadian T1 strains the methionine residue was found at site 103, while had been previously described as one residue to the left of the second leucine base, at site 104 [Rao et al. 1995]. The amino acid sequence L-X-M-X-L-S-X-X-G was found in all Canadian T1 rotaviruses. The T2 RNA binding motif was similar; except in T2 rotaviruses, where the methionine at position 103 became a leucine and was followed by another leucine. Apart from genotype-specific variations, the RNA binding region and leucine zipper region were highly conserved.

Two completely conserved cysteine residues were found at residues 120 and 136. The other structural cysteines were highly but not completely conserved amongst T1 Canadian strains, while T2 strains contained no additional cysteine residues [Rao et al. 1995].

The oligomerization domain between residues 181 and 236 has conserved hydrophilic and hydrophobic domains with the T1 and T2 genotypes with few exceptions. However, the sequences are genotype specific, differing at many sites and definitely altering polarity between the two types [Rao et al. 1995].

NSP4:

Phylogenetic analysis of the NSP4 genome segment demonstrated 2 sub-genomic E1 lineages and three different Canadian sub-genomic E2 lineages. Most E1 strains



follow sub-genotypic lineage 1, which is most closely related to Matlab13-03 rotavirus. Lineage 2 E1 viruses are most closely related to the genotype-defining strain, Wa.

The E2 genotype strains were more diverse than the E1 strains. Strains RT036-07 and RT008-09 cluster together in sub-genotypic lineage 1, which is most closely related to the G8P[10] strain 69M. RT128-07 falls into Canadian lineage 2 by itself, being most closely related to the G12P[4] strain L26. The remaining five E2 rotaviruses fall into Canadian lineage 3, which is most closely related to the G6P[6] type strain B1711. A phylogenetic tree for NSP4 can be seen in Figure 21.

Domains analyzed within the NSP4 genome segment included hydrophobic domains, the oligomerization domain, the VP4 binding domain, the core binding domain and the enterotoxin domain. All of these described domains were not completely conserved, especially the VP4 binding domain. Several of the changes in the amino acid sequence were genotype specific.

All hydrophobic regions described were highly conserved. Variations were spontaneous and rare except for the E2-specific serine residue at site 19.

Analysis of the enterotoxin domain spanning residues 114-135 showed high homology in all NSP4 genotypes. Three residues showed minor variation from the dominant amino acid sequence, but these changes and slight variability had been noted in

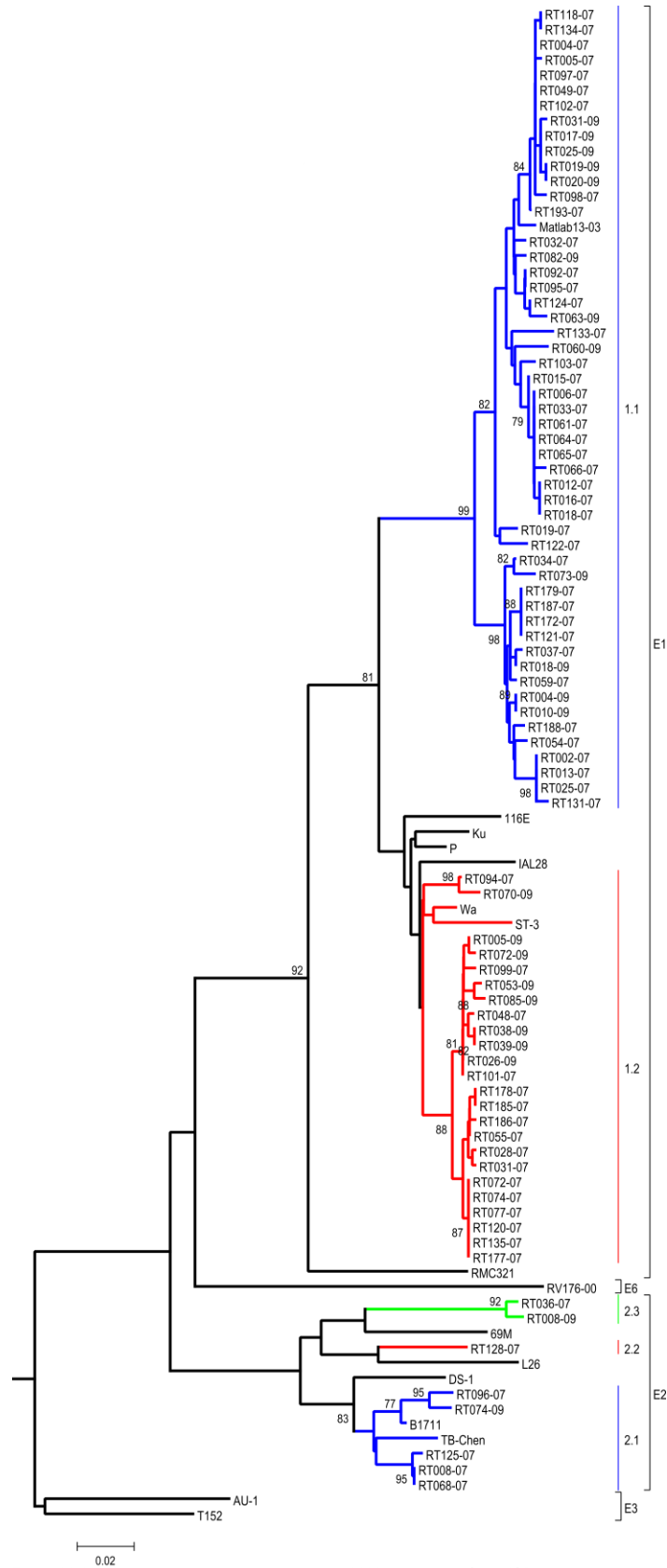


Figure 21: Neighbor Joining phylogenetic tree with 2000 bootstraps of the NSP4 genome segment. Square brackets represent genotypes, while coloured lines represent sub-genotypic lineages. Measured in units of base substitutions per nucleotide.

the past [Horie et al. 1997]. The final variations found in the enterotoxin region, N134K, were genotype specific, occurring only in E2 rotaviruses. The enterotoxin domain lies within the VP4 binding domain, which is fairly variable, in contrast to the enterotoxin domain. The variation of the domain spanning from amino acid 112-148 is greatest between sites 135 and 143, as shown in Figure 22. The variable region can be divided into groups by phylogenetic tree analysis. Patterns such as the aspartate-lysine combination, at sites 136 and 137 only occur in subcluster 2 of E1 genotypes, while the valine residues commonly occur at site 142 for subcluster 1 of E1 genotypes. E2 genotypes were highly conserved within the genotype but varied greatly from E1 strains between residues 135 and 143. The pattern of sub-clustering in the NSP4 E1 genotype strains (correlating with the VP4-binding domain variations) was used to group VP4 strains and predict which amino acids in VP4 may play a role in binding NSP4. Only one VP4 site, site 281, was synchronous with the variation pattern for the VP4-binding region of NSP4.

The oligomerization domain was also highly conserved, as shown in Figure 22. There were six sites of variation between residues 86 and 106, three of which were E2 specific. Variation within the oligomerization domain aside from the genotypic differences appeared rarely and strains did not contain more than one of the noted variations [Horie et al. 1997].

A final binding domain, the inner-capsid binding domain, was also analyzed for conservation. Unlike the VP4 binding domain, this one was conserved. There were five sites of variation noted, one of which was solely genotype specific. Other variations showed bias toward the E2 genotype, but also included some E1 genotype strains. These

mutations were commonly asparagine or isoleucine to serine conversions. Two of the variation sites changed rarely [Horie et al. 1997].

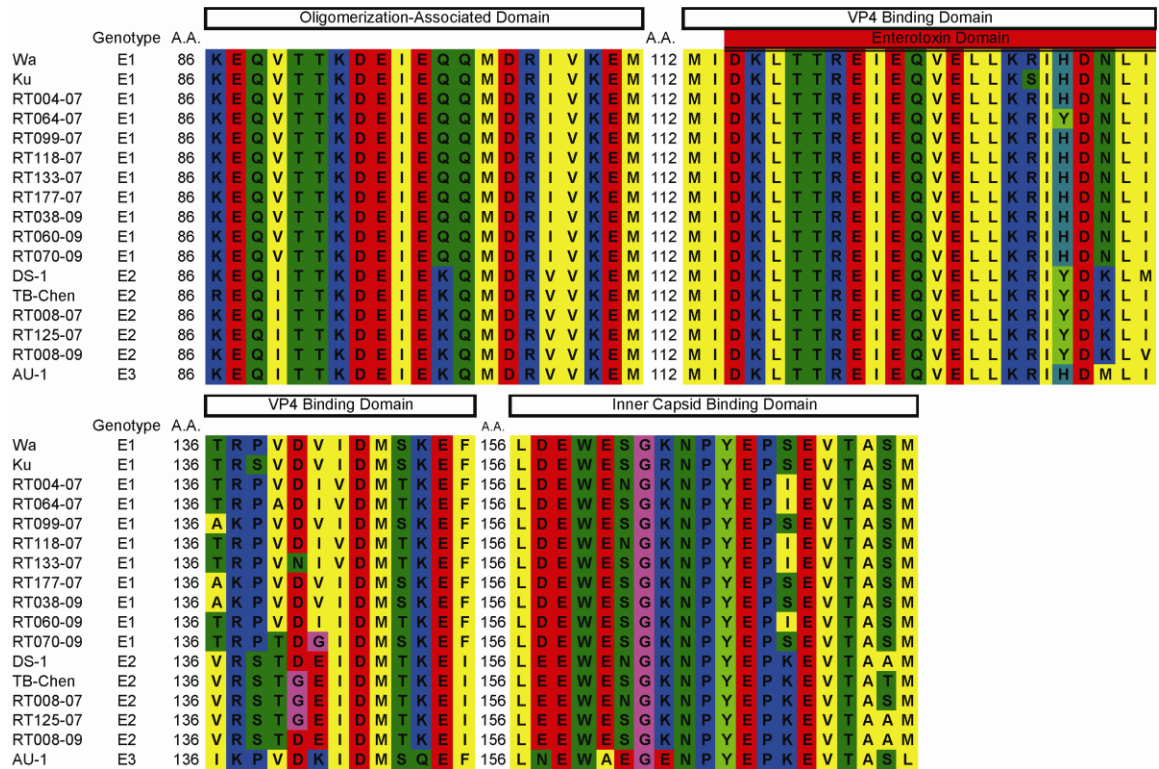


Figure 22: A representative amino acid alignment of five important NSP4 domains. The enterotoxin domain within the VP4 binding domain is depicted above, while the rest of the domain is depicted below. The oligomerization domain is important for functionality of the NSP4 protein. The inner capsid binding domain plays a large role in viral assembly during viral replicaion. Strains are organized by NSP4 genotype.

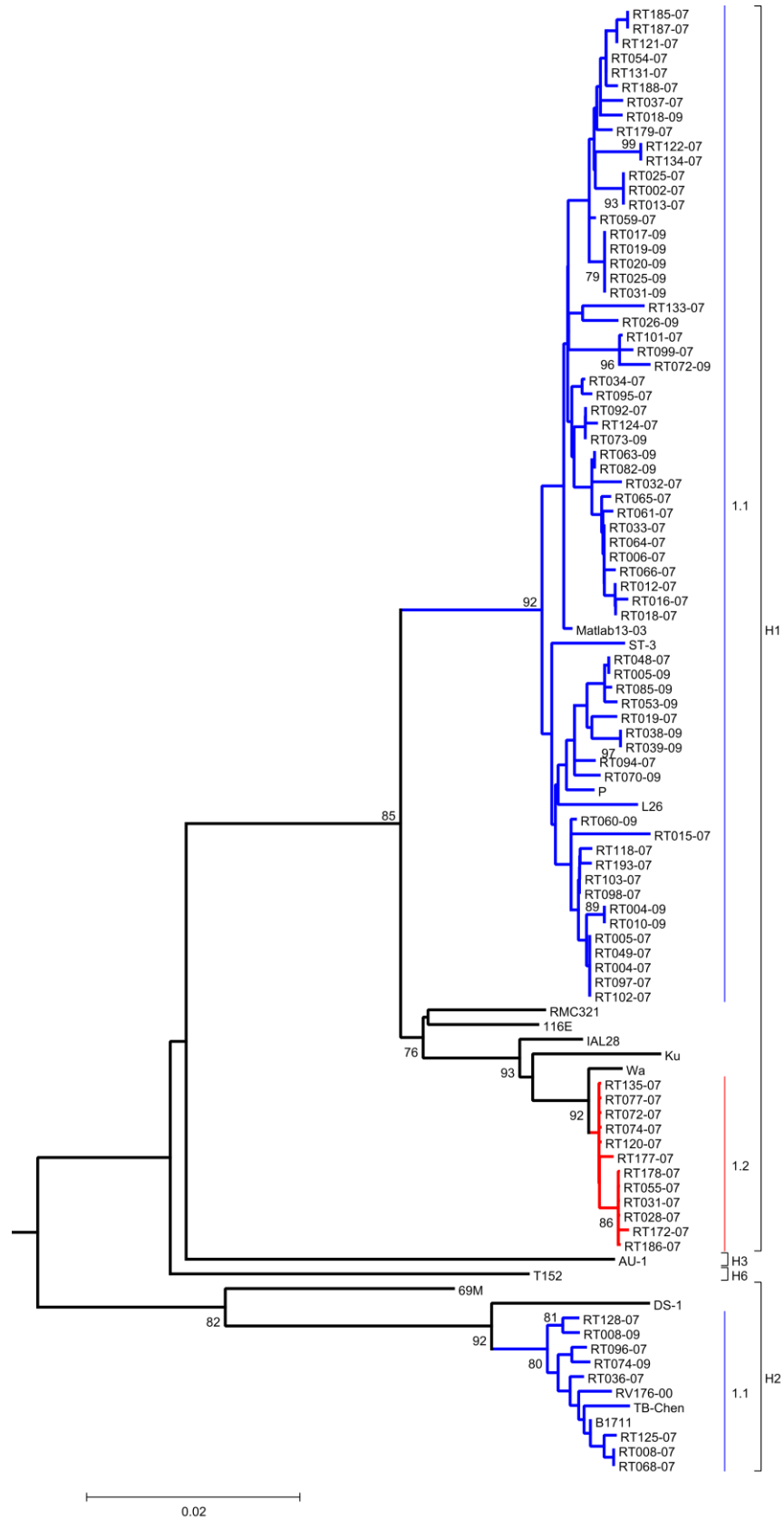
NSP5:

Phylogenetic analysis of the NSP5 genome segment revealed 2 H1 sub-genotypic lineages with 64 and 12 strains, respectively, as shown in Figure 23. The larger of the H1 lineages was highly identical to Matlab13-03, while the smaller was highly identical to the Wa type-strain. The two lineages appear quite distinct on the phylogentic tree, but differ by as little as 0.0338 nucleotide substitutions per site. The distinction

between the lineages is more attributable to the high identity between the strains within each sub-genotypic lineage.

Only one Canadian H2 lineage was shown in phylogenetic analysis. The H2 lineage was most highly identical to B1711, but also very highly identical to RV176-00 and TB-Chen. Canadian H2 strains are divergent from DS-1 in comparison to the other H2 type-strains included in the study, as seen in Figure 23.

The NSP5 genome segment was described as having two conserved structural cysteine residues and a conserved multimerization site. The functional cysteine residue at amino acid position 173 was completely conserved in all genotypes. However, position 170 was only conserved in the H1 genotype strains. The H2 lineage strains had an alternative cysteine residue at site 176. There were also H1 specific cysteine residue sites at positions 105 and 125 and an H2 specific cysteine residue at site 148. The previously defined multimerization domain, required for proper function of the NSP5 protein, was completely conserved in a genotype-specific manner. Two residues differed between H1 and H2 genotypes, within the residues 147-157 [Torres-Vega et al. 2000].



### 3.2.3 Analysis of Heminested Multiplex PCR Assays:

Genotypes were distinguished by heminested multiplex PCR assays, as described earlier. Genotypes were confirmed by partial sequence analysis from the genotyping incidence study and were often confirmed again by full genome sequencing. The whole genome study sequences were separated into genotypes in Mega 5.0 to assess the primer binding sites for those genotypes.

Only five G genotype-specific primers and two P genotype specific primers were assessed, as well as the end primers VP4F, VP4R, VP7F, VP7R, Con2 and Con3 [Gentsch et al. 1992, Gouvea et al. 1990, Iturriza-Gomara et al. 2000, Iturriza-Gomara et al. 2004, Simmonds et al. 2008]. Primers analyzed from the heminested multiplex PCR assay include those found in Table 11.

Upon assessment of the end primers, only a few mismatches were noted. Single mismatches to the VP7F primer were found in a G9 strain and a G3 strain at different sites. These strains were both 2007 isolates and the mutation did not reappear in any more recent isolates of any genotype. The VP7R primer had three common mismatches beginning with a transversion at the first nucleotide at the 5' end of the primer. The other two mismatches are transitions at sites complementary to the fourth and sixth thymine residues of the primer and are quite common.

The most common G type was G1. The alignment of Canadian G1 strains indicated that the target region for the currently used aBT1 (G1) primer is fairly conservative. However, there were 6 sites found with mismatches to fully sequenced G1 strains. Five sites were rarely mismatched, with four or fewer strains containing the

mismatch. The sixth site corresponded to the seventh base from the 3' end of the primer. The mismatch was very common, occurring in 67% of G1 strains.

The primer for G2 strains, aGT2, was assessed on only 8 Canadian strains but contained several mismatches. The target site for the G2 primer was highly conserved amongst Canadian G2 strains, containing only one variable site in the 25 base pair target region. However, the target site contained up to four primer mismatch sites, three of which were mismatches for all G2 Canadian isolates. In addition, there were only four strictly G2 genotype-specific sites in the target region. There were some G2-specific sites in the target region rarely shared by strains with other genotypes.

The G3 primer targets a highly conserved VP7 region among Canadian G3 strains. However, there is one site where the primer and Canadian strains are mismatched. The mismatch is a transition mutation at the sixth adenine in the primer, which occurs near the middle of the primer. There are three G3-specific sites in the target region, as well as several sites that are contained in all G3 isolates and only occur in other genotypes at low frequency.

Only three G4 genotype isolates were available to be analyzed for this study but they are highly conserved in the primer target region. The fifth guanine residue in the target region is the only variable nucleotide amongst the Canadian G4 isolates. Aside from the variable site in the G4 strain, there are three primer mismatches. The three mismatches occur in the first three residues at the 5' end of the aDT4 (G4) primer sequence.

The G9 primer targets a region that is highly conserved among Canadian G9 isolates. While there are only three genotype specific residues, there is high variability



among other genotypes in the target region. There are two bases in which there is variability in the Canadian G9 primer target regions. The variability only occurs in one strain, which could be an anomaly. There are no mismatches to the G9-specific primer if the two rarely-variable sites are not considered.

Similar analysis was used to assess the VP4 primers used to genotype Canadian strains. Two sets of primers were used for the RT-PCR step, and both were analyzed for mismatches. First, Con-3 is a VP4 forward primer developed based on the Ku strain of rotavirus, by John Gentsch and colleagues at the CDC. Upon comparison to an alignment of Canadian strains it was shown that the Con-3 target site is highly conserved. The Con-3 target site also contains very few mismatches in its target region. However, the target sequence is offset by a thymine residue found 10 bases from the 5' end of where the primer should bind. In addition, the target sequence lacks the fifth thymine residue. The sequence seems to have shifted in the middle of the primer target site. Two of the five residues between inserted and deleted residues are mismatched due to the shift. Other than the shift, the target sequence is highly conserved, with only three isolates containing rare single mismatches to the primer.

The Con-2 primer targets a fairly genetically stable sequence. Alignment of the Canadian isolates revealed five sites where primer-target mismatches occurred. Four of the sites were rare, and did not occur with other mismatches. However, one mismatch was more common. The second cytosine base in the target sequence often underwent a transition mutation to thymine.

VP4-F targets a highly conserved region in the VP4 genome segment. The targeted region contained only one rarely occurring mismatch near the 5' end. VP4-R

however, has five sites where transition mutations occur from adenine to guanine in the target sequence. Only one of the five potential mismatched sites is common for the VP4-R target region.

Two P-typing primers were analyzed for matching Canadian genotyping target regions, 1T-1DCDN and 2T-1 for P[8] and P[4] strains, respectively. The P[8] target site is variable in comparison to most genotype-specific target sites. The P[8] target site contains seven variable sites, five of which exist in more than 10% of Canadian P[8] strains. While the P[8] primer does have 4 degenerate bases, there are common mismatches that are unaccounted for. The area chosen for this primer has recently evolved to become fairly variable. P[4] strains analyzed show more variation in the P[8] target sequence and are more conserved, so it is not likely there are any false P[8]s.

The 2T-1 target site is far more stable than the P[8] target site. There are seven genotype-specific bases and the site is completely conserved among Canadian P[4] strains. The primer is well matched to the target site, and differs at only a single base.

Both G-typing and P-typing assays contained primers in the multiplex assay that were not fully analyzed in target sites since no Canadian strains contained genotypes matched to those primers. Other primers not analyzed but in the multiplex include: aAT8 (G8), G10, G12, 3T-1 (P[6]), 4T-1 (P[9]), 5T-1 (P[10]) and ND2 (P[11]).

### 3.3 Real Time PCR Assay:

A novel real-time PCR assay was developed for detection of rotavirus in stool sample extracts. Standard curves were created using 10-fold serial dilutions of RNA transcripts. RNA transcripts were 586 bp in length, derived from strains RT006-07 (C1)

and RT036-07 (C2). They were successfully amplified by the primers and Taqman probes designed for the assay. The serial dilution used for both standard curves was from  $5 \times 10^{10}$  copies to  $5 \times 10^3$  copies per reaction. The C1 probe detected the lowest dilution on the standard curve, while the C2 probe was able to detect standard templates of  $5 \times 10^4$  copies per reaction. Two positive rotavirus samples, RT018-07 and RT068-07, were used as positive controls for C1 and C2, respectively.

The real-time assay successfully detected rotavirus in all rotavirus-positive samples tested (rotavirus-positives were determined in the heminested multiplex PCR assay). All positive samples were those with Ct values falling within or greater than the range of the standard curve. Of 136 samples tested, which were previously genotyped as either G1P[8], G3P[8], G4P[8] or G9P[8], 136 were detected by the C1 probe in the VP2 real-time assay. Of 19 G2P[4] or G9P[4] genotyped rotaviruses, all were detected by the C2 probe in the real-time assay. Five samples with multiple genotypes were tested with the real-time assay. Two samples that were genotyped by heminested PCR containing G1, G2, P[8] and P[4], were detected only by the C1 real-time probe. Two others with the same multiple genotypes were detected by both the C1 and C2 probes. Two samples with G1 genotype and both P[8] and P[4] VP4 genotypes were detected with only the C1 probe. Finally, one sample that was genotyped, but could not previously be confirmed by sequencing to be G2P[4], was detected by the C2 probe in the real-time assay.

The assay was also validated using different reaction reagents. Ten successful real-time runs were completed in total. Four utilized the Ambion Ag-Path enzyme kit (Ambion, Carlsbad, Ca), made for a fast-real time method, while six runs were done using two Qiagen One-Step RT-PCR kits (Qiagen, Hilden, Germany), using the standard

real-time method. Qiagen One-step kits with different lot numbers were used to ensure there was no lot-to-lot variability. All assays except one successfully created standard curves with R-squared values of more than 0.99 on Ct vs log [copy number] graphs and replication efficiencies of greater than 90% and less than 110%. Examples of successful C1 and C2 standard curves and amplification charts can be found in Figures 24 and 25, respectively. One assay using the Qiagen One-Step enzyme kit failed to produce a curve for the C2 probe, likely due to pipetting error. Positive controls were included in every run and often with multiple replicates to ensure detection consistency. The C1 positive control, RT018-07 was run successfully 45 times and as many as 12 times in a single run. Similarly the C2 positive control, RT068-07, was run 43 times and as many as 10 times in a single run. Running several positive controls in a single run showed high consistency in detection of positive controls. Standard deviation of Ct values in triplicate standard curves was less than one in all runs. Average Ct values from run-to-run varied slightly, as the threshold level of each run was set to optimize standard curves.

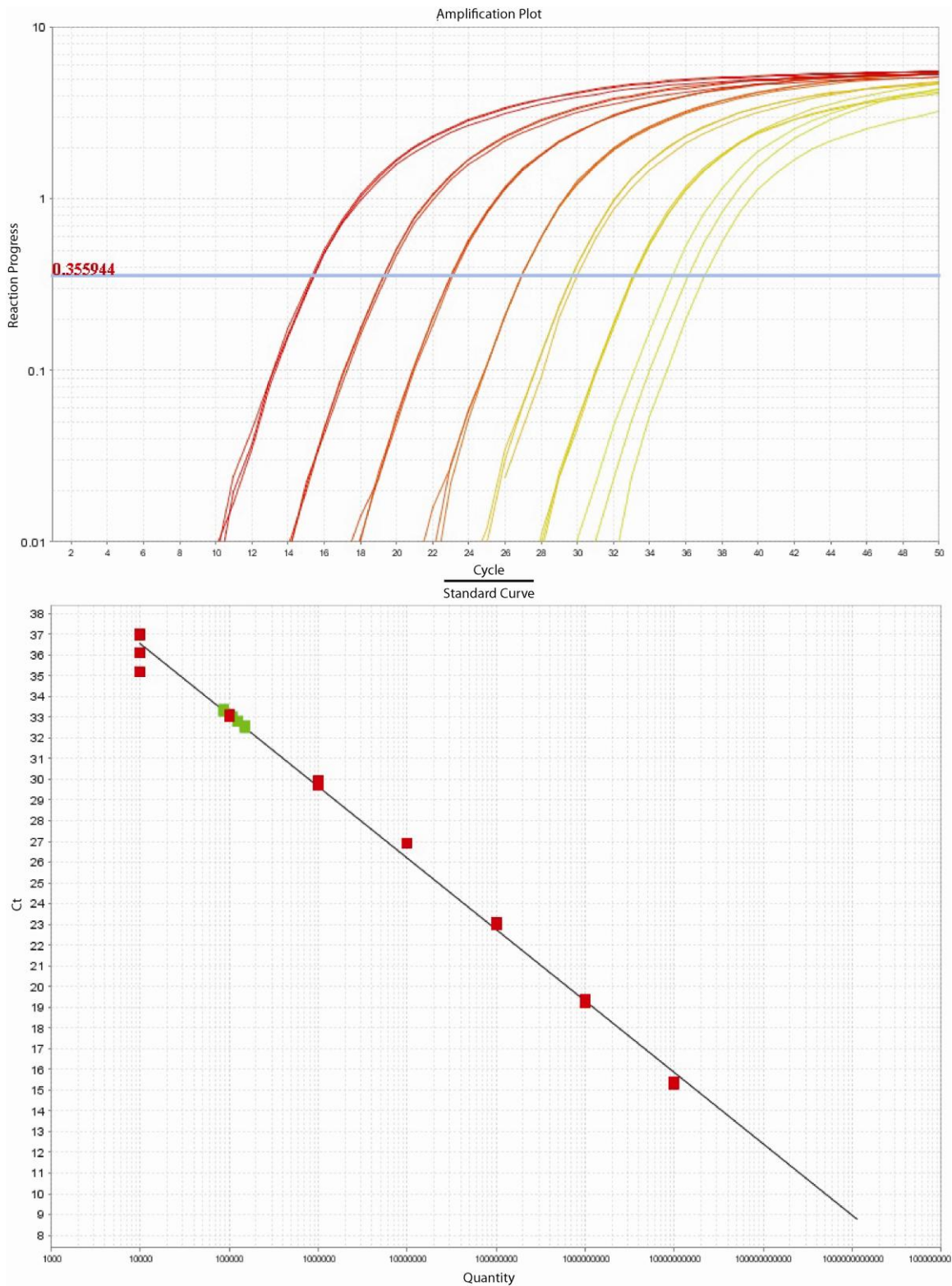


Figure 24: Amplification plot and Ct values for C1 standard curve samples. Above: amplification curves are smooth and well rounded and evenly spaced. Below: red boxes indicate standard curve samples while green boxes depict positive control samples.

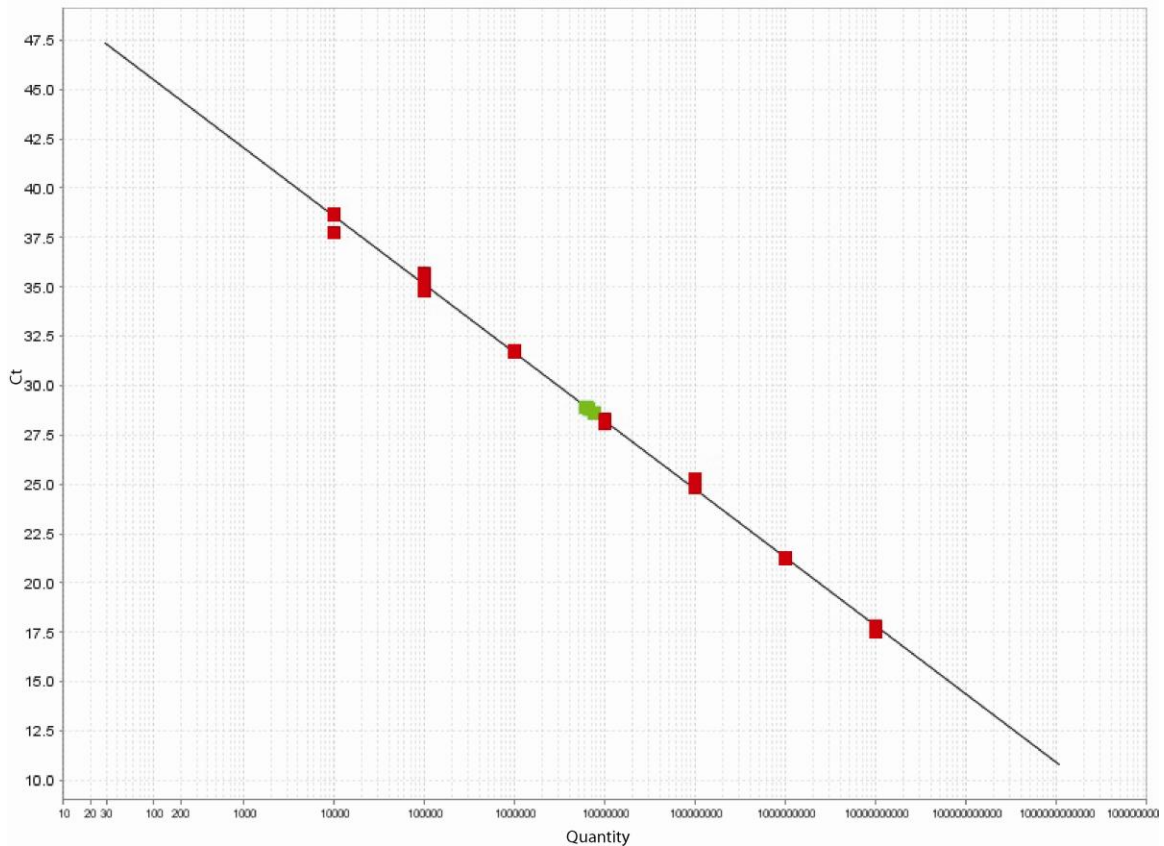
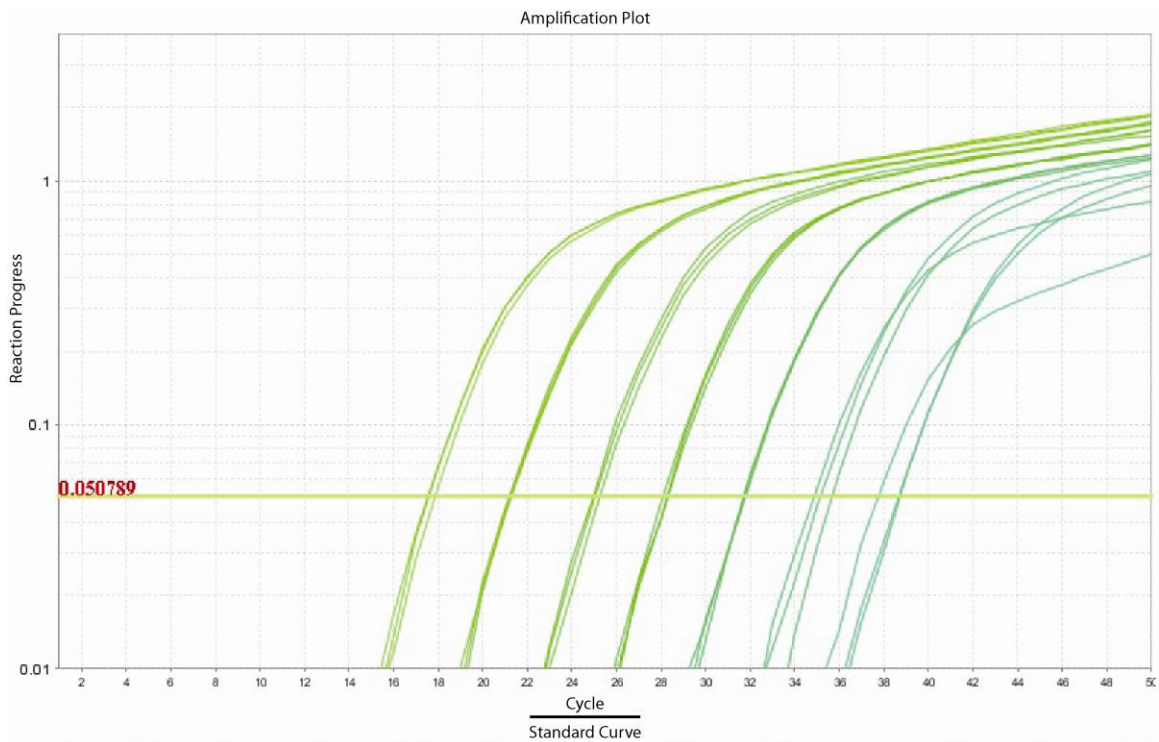


Figure 25: Amplification plot and Ct values for C2 standard curve samples. Above: amplification curves are smooth, well rounded and evenly spaced. Below: red boxes indicate standard curve samples while green boxes depict positive control samples.

## 4. Discussion

### 4.1 Heminested Multiplex PCR Genotyping:

Canadian rotavirus surveillance has revealed several trends of rotavirus genotype prevalence in Canada. The previous surveillance studies both showed that G1 strains were the most prevalent strains in Canada, followed by either G2 or G3 strains [Koustouros et al. 2003, Pang et al. 2004]. In the present study, 348 samples from 2007 to 2010 from a much wider range of sites across Canada were tested for the presence of rotavirus and genotyped. Samples came from a total of eight paediatric centres, but only two, Ottawa and Halifax, were surveyed in all four years. Additionally, Edmonton was surveyed in 2007, Winnipeg in 2007 and 2010, Quebec City in 2007 through 2009, Vancouver in 2010, Regina in 2010 and Toronto in 2010. Sampling from different sites in Canada resulted in a broader national picture of rotavirus strain prevalence. Of the 348 strains tested, 258 were positive for rotavirus, genotyped and confirmed by sequencing.

Surveillance showed that G1P[8] strains were the most prevalent genotype found in Canada, making up roughly 67% of all strains tested. G1P[8] strains were found in the highest prevalence in all years of surveillance, although that prevalence fluctuated greatly. The decline of G1P[8] prevalence, occurring in 2008 to 2009, from nearly 87% to 41%, is an excellent example of how different strains can emerge and decline on a season-to-season basis.

Genotype fluctuations were more evident in single city populations on a season-to-season basis. Halifax experienced a surge of G3P[8] strains in 2009. G3P[8] strains went undetected in the surrounding 2008 and 2010 seasons, but 17 of 22 (77.3%) samples collected from Halifax in 2009 were positive for G3P[8] rotavirus. On a related note,

G1P[8] strains were detected in 100% of 2008 Halifax samples (n=29). One year later, G1P[8] rotaviruses were found in only 3 of 22 (13.6%) samples. Likewise, G9P[8] rotaviruses were detected in Ottawa in 2009 at a prevalence rate of 34.3% (n = 35). Otherwise, from 2007-2010 G9 strains were only detected in 3 samples, none of which were from Ottawa.

Overall, six different genotype combinations of rotavirus were found in 258 positive samples. G1P[8] prevailed, being detected in as many as 86.7% of samples in 2008 (n=60) to as few as 41.4% of samples in 2009 (n=70). Canadian G1P[8] strains were also the only strains to be detected in every sampled city in every year. Furthermore, G1P[8] strains were the only samples found in 100% of samples at certain sites in certain years. Samplings with 100% G1P[8] detection include Winnipeg in 2007 (n=3), Halifax in 2009 (n=29) and Toronto in 2010 (n=3). G1P[8] strains are, by far, the most prevalent strain in Canadian samples and should be considered the main vaccine target in the Canadian population.

G2P[4] was consistently detected in all years. Although, G2P[4] frequency has not fluctuated as much as other genotypes, its prevalence has increased every year from 2008-2010. Perhaps this increase is due to reduced protection by the vaccines compared to other genotypes, which are creating mild herd immunity in Canadian populations, although vaccine uptake has likely been lower than 5% in most of Canada prior to 2011 [Dube et al. 2010]. G2P[4] strains were detected in a disproportionately low amount in Halifax throughout the study, being found in only 1 sample of 70 tested. In Edmonton however, prevalence of G2P[4] strains was 25% (n=16). There was no distinct pattern, as far as geography or time indicating how G2P[4] prevalence increased or decreased. Such



a pattern would be valuable during vaccine implementation, as G2P[4] does not have corresponding P and G types in Rotarix and only the corresponding G type in RotaTeq.

G3P[8] rotavirus strains have consistently been second in prevalence to G1P[8], but fluctuate substantially and frequently. The G3 strains were particularly prevalent in 2009, mostly to do with their emergence in Halifax. However, the emergence of G3P[8] in Halifax was short lived, and overall Canadian prevalence shrank to 20% (n=50) in 2010, from 31.4% (n=60), in 2009.

G4P[8] strains were found in low frequency in 2007 and 2009 and not detected in 2008 and 2009. The once common strain G4P[8] has seldom been detected in the Canadian rotavirus landscape. Canadian G4P[8] strains were only rarely found in other Canadian studies and only 2 of 258 samples were detected in this study [Koustouros et al. 2003, Pang et al. 2004]. While continuing to appear on occasion in Canadian surveillance, G4P[8] should not be considered a major rotavirus strain in Canada.

G9P[8] strains, although detected in 2007, only seemed to emerge as a prominent strain in 2009. The emergence was not sustained, as no G9P[8] strains were found in 2010. The lack of presence of G9P[8] strains quelled the suspicion that G9P[8] strains were emerging as a dominant strain due to an increasing effect of vaccine usage and herd immunity. This suspicion dawned from the fact that G9 was the only component not in the RotaTeq vaccine, when considering circulating Canadian G genotypes.

The final rotavirus strain detected, G9P[4], was found in Canada for the first time in 2010. Future surveillance will determine if this strain is going to emerge as the next major strain circulating in Canada. The threat of emergence is particularly concerning,

considering neither the G nor the P component of this strain are shared by presently available vaccines.

When all 258 genotyped Canadian rotavirus samples are considered in all years, G1P[8] was detected in approximately 67%, G2P[4] in 8.5%, G3P[8] in 17.5%, G4P[8] in 1.25%, G9P[8] in 5% and G9P[4] strains were found in 0.75% of samples.

Vaccine effectiveness against G9 strains remains to be seen. As sustained occurrences of G9 have not been established in Canada, it is not a reliable environment to study cross protection of either vaccine against G9. However, G9 strains seem to be appearing increasingly as they have been detected in 2009 and 2010. Also, a recent study in Detroit, a city bordering South-eastern Ontario, found G9 strains were the predominant circulating G-genotype between 2007 and 2009 [Abdel-Haq et al. 2011]. Of particular concern is the effectiveness of vaccines on G9P[4] strains. Neither vaccine contains the G or P genotypes of the newly detected Canadian G9P[4] strains. Immunity to rotavirus is strongly dependant on the two outer capsid proteins [Laird et al. 2003, Martella et al. 2006, Ward. 2008]. Surveillance of Canadian strains should be used to monitor emergence of any G9P[4] strains as vaccine use grows across Canada. If G9P[4] can escape vaccines, a new vaccine or vaccine component should be considered for future use. Future surveillance of Canadian strains will be necessary to determine if G9, especially G9P[4], will emerge permanently as one of the predominant Canadian strains.

Fluctuations in rotavirus prevalence are common and the causes are unclear, making it difficult to properly analyze the effects of a vaccine on specific genotypes of rotavirus. Only longer and more comprehensive surveillance studies in a vaccinated population will be able to determine if the vaccines will have an effect in reducing the

incidence of their respective genotypes in the human population. Unfortunately, vaccine coverage in the Canadian population is low, so a direct analysis of vaccine effectiveness will be difficult [Dube et al. 2010]. Low uptake may change in subsequent years as provincial health organizations are beginning to fund use of RotaTeq. Five of the thirteen provinces and territories are funding rotavirus vaccine as of December 2011 [PHAC and CNCI. 2011]. Presently, certain Canadian cities may be ideal populations to study herd immunity effects in the relatively highly vaccinated United States of America, especially in highly populated border areas such as South-eastern Ontario [Abdel-Haq et al. 2011, Tate et al. 2009].

RotaTeq and Rotarix vaccines have been found to have high efficacy in preventing serious rotavirus disease against G1, G2, G3, G4 and G9 strains [Dennehy. 2008]. In the current study, presently circulating rotavirus strains from 2007-2010 are all strains for which vaccines have been shown to provide protection. Likewise, past Canadian studies have shown prevalence of G1 strains, with no strains other than G1, G2, G3, G4 and G9 being detected [Kostouros et al. 2003, Pang et al. 2004]. These and other studies have demonstrated that licensed rotavirus vaccines should be effective in preventing rotavirus disease in Canada [Le Saux et al. 2011, National Advisory Committee on Immunization (NACI). 2008a].

## 4.2 Full Genome Analysis

### 4.2.1 Reassortment:

Reassortment of rotavirus genome segments in mixed infections is known as one of the principle mechanism by which genetic diversity can increase in rotaviruses [Ghosh

and Kobayashi. 2011]. Also, some genome segments, including VP4, VP6 and VP7, are more prone to reassortment than others [Iturriza-Gomara et al. 2003]. However, reassortment requires a coinfection to take place and the event is still not well understood. In this study, full ORF analysis of 84 fully genotyped rotavirus strains from 2007-2009, showed no unusual reassortant strains. Using McNemar's test it was shown that based on the genotype of the VP2 genome segment, there were no significantly different non-VP7 genome segments in reference to their respective genogroup (no evidence to reject the null hypothesis that the genogroups are the same as those found by VP2 genotyping). Strains that had all of their non-VP7 and non-VP4 genome segments clustering within a single genogroup are defined as belonging to a "pure genogroup" [McDonald et al. 2009]. All 84 Canadian strains were found to be "pure" rotavirus strains, despite finding 3.7% of samples coinfecting with both genogroups in the heminested multiplex PCR. However, many VP7 genome segments in Canadian strains were found to have varying genotypes, despite non-VP7 segments remaining in the Wa-like genogroup. VP7 genotypes found with Wa-like constellations included G1, G3, G4 and G9.

Amongst the 8 fully sequenced viruses with DS-1-like lineages, only G2 VP7 genotypes were found. However, two samples from the 2010 set of isolates had DS-1-like VP4 segments, but their VP7 segments fell into G9 rather than G2. These strains were tested in the real-time PCR assay developed for the VP2 genome segment. The G9P[4] samples contained only the C2 genotype which was found to have a 100% correlation with all other non-VP7 genotypes. The two G9P[4] strains found in 2010 were the first DS-1-like strains in this study not to have a G2 VP7 genotype. The two strains are either

the result of a reassortment event or have been undetected previously in Canadian surveillance studies. They were found in Saskatoon and Halifax, which are separated greatly geographically, yet the G9 sequences found have a higher identity to each other than any other G9 strains in this study. Also of note, the 2010 G9 strains are more identical to the only G9 strain detected in 2007 than any of the 12 G9 strains found in 2009. Although the strains are separated by a greater time period, the G9 strain from 2007 and the one from 2010 were detected in Halifax samples, while the 2009 group were found in Ottawa. The high identity of the G9 strains from the same city stresses the importance of surveying many areas in Canada, which may contain different lineages of rotavirus genotypes that may evolve in a divergent manner within the region. As 2010 was the first time Saskatoon was included in the genotyping study, it is uncertain whether this strain is a common circulating strain or a new virus in the area. Travel and population movement likely has a large role to play in the dispersement of new and unusual human rotavirus strains.

#### 4.2.2 Full Open Reading Frame Phylogenetics and Motifs:

Phylogenetic analysis and motif analysis of the whole rotavirus genome are powerful tools in fully characterizing Canadian rotavirus strains. Although no highly unusual strains or mutations were discovered, patterns in phylogenetic groupings and amino acid changes can be associated with each other. Rotaviruses from different times or areas can also be compared, in respect to phylogenetic and motif analysis. Genotype-specific amino acid arrangements were found to be present in all Canadian strains to some degree. For most genome segments (VP1-VP4, VP6, NSP1-NSP5) only two

genotypes exist. Genotype-specific amino acid patterns in these genome segments are either Wa-like strain-specific or DS-1-like strain-specific. As stated previously, the genotypes of all segments match in regards to being either Wa-like or DS-1-like. The strong genotype-specificity and consistency throughout segments could indicate two things: Canadian rotaviruses have evolutionarily distinct ancestral strains that rarely reassort, or all proteins interact with each other in a genotype-specific manner to function, organize and/ or assemble.

While reassortment events were not detected between genotypes, it is likely that they did occur between lineages found within genotypes. After organizing phylogenetic trees from every segment into sub-genotypic lineages, strains could be grouped together. Strains that matched in every lineage and VP7 genotype were considered clonal strains, while those with differing lineages in any segment were considered a new strain. While various cut-off values of nucleotide substitutions per site were used for every segment, the clusters of each sub-genomic lineage differed enough to be distinguished from other lineages. In total, the 76 Wa-like genogroup strains could be placed into 28 different clonal groups, while the 8 DS-1-like strains could be put into 4 different groups (Table 7). There were few instances when a strain clustered on its own. Since relatedness between two strains varied from one segment to another causing variable clusters within the genotype, it is likely that the variation is caused by ancestral reassortment events. These data supports the idea that reassortment events are a more common method of rotavirus genetic variation than genetic drift.

VP1:

Rotavirus VP1 proteins are the largest of the 11 or 12 proteins encoded by the genome. VP1 is the viral transcriptase and replicase protein, involved in both viral mRNA production and dsRNA genome synthesis. Evidence for these functions lies within several motifs that correspond to domains of RNA-dependant RNA-polymerases found in other viruses with positive sense RNA or double stranded RNA genomes. Moreover, evidence exists in the proposed structure of VP1, which seems to correspond to the polymerase core of many other viral polymerases with palm, fingers and thumb regions. A recent bioinformatics study revealed several motifs within the core region likely responsible for sugar selection and polymerization [Vasquez-del Carpio et al. 2006].

Five motifs were analyzed to compare previously described functional sites to those found in Canadian strains. The motifs described were in the catalytic core of the enzyme and should be completely conserved to preserve functionality of the protein. Three particularly important motifs, A, B and C, are found in almost all RNA dependant RNA polymerases [Vasquez-del Carpio et al. 2006]. In Canadian rotaviruses, motifs A, B, and C were completely conserved. It is not surprising that circulating Canadian rotaviruses are highly conserved in regions required for replication and transcription. Proposed functions of motifs A, B and C are magnesium ion coordination and sugar selection. Analysis of motif D revealed variation at two sites in three different strains. RT032-07 and RT048-07 vary at amino acid residue 665, while RT122-07 differs from the classical motif at site 666. The lack of conservation of this site may indicate that there is room for mild variation within the site, without compromising functionality. As only 3 Canadian strains of the 84 were found with variation in motif D and the mutations were

not sustained in future related strains, the 3 are likely mutagenic anomalies. Conservation within motif D, following the majority of strains, will likely be sustained in prevalent strains in the future. Motif F was the final motif found in Canadian strains, and was completely conserved in Canadian strains and matched previously described amino acid sequences and locations [Vasquez-del Carpio et al. 2006].

#### VP2:

The VP2 genome segment is an essential scaffold protein in rotavirus cores. It interacts with many of the rotavirus structural proteins, in particular VP1, VP3 and VP6. It has been proposed that VP2 interacts with VP1 and VP3 in its N-terminal domain. Analysis of the N-terminal domain between residues 1-156 showed a lot of genotype-specific interaction. As this protein is an essential scaffold protein and must interact with structural proteins VP1, VP3 and VP6, these genotype-specific changes likely correspond to genotype specific changes in both VP1 and VP2. Analysis of VP1 showed high conservation of the segment in several core functional domains, while genotype-specific changes occurred elsewhere. Folding of the VP1 protein allows several areas, all of which are likely genotype-specific, to interact with VP2 to form rotavirus cores. In addition, the high genotype-specific variability in VP3 proteins probably corresponds to the variability in the VP2 N-terminal domain, as well. Genotype-specificity is emphasized by several deletions in the C2 rotavirus N-terminal regions before the 50<sup>th</sup> residue. The specific interactions between VP2, VP3 and VP1 proteins are still unknown, but the variability in the interacting region is likely a key factor in the lack of reassortment in rotaviruses, especially in these proteins [Heiman et al. 2008].



Interestingly, when C1 rotaviruses were analyzed according to their G genotype it was found that there were four genotype specific changes between the G4 rotaviruses and the other G genotypes associated with C1 strains. G genotype-specific variation existing in the N-terminal region of the VP2 amino acid sequence could be a random evolutionary variation, or based on a selective relationship between the VP7 genotype and VP2 genotype. A VP6 amino acid alignment failed to reveal any G4 specific amino acid variation, which makes it likely that Canadian G4 strains vary from other C1 strain residues due to genetic drift rather than selective pressures provided by genotype specificity.

VP3:

The VP3 GTP/S-adenosyl-L-methionine binding region was conserved in the latter half of the domain. It is likely that this domain is highly conserved between amino acids 460-472 because these amino acids are functionally more important than the amino acids at 447-459. The first amino acids were more susceptible to general variation and genotype-specific variation. Past studies have confirmed this variation pattern among many rotavirus A strains and strains of different species [Chen et al. 2009]. Other motifs studied include the casein kinase II functional region, which is completely conserved in all strains. Interestingly, comparison of M1 and M2 strains showed genotype-specific amino acid variation at the second residue of this site, while all of the previously described functional residues were conserved throughout genotypes [Cook and McCrae. 2004]. This may indicate an evolutionary divergence between M1 and M2 segments in

Canadian strains. Finally, the guanylyltransferase motif was completely conserved in all Canadian strains, confirming the functional importance of this motif.

Structural proline and cysteine residues were also analyzed. The only proline not found of the 15 described was at position 502. Proline-502 likely represented a non-essential residue to the VP3 protein make-up. All others were completely conserved and three additional conserved prolines were found. All previously described cysteine residues were found, in addition to two new ones at sites 78 and 528 [Cook and McCrae. 2004]. The secondary and tertiary structure of this core protein is important for proper interactions with the two other core proteins. Structural amino acids such as the rigid proline or the sulphide-bond forming cysteines play key roles in VP3 structure and interaction. Additional studies based on full genotype surveillance or mutagenesis will reveal whether these residues play a large role in the structure of the VP3 protein.

VP4:

Functional sites previously described by Chen et al in 2009 including cysteine and proline residues shifted slightly in human rotavirus strains when compared to the lamb strains. The movements of the residues were not drastic and their importance is highlighted by their complete conservation throughout both human genotypes in this study. The internal fusion domain described as variable by Chen et al in 2009, was slightly variable in the human strains. There were four genotype-specific amino acid changes in addition to three rare variations throughout all strains. In this domain, conservation does not seem to be required at all sites, and the variation does not seem to

inhibit the essential functions of this domain. The variations noted in the Canadian strains in the current study were also noted previously [Chen et al. 2009].

Another important functional region is the cleavage domain. During the rotavirus life cycle, a complete rotavirus must undergo a trypsin-dependant cleavage between the VP5\* and the VP8\* regions of the VP4 protein. This cleavage enables efficient penetration after binding. Cleavage sites are found at arginine residues found at positions 230, 240 and 246 in Canadian rotaviruses. It has been shown using site-directed mutagenesis that only cleavage at site 246 is required, although all three sites are susceptible to cleavage by trypsin [Gilbert and Greenberg. 1998]. In Canadian strains arginines at sites 230 and 240 were completely conserved. However, the site required for cleavage, 246, varied in one Canadian strain. While strain RT120-07 did not contain the required arginine residue, it did have a lysine residue at site 246. Lysine is also susceptible to trypsin cleavage, so this mutation is likely to preserve the cleavage function of the VP4 protein [Gilbert and Greenberg. 1998]. Although rare, the lysine at site 246 has been noted in past studies [Arias et al. 1996]. Conservation of trypsin-susceptible amino acid residues at these sites signifies the importance of cleavage to enable viral entry into the cell. Although rare, the variation shown at site 246 to another trypsin-susceptible amino acid confirms the functional requirement for this site.

Initial attachment of rotavirus to the surface of cells may be sialic-acid dependant or sialic acid independent. Studies have shown that many animal rotaviruses contain a galectin fold with a carbohydrate binding site responsible for sialic acid binding [Dormitzer et al. 2002, Dormitzer et al. 2004]. However, most human strains are sialic-acid independent, including Wa and DS-1. As Canadian strains may have evolved from

Wa-like and DS-1-like strains in the past, it is likely that they are also sialic acid independent. Sialic acid independent rotaviruses likely require the galectin fold and neuraminidase-insensitive modified sialic acids or glycolipids for cellular attachment [Baker and Prasad. 2010, Dormitzer et al. 2002]. Analysis of the binding groove is important to monitor differences in sialic acid insensitive strains, and to determine how these relate to the type of putative receptors involved in cell attachment and penetration.

In the case of the Canadian strains, the amino acid sequence alignment of antigenic sites previously implicated in escape mutations in the VP8\* region is similar to that of other human strains [Zeller et al. 2011]. There is a shift in the amino acid sequence number due to a deletion early in Canadian amino acid sequences, but the amino acid positions are otherwise conserved with respect to each other. Conservation of amino acid position is an indicator that Canadian strains likely follow the function of other human strains, in that they have sialic acid independent hemagglutinin acting properties.

The outermost portion of the VP4 protein is the VP8\* portion of the protein. Antigenic binding epitopes were analyzed. Although VP8\* should not be sialic acid-dependent, the binding of rotavirus to intestinal cells is probably still dependent on this region. The antigenic regions analyzed in this study show a similar pattern of conservation and variation as an earlier study analyzing Belgian strains in the same way [Zeller et al. 2011]. There is high conservation of amino acid sequence in 5 of 25 antigenic amino acids in the VP8\* region. There are also sites with temporal trends of variation, such as 125, 131, 135, 149, 194 and 195. These trends can be convergent or divergent with the amino acid sequences of RotaTeq, as well as their tendencies to be

associated with a particular genotype. As these variations can be either convergent or divergent to the RotaTeq vaccine strain, it is probable that the observed variation was not caused by selective vaccination. This observation could also be due to the low vaccine coverage in Canada until 2012 [Dube et al. 2010, PHAC and CNCI. 2011]. Residue 182 was varied in only a single rotavirus strain. The lack of variation in site 182 and the 5 conserved sites (86, 87, 108, 179 and 193) is probably related to an important function related to cellular attachment. Many other amino acid substitution sites are genotype-specific. As RotaTeq and Rotarix contain only non-P[4] VP4 segments, it is not a surprise that many of the genotype-specific amino acid arrangements in Canadian P[4] strains are not the same as in the vaccine strains. In fact, in P[4] strains, three of four antigenic regions of the VP8\* protein contain at least one amino acid change known to cause immune escape when compared to either vaccine sequence. As vaccine use increases in Canada, P[4] strain amino acid sequence and prevalence should be monitored to ensure that the vaccines remain sufficiently well matched antigenically to provide protection.

The portion of the VP4 protein that makes contacts with the outer capsid layer, and with the rest of the virion, is the VP5\* segment. The VP5\* (or “foot” portion) of the VP4 protein is thought to be partly responsible for integrin binding and cellular entry, along with anchor functionality for the spike of the virion. Analysis of Canadian strains shows that VP5\* epitopes are completely conserved in 3 of 5 epitopes. This is in agreement with previous data showing that VP5\* is more highly conserved than VP8\*, indicating functional constraints in amino acid variability [Dormitzer et al. 2004]. Epitope 5-5 varies only in P[4] strains. This variation has been implicated as an antibody-escape mutation and differs from the 5-5 epitope of vaccine viruses. The 5-5 motif is the

beginning of a putative  $\alpha\beta$  heterodimeric, transmembrane glycoprotein (integrin) binding domain found in many group A rotaviruses. It has been implicated as essential for attachment and entry of rotaviruses into cells [Coulson et al. 1997]. Similar to the 5-5 domain, sites 387, 392 and 439 of the 5-1 epitope are genotype-specific and differ from vaccine strains. In addition, epitope 5-1 contains two residues that are conserved in Canadian strains but differ from RotaTeq and one that differs from Rotarix. The difference between vaccine strains at completely conserved sites of Canadian strains is evidence that vaccine strains are based on older strains. Monitoring of antigenic epitopes in future studies will be important as vaccine strains become increasingly outdated.

#### VP6:

The middle capsid layer of rotavirus is made up of VP6 trimers. Trimerization of the VP6 protein is important for proper protein function and antigenicity. Antibodies involved in detection and neutralization of rotavirus are capable of binding the trimerized form of the amino acid. VP6 is an important structural protein, since it interacts with both inner core and outer capsid proteins [Affranchino and Gonzalez. 1997, Heiman et al. 2008].

Trimerization of the VP6 protein is dependant on residues between 246 and 314. These amino acids are highly conserved in Canadian rotaviruses. There are only a few genotype specific differences between the two genogroups. Canadian strains follow a typical pattern of conservation in this essential domain.

Protein-protein interaction domains are essential for VP6 to properly interact with specific structural proteins in the rotavirus. Specifically, domain B is responsible for

interacting with VP2 and VP4, while protein H is responsible for interacting with VP7 and VP4. Throughout both of these domains and within the protein-protein interacting domains there are a steady number of genotype specific variations within the VP6 genome segment. The genotype-specific variation observed in VP6 is another example of an essential protein that is conserved within its genogroup, probably because mutations are constrained by the necessity to maintain an essential structural function.

VP7:

Phylogenetic analysis of the VP7 genome segment ORFs placed all Canadian strains within the five genotypes predicted by heminested multiplex PCR. G1 strains could be divided into two sub-genotypic lineages. Most Canadian strains fell into subgenotypic lineage 1, which was more divergent from the Wa and Ku type strains than lineage 2. G1 strains were most closely related to G4 strains and least closely related to G2 strains. The high variability between G1 and G2 is representative of the differences found in antigenic sites. G2 also diverges most from all other Canadian strain G genotypes. It is not surprising that Rotarix, which only contains a G1P[8] strain is less efficacious against G2 strains in clinical trials, especially because G2 is strongly associated with P[4] VP4 genotypes [Zeller et al. 2011].

Neither of the currently licensed vaccines carries G9 components. Canadian G1 viruses differ by roughly 0.35 nucleotide substitutions per site when compared to G9 strains, while G3 strains differ by roughly 0.31 nucleotide substitutions per site compared to G9 strains. In more recent years, G9 has been shown to be one of the most prevalent Canadian rotavirus strains. The fact that G3 is more closely related to G9 could mean that

some cross-protection is offered by G3 strains. When comparing vaccines, Rotarix contains a G1 virus, while RotaTeq includes both G1 and G3 genotypes. Both vaccines have been shown to be efficacious against G9 strains. However, efficacy against G9 may have some dependence on the VP4 genotype associated with the G9 strains, which is most commonly P[8].

It has been proposed that G9 reassorts more readily than other G types [Kirkwood et al. 2003]. Phylogenetic analysis reveals that G9 strains are more closely related to G2 strains, in comparison to the identity of G1, G3 or G4 strains to G2 strains. G9 is also relatively closely related to G1 strains, in comparison to other genotypes. The ability of G9 to reassort more readily may be due to its greater homology with both G1 and G2 than with other genotypes, which may allow the lineage 2 genogroup VP4 and VP6 proteins to associate more easily. More evidence for this comes from two 2010 G9P[4] strains which are highly identical to the 2007 G9 Canadian strain. It is possible that G9 may be the common ancestor of the other G types, as indicated by its place on a rooted phylogenetic tree. The G9 branch appears to be ancestral to all other rotavirus types circulating in Canada, as indicated by its shallow branch on the VP7 phylogenetic tree.

Antigenic motif analysis shows that VP7 genotypes differ greatly in their antibody binding regions. Although some variation occurred within genotypes, most variations were between different genotypes. Antigenic motifs were divided into three epitopes: 7-1a, 7-1b and 7-2. Variation in 7-1a was most evident when comparing the G1 genotype-containing Rotarix vaccine strain to Canadian G2 strains, for which there were 9 variable sites of 14 sites analyzed in the 7-1a region. Included in the 9 G2 variable sites were 6 variations at sites known to cause immune escape from monoclonal neutralizing



antibodies. A comparison of Canadian G2 rotaviruses and the G2 component of RotaTeq revealed 4 variations, 2 of which were known to be escape mutations that were conserved throughout the genotype. Other strains also contained conserved variations with respect to their RotaTeq components. In fact, all Canadian strain genotypes, except G3 strains, contained a conserved site which varied from RotaTeq at antigenic residues in the 7-1a antigenic region. Antigenic amino acid differences between RotaTeq components and matching genotypes, such as G2, demonstrate the possibility of evolution causing immune escape as the vaccine strains diverge over time from their respective genotypes circulating in the population. With the exception of some G1 strains, all Canadian strains contained at least two neutralizing monoclonal antibody escape mutations when compared to the Rotarix vaccine. It is important to note that vaccine components induce protection from rotavirus with different groups of immunoglobulins, so while immune escape mutations are important to monitor, sufficient serological cross-protection may still be induced by a vaccine.

The second of the G type antigenic epitopes analyzed was 7-1b, in which 4 of the 6 amino acid sites are known to be sites of escape mutations. Canadian G1 strains were well matched to both vaccines in this epitope. Only a single Canadian strain (RT072-07) varied from the vaccine strains in the first 5 amino acids of this epitope. Furthermore, RT072-07 had a substitution at site 212, which is not associated with immune escape. Strains from other genotypes had fewer matching amino acids in their 7-1b epitopes as compared to vaccine strains, but none differed more from the vaccine strain than the Canadian G3 strains. G3 strains differed from their respective RotaTeq component in three sites and from the Rotarix 7-1b component in four sites. G2 strains varied the most

within their genotype and had up to three differences from Rotarix and two from RotaTeq, and at least one of each was an immune escape mutation. With the exception of most G1 strains, all Canadian rotaviruses varied from their respective RotaTeq vaccine components with one immune escape mutation. Once again, many variations were noted in circulating rotavirus strains in comparison to the RotaTeq vaccine components. This is an indication that rotaviruses may have evolved towards vaccine escape since the introduction of the vaccine.

Finally, analysis of the 7-2 epitope showed that Canadian G2 and G4 strains differed the most from Rotarix in their 7-2 epitope, while many G1 strains were well-matched. With the exception of some G1 residues, epitope 7-2 was highly conserved within respective genotypes and differed highly between genotypes. In general, RotaTeq components were well-matched to most genotypes with the exception of one or two mutations in some Canadian G1 and G4 strains.

Analysis of the three main antigenic epitopes showed that G9 strains were more closely related to G3 strains than other G genotypes, differing in 8 of 29 epitope sites. Protection against G9 strains by Rotarix should be approximately equal to its protection against non-G1 strains. Cross protection by RotaTeq is more difficult to estimate because components of RotaTeq match portions of G9 strains. Cross-protection against G9 strains is likely provided by the G4 component of RotaTeq which shares 19 of 29 residues with G9 strains, mostly in the 7-1a and 7-1b epitopes.

As previously stated, RotaTeq and Rotarix were highly efficacious against circulating strains of rotavirus during their clinical trials. This study shows that there are sites where circulating strains may contain escape mutations that could reduce immunity

induced by vaccines. Although it is expected that vaccines still provide good cross-protection against all rotaviruses regardless of group A genotype, it is important to monitor the accumulated changes in antigenic regions of rotavirus over time in order to identify further escape mutations. Presently, Canadian strains are likely to be highly susceptible to neutralization by neutralizing antibodies induced by the current vaccines, especially RotaTeq. Future studies will be useful to determine if and when new vaccines should be updated.

Heminested multiplex PCR genotyping has indicated that G1 are the most prevalent strains circulating in Canada, while G3 is the second in prevalence. Both vaccines provide very good protection against these strains and it is unlikely that the minor mutations identified in this study will result in immunological escape. However, G2 strains are highly divergent from Rotarix and G9 strains are not well matched by either vaccine. In addition, G2 strains are almost exclusively associated with P[4], a VP4 genotype not contained in either vaccine. G9 strains are known to recombine more easily than other genotypes and have recently been found with P[4] genotypes in Canada [Kirkwood et al. 2003]. G2 and G9 strains should be monitored closely as vaccine usage increases. If surveillance demonstrates a sustained increase in G9 or G2 prevalence, re-evaluation of the antigenic composition of currently used vaccines and possibly updating the vaccines should be considered.

NSP1:

NSP1 is the largest of the non-structural proteins. While found throughout the cytoplasm during infection, it is considered the only non-essential protein because

rotavirus can replicate in its absence [Graff et al. 2002], and is subject to a high mutation rate that may or may not affect viral function. Although the function of NSP1 is not entirely clear, there is speculation that it is involved in viral RNA-binding capabilities, as well as possibly having an effect on interferon regulatory factor 3 (IRF3) and E3 ubiquitin ligase [Heiman et al. 2008]. It is known that NSP1 has zinc binding activity and can bind to virus-specific RNAs [Ito et al. 2001]. It is thought that the RNA-binding domain may be a cysteine-rich motif. Analysis of an alignment of Canadian strains showed that there is a conserved cysteine-rich motif within the NSP1 genome segment, although it differs slightly from those previously described [Hua et al. 1993, Ito et al. 2001]. There is possibly room for variation to occur within this site without the loss of function. The cysteine-rich domain pattern, C-X<sub>2</sub>-C<sub>2</sub>-X<sub>7</sub>-C-X<sub>2</sub>-C-X<sub>3</sub>-H-X-C-X<sub>2</sub>-C-X<sub>5</sub>-C, is conserved in all Canadian strains. The purpose of this binding region may be to associate the viral RNA with host translational proteins.

It has also been proposed that NSP1 interacts with the cellular factor IRF3 at its C-terminal end, in combination with a zinc finger at the N-terminal end. This interaction may play a role in reducing the interferon-mediated antiviral response [Graff et al. 2002]. In Canadian rotavirus strains, the zinc finger, which is defined by cysteine residues at the N-terminal end between residues 42 and 72, is well conserved [Graff et al. 2002]. The proposed zinc-binding cysteine residues at sites 54, 57, 63 and 66, in particular, are completely conserved among all genotypes. The finger contains several sites that rarely contain mutations and three that vary more frequently in Canadian strains. There are also 5 genotype-specific variations within the zinc finger. The C-terminal domain acts in

concert with the zinc finger to bind IRF3, which models the IRF3 interacting  $\alpha$ -helix. The C-terminal domain is not well defined and not well conserved in Canadian strains.

NSP2:

Sites 205 to 241 of the NSP2 protein are thought to be an RNA binding domain based on the proposed  $\beta$ -strand formation of the sequence [Patton et al. 1993]. In addition, the region is highly conserved with basic residues and only 6 variable sites, one of which was found to be genotype-specific in Canadian strains. High conservation indicates that this site may play an important role in viral function. Within this site is a putative Histidine-triad (HIT-like) motif proposed to have magnesium ion binding capabilities. The HIT family of proteins usually function as nucleoside hydrolase/transferase proteins [Kumar et al. 2007]. The motif between bases 221 and 227 does not follow the traditionally required HIT signature motif, His-X-His-X-His-X-X, and does not have high identity to other HIT proteins. However, when structures are superimposed upon each other, it does contain a similar histidine site at residue 225 that is apparently similar to histidine-112 of the protein kinase C interacting protein (A HIT protein) [Jayaram et al. 2002]. The NSP2 protein functions in covalently binding  $\gamma$ -phosphate from nucleotides, in a magnesium-dependant manner, unlike HIT proteins, which bind  $\alpha$ -phosphate independent of magnesium. It has been suggested that this nuclease activity is an energy transduction mechanism acting as a molecular motor for dsRNA synthesis in the viroplasm, as NSP2 is the only rotavirus protein with NTPase activity [Jayaram et al. 2002, Kumar et al. 2007]. The NSP2 protein sequence is potentially very important in coordinating genome replication and packaging. Alignment of Canadian strains confirms

high homology in the putative RNA-binding region, as well as in the putative HIT-like protein region, despite relatively high variation within and between strains in areas surrounding this domain. The conserved sequence found in Canadian strains between residues 221 and 227 is H-X-K-X-H-X-R. The lysines at positions 188 and 223, as well as arginine at position 227 have also been shown to be involved in the function of NSP2 [Kumar et al. 2007]. These residues are also conserved in Canadian strains. The magnesium binding sites involved in NSP2 function are also completely conserved at sites 153 and 171. Canadian NSP2 proteins are conserved at all amino acids thought to be required for NTPase function. Although the exact function and mechanism of NSP2 is not fully understood, the conservation of these sites suggests that proper function of NSP2 is dependant on these amino acid sequences.

NSP2 is known to have conserved structural cysteine residues at positions 6, 8, 85 and 285 [Patton et al. 1993] and all of those cysteines were completely conserved in the viruses analysed in this study.

NSP3:

The NSP3 protein showed a high degree of conservation in important motifs despite the variation between genotypes. One genotype-specific amino acid substitution was found in the RNA binding motif, but it was otherwise highly conserved. The RNA binding motif is thought to bind cellular RNA; therefore, it is likely that the site should be conserved intergenotypically. The genotype-specific residue within the motif is likely to be ancestral rather than mutation related.

Complete conservation of previously described cysteine residues highlights the importance of these residues to proper function of the protein. Two other cysteine residues were also found in T1 genotype strains, but it is unclear if they have any functional significance.

Oligomerization motifs were highly conserved within genotypes, but differed between them. This is possibly another example of how differences in protein-protein interactions could inhibit intergenotypic reassortment.

#### NSP4:

The NSP4 protein has many important domains. NSP4 has a hydrophobic trans-membrane domain for integration into the endoplasmic reticulum. It also contains domains for association with the other viral proteins such as VP4 and VP6. Part of the VP4 binding domain is the enterotoxin domain. It has also been suggested that NSP4 may play roles in viral pathogenesis such as intracellular calcium recruitment, tubulin interaction, recruitment of DLPs to the ER, chloride secretion and VP4-related TLP assembly [Arnoldi and Burrone. 2009, Ball et al. 2005, Heiman et al. 2008, Horie et al. 1997, Okada et al. 1984].

The hydrophobic domain is highly conserved except for residue 19, which demonstrates genotype-specific variation. The enterotoxin domain is also highly conserved. The only mutations noted in this domain were rare and similar to those reported by Horie et al [Horie et al. 1997]. The enterotoxin domain plays a role in viral shedding and spread, and is highly conserved, possibly in order to allow binding to specific host receptors during pathogenesis. Unlike the enterotoxin portion of the VP4

binding domain, the latter 13 amino acid portion (136-149) is variable in many sites. Several sites appear to be genotype-specific, but some genotype-specific variations are not shared by all E2 strains. Once again, genotype-specificity in the latter portion of the VP4 binding domain could be the reason for the lack of reassortment of the NSP4 genome segment between Wa-like and DS-1-like strains. However, evidence of NSP4 reassorting independently of other genome segments has been shown, and many interspecies reassortants have been found [Ball et al. 2005].

The final 20 amino acids in the NSP4 protein represent the VP6 binding domain. Although the domain in the VP6 protein, which interacts with NSP4 has not been identified, the association between NSP4 and VP6 that occurs during viral assembly is well documented [Ball et al. 2005, Heiman et al. 2008]. The VP6 domain is also part of the tubulin binding domain. In the case of Canadian NSP4 strains these domains are highly conserved. There is one NSP4 lineage-specific mutation at residue 169 within the VP6 binding domain. In addition, there are a few genotype-specific variations, but the residues are highly conserved within binding domains.

#### NSP5:

NSP5 is the shortest of the genome segments with the shortest amino acid sequence in this genetic analysis of Canadian rotaviruses. The NSP5 genome segment sometimes includes a second ORF encoding the NSP6 protein. However, NSP6 is non-essential and absent in many rotaviruses. This study showed only genogroup I viruses (DS-1-like) had NSP6 ORF regions. NSP5 is very important to proper viral function, particularly during assembly of core proteins within the viroplasm.



NSP5 proteins must multimerize and associate with NSP2 and ssRNA, to carry out their task of viral protein accumulation within the viroplasm. Multimerization is thought to give the NSP5 kinase function and the ability to become hyperphosphorylated and to interact with viral components. NSP5 has a multimerization domain defined by amino-acid residues 187-197, at the C-terminal end of the protein [Torres-Vega et al. 2000]. In Canadian strains, these residues were conserved within genotypes, but contained two genotype-dependant differences. There were two variable sites between genotypes. The high conservation was not surprising because of the suggested requirement for this motif to oligomerize. Oligomerization activates kinase activity, to hyperphosphorylate and to bind to NSP6. As suggested previously, there were no threonine or serine residues in the multimerization domain despite the high frequency of those amino acids elsewhere in the protein [Torres-Vega et al. 2000].

#### 4.2.3 Summary of Phylogenetic Analysis:

Canadian strains represent a unique population of rotaviruses that vary in relatedness to previously described strains. The majority of genogroup II genome segments are related to Matlab 13-03, while genogroup I segments are more frequently related to RV176-00. However, Matlab 13-03 differs from most Canadian strains in the VP3 segment, falls into the T2 genotype for the NSP3 genome segment and is a G12 strain, when analyzing the VP7 genome segment. RV176-00 is also a G12 strain, and diverges substantially from Canadian strains in the NSP4 phylogenetic tree.

The use of Wa, DS-1, P and ST-3 as type strains in heminested multiplex PCR are useful as positive controls. However, in the future the assay may need to change to

accommodate genetic drift. These strains are suitable for VP7 and VP4 genotyping positive controls, but are not representative of currently circulating Canadian rotaviruses.

#### 4.2.4 Assessment of Current Heminested Multiplex PCR Methods:

Heminested multiplex PCR genotyping is a powerful and fast tool for genotyping rotavirus strains. However, this technology relies on highly conserved genotype-specific regions and suitable primers to complement these regions. Using the sequencing data from full genome analysis, the primer-binding sites were assessed to elucidate whether the currently used assays are suitable for continued use.

The first round of genome replication involves two “parent” primers that encircle the unidirectional genotype-specific primers used in the second stage. The primer VP7F is still a very good primer, with no consistent mismatches. However, VP7R contains three mismatched sites that are quite common. The VP7R binding site should be monitored periodically in the future to ensure it still matches the primer.

Con3 is another primer with a highly conserved binding site. Con3 binds the VP4 genome segment with an extra base. As the mismatch does not relate to a frameshift of the translated VP4 sequence, it is part of the original primer. Con3 also contains a common mismatch located 10 bases from the 5' end. Mismatches in the middle of a long primer like Con3 are less worrisome than mismatches at the end. Primer stability can be maintained as long as the ends of the primers bind well. Therefore, the common mismatch in Con3 is not considered a serious problem. Likewise, Con2 has a common mismatch at the second cytosine base, which lies 9 bases from the 5' end of the primer. While additional mismatches could cause primer failure, this mismatch is unlikely to

cause any primer problems due to its proximity to the primer ends. Four rarer mismatches occurred in Con2 and three in Con3. The rare mismatches are not recurring and likely just anomalies in the genome segments that should be noted, but should not affect primer function.

The final two parent primers are VP4F and VP4R; both target highly conserved regions. VP4F, in particular, targets a site with only one rare mismatch. The target site for VP4R is more questionable, containing 5 potential mismatch sites, one of which is common. A single fairly common mismatch and several uncommon mismatches will not affect the primer. These newer versions of VP4 parental primers are highly effective, although, they negate the detection of P[11] strains, since the P[11] primer binds too close to VP4F.

Mismatches found in parental target sites are of particular concern in the detection and genotyping of rotaviruses. If target sites of parental strains are mutated to the point of primer failure and the assay is used for confirmation of rotavirus presence, rotavirus may go undetected. Currently analyzed mismatches are occurring in successfully genotyped strains, so the primers still appear to work. However, in using parental strains as rotavirus detection tools an alternative assay should be used as a confirmatory assay to ensure no false-positives were missed due to a greater degree of mismatch or lower virus concentration. A real-time PCR assay was designed to provide a sensitive test for rotavirus detection.

Typing primers were analyzed for mismatches to determine if they were still specific enough to be used in genotyping assays. The target for the G1 primer contained five rarely mismatched bases in the 84 strains analyzed. While the rarely mismatched

bases are not likely to cause primer failure, a combination of one or more of them perhaps in conjunction with the common mismatch may lead to mistyping. The common G1 mismatch occurred at 7 bases from the 3' end of the primer. Presently, this mismatch has not caused mistyping and is not a concern. Future surveillance studies should consider the use of degeneracy at the site to prevent mistyping G1 strains, if mistyping of G1 is suspected. Non-G1 genotypes from the fully genotyped group were highly mismatched to the G1 primer, so false G1 positives among strains are not a concern.

The G2 target site was greatly mismatched to the primer. Three sites conserved in the Canadian targets did not complement the primer. In addition, a fourth common site in the primer target varied. This primer was developed to detect the DS-1 rotavirus strain which was in danger of failure due to primer mismatch with Canadian strains. Future surveillance studies should consider changing the primer by adding degenerate sites or by evaluating which conserved target sequences should be used. DS-1 should not be used as the basis for new primers, since it is no longer representative of the circulating G2 population of rotaviruses.

Canadian G3 rotavirus strains were detected with a completely conserved target site; however, the primer contains one mismatched base. Fortunately, the mismatched base occurs near the middle of the primer and is unlikely to affect the detection of G3 strains. False-positive G3 detection is unlikely since G3 target sequences contain 3 exclusively genotype-specific bases, as well as several genotype-specific bases that are seldom shared by other genotypes.

The G4 target site is highly conserved among the three Canadian strains sequenced. There are three mismatches with the primer sequence at the 5' end. These

mismatches are a concern for effective G4 genotyping. This primer should be re-evaluated for redevelopment before use in another surveillance study.

The G9 specific primer is new compared to the primers for other genotype targets. Therefore, the primer matches the target site quite well. The target site contains two bases that vary among Canadian G9 strains, but these variations are rare. When the variable sites are excluded, the G9 primer suits the target site perfectly. The target site is also one of high variability amongst other genotypes, so false-positive G9 detection is unlikely.

The P[8] primer was found in a highly variable part of the target sequence. Despite degeneracy in the primer, the number of commonly variable sites makes this target a poor choice for continued surveillance. More specifically, strains with four conserved sites that differ from the majority of Canadian P[8] strains, are found in sub-genotypic lineage 2 in the P[8] phylogenetic tree. These strains are more closely related to the Wa-type strain the primer is based on. Three of the four sites that differ are those that correspond to degenerate bases on the primer. Currently, the degeneracy allows a suitable number of variations in nucleotide sequence to be matched, so that the primer works effectively. However, as these lineages diverge, it may be more suitable to have a genotype-specific primer somewhere less genetically variable. This is an especially sound strategy considering the implications of adding more degenerate sites as mutations accumulate. More degeneracy will result in lack of specificity, more dilution of specific primer components and the possibility of false-P[8] detection.

The 2T-1 primer for detection of P[4] is well suited for Canadian P[4] strains. The target site for the 2T-1 primer is completely conserved in Canadian strains at the target

site. In addition, there is only one base that mismatches the primer. False-P[4] detection is unlikely, as there are seven genotype-specific sites in the P[4] primer site.

#### 4.3 Real Time PCR Assay:

Real-time PCR utilized a 184 base pair target sequence (size of the segment of the Wa VP2 segment may vary if there are insertions or deletions in the future) near the 3' end of the rotavirus VP2 genome segment. Of the 84 samples previously genotyped by full ORF sequencing analysis, 76 were C1 and 8 were genotype C2. VP2 genotype strains matched non-VP7 segment genotypes 100% of the time, and fell into either Wa-like (genogroup II) or DS-1-like (genogroup I). Therefore, this assay was developed to detect all rotaviruses, while being able to discriminate between C1 and C2 rotaviruses. This VP2-based real-time detection assay should prove highly useful in future studies.

The real-time PCR assay was successfully validated and is a sensitive and specific test capable of discerning between the two most commonly occurring VP2 genotypes circulating in Canada today. The panel of samples tested included 136 C1, 20 C2, 37 negative and 2 multiple genotype samples. Three samples that had multiple G or P types were detected by the C1 probe only. Samples tested came from a variety of years and were of a variety of G genotypes to ensure the test was robust enough to capture all rotavirus samples. Of the 195 samples, 73 were from 2007, 61 were from 2008, 38 were from 2009 and 23 were from 2010. Previous genotyping and confirmation showed that 101 samples were G1P[8], 17 were G2P[4], 24 were G3P[8], 3 were G4P[8], 4 were G9P[8], 2 were G9P[4], 6 samples contained multiple genotypes, and one was previously unconfirmed by sequencing but genotyped G2P[4]. Several attempts to confirm the

genotype of sample RT029-07 were made but sequencing was inconclusive. In this instance, the number of copies detected in RT029-07 by the assay was about 31725, which is on the low end of the standard curve. Detection of the unconfirmed strain shows that this real-time assay is more sensitive and can confirm a lower copy number of genomic material than can traditional reverse-transcriptase PCR and sequencing. RNA degradation due to multiple freeze thaw cycles and time spent in -20° C may account for the real-time assay failing to detect both C1 and C2 genotypes in multiple-genotype samples, previously found to have G1, G2, P[4] and P[8] genotypes. RNA degradation may have also played a role in the inability to confirm the RT029-07 genotype in earlier studies.

The real-time assay is likely to prove most valuable when used in conjunction with VP4 and VP7 heminested PCR genotyping. The real-time assay is fast and is useful to confirm results from the heminested assays. Drawbacks of heminested multiplex PCR include non-specific amplification causing false positives, misreading of genotypes due to band smearing, or lack of separation and the possibility of misgenotyping or genotyping failure due to mutations in genotype-specific primer binding sites. Previously, genotyping results were confirmed using partial genome sequencing of the VP7 and VP4 genome segments. Sequencing, while highly confirmatory and helpful, is only as strong as the parent primer set used in the heminested multiplex PCR assay. Primer mismatch or low template concentration in sequencing could lead to a false negative test result. In addition, sequencing lacks the ability to detect multiple genotypes in a single sample, without further steps to culture, clone and isolate multiple strains from a specimen. The real-time PCR assay is both highly sensitive and robust, in that it will detect strains of

rotavirus within a sample to a lower limit of around  $5 \times 10^3$  copies. As rotavirus related illness is unlikely to occur when such a low copy number is shed in the stool, this sensitivity is more than sufficient to detect rotaviruses causing disease. The real-time PCR assay also utilizes a different genome segment and primer set. Thus, it requires at least one of two probes to bind to the amplifying DNA, therefore it is less susceptible to contamination than standard PCR. Since this real-time assay has been developed, and demonstrated to discriminate between C1 and C2 rotavirus strains, it will be useful as an alternative assay to detect multiple genotype samples.

In comparison to sequencing, the use of the real-time assay as a confirmatory assay is more time efficient, uses fewer materials and reagents and can detect more than one genogroup of virus within samples. The real-time PCR assay also has few shortcomings compared to traditional sequencing. The biggest shortcoming is the inability to discern between G genotypes. If a genotyping band is interpreted in the wrong position due to gel inconsistencies, or smearing, the real time assay can not detect the error. In addition, with conventional heminested multiplex PCR, any non-specific amplification that could be misconstrued as a positive result for a rotavirus genotype can be confirmed by sequencing the non-specific amplicon. Searching for the sequence using internet based genetic tools such as BLAST can identify the false target amplicon causing the amplification. During the course of this study three adenovirus samples were identified by sequencing non-specific amplicons. Finally, in some cases, sequences can also be aligned to allow phylogenetic analysis, without further sequencing.



#### 4.4 Future Directions:

Additional genotype surveillance will be necessary as vaccine uptake increases. Regular and broad surveillance standards should be set to enable adequate surveillance of evolving and emerging rotavirus strains in Canada. This is especially important due to the possible effects of selective pressure on viral evolution caused by increased use of vaccines.

Motif analysis from this study can be used to follow up on any rare amino acid variations found in Canadian rotavirus strains. Protein and cell culture studies can answer questions about protein function, protein interactions, viral fitness, antigenicity and the roles of particular variable motifs during infection. Comparison of individual protein motifs implicated in pathogenesis may answer the question of why G1P[8] strains have been the most prevalent strains in humans over time.

Rotavirus genotyping assays require regular updating in order to ensure they are effective in the genotyping of particular rotavirus genotypes. Monitoring of these assays should continue and primers should be updated and tested accordingly.

## 5. Conclusion

Rotavirus remains a substantial cause of morbidity in Canada today. This study has shown that circulating rotaviruses in Vancouver, Edmonton, Saskatoon, Winnipeg, Toronto, Ottawa, Quebec City and Halifax, are predominantly G1P[8], G2P[4], G3P[8], G4P[8] or G9P[8] rotaviruses. Other viruses make up less than 1% of strains detected. Efficacy studies have shown that RotaTeq protects against severe rotavirus gastroenteritis for all of the prevailing strains found circulating in Canada. Phylogenetic trees and motif analysis showed that Canadian rotaviruses differ at a sub-genotypic level, but are not abnormal rotavirus strains. Rotavirus vaccines have only recently been introduced as part of the Canadian vaccination schedule in 4 of 13 provinces and territories. Further roll-out of vaccination programs are expected to reduce the burden of rotavirus on the health of infants and reduce the costs to the healthcare system and economy of AGE.

This study analysed standard rotavirus protein motifs. In some cases, there were changes in motif location, which demonstrates the genetic flexibility of rotavirus proteins. While motifs were highly conserved in some functional regions, they showed variability in other regions, which may not be as strongly constrained by functional factors. Variability most often occurred between the two Canadian genogroups, the Wa-like genogroup and the DS-1-like genogroup. Differences in important motifs, especially those for viral protein-protein interactions, may explain the extreme bias these proteins have when assorting with one another, which discourages reassortment events between genogroups.

Currently used rotavirus genotyping assays should be regularly validated to ensure they have adequate performance. Several aspects of the currently used heminested

multiplex PCR assay were evaluated and while current performance is good, there could be the possibility of primer failure due to genetic drift in specific strains. Close monitoring of the 1T-1DCDN primer for P[8] typing, in particular, the aCT2 primer for G2 typing, and the aDT4 primer for G4 typing, should be done to avoid the possibility of genotyping failures.

A real-time assay for detection of rotavirus was developed as an alternative to sequencing for confirmation of rotavirus presence and genotype. However, confirmation of genotype with this assay is limited to genotyping between genogroup II (highly associated with G1, G3, G4, G9 and P[8] genotypes) and genogroup I (highly associated with G2 and P[4] genotypes). The real-time assay also has the advantages of high sensitivity and being quicker and easier to carry out.

Surveillance of rotavirus on an annual basis should be continued to monitor changes in rotavirus prevalence, as well as to detect the possible emergence of new rotavirus genotypes. Strain surveillance should become routine as vaccinations are likely to change the rotavirus landscape in Canada and the rest of the world

## 6. References

- Abdel-Haq N, Amjad M, McGrath E, Chearskul P, Amer A, Salimnia H, Asmar BI. 2011. Emergence of Human Rotavirus Genotype G9 in Metropolitan Detroit between 2007 and 2009. *J Med Microbiol* .60:671-677.
- Affranchino JL, Gonzalez SA. 1997. Deletion mapping of functional domains in the rotavirus capsid protein VP6. *J Gen Virol* 78 ( Pt 8):1949-1955.
- Aladin F, Nawaz S, Iturriza-Gomara M, Gray J. 2010. Identification of G8 rotavirus strains determined as G12 by rotavirus genotyping PCR: updating the current genotyping methods. *J Clin Virol* 47:340-344.
- Arias CF, Romero P, Alvarez V, Lopez S. 1996. Trypsin activation pathway of rotavirus infectivity. *J Virol* 70:5832-5839.
- Arnoldi F, Burrone OR. 2009. Role of viral nonstructural proteins in rotavirus replication. *Future Virol* 4:185-196.
- Baker M, Prasad BV. 2010. Rotavirus cell entry. *Curr Top Microbiol Immunol* 343:121-148.
- Ball JM, Mitchell DM, Gibbons TF, Parr RD. 2005. Rotavirus NSP4: a multifunctional viral enterotoxin. *Viral Immunol* 18:27-40.
- Banerjee I, Ramani S, Primrose B, Iturriza-Gomara M, Gray JJ, Brown DW, Kang G. 2007. Modification of rotavirus multiplex RT-PCR for the detection of G12 strains based

on characterization of emerging G12 rotavirus strains from South India. *J Med Virol* 79:1413-1421.

Boom JA, Tate JE, Sahni LC, Rench MA, Hull JJ, Gentsch JR, Patel MM, Baker CJ, Parashar UD. 2010. Effectiveness of pentavalent rotavirus vaccine in a large urban population in the United States. *Pediatrics* 125:e199-207.

Both GW, Siegman LJ, Bellamy AR, Ikegami N, Shatkin AJ, Furuichi Y. 1984. Comparative sequence analysis of rotavirus genomic segment 6--the gene specifying viral subgroups 1 and 2. *J Virol* 51:97-101.

Braeckman T, Van Herck K, Raes M, Vergison A, Sabbe M, Van Damme P. 2011. Rotavirus vaccines in Belgium: policy and impact. *Pediatr Infect Dis J* 30:S21-4.

Breiman RF, Zaman K, Armah G, Sow SO, Anh DD, Victor JC, Hille D, Ciarlet M, Neuzil KM. 2012. Analyses of health outcomes from the 5 sites participating in the Africa and Asia clinical efficacy trials of the oral pentavalent rotavirus vaccine. *Vaccine* 30 Suppl 1:A24-9.

Carreno-Torres JJ, Gutierrez M, Arias CF, Lopez S, Isa P. 2010. Characterization of viroplasm formation during the early stages of rotavirus infection. *Virol J* 7:350.

Centers for Disease Control and Prevention (CDC). 2009. Reduction in rotavirus after vaccine introduction--United States, 2000-2009. *MMWR Morb Mortal Wkly Rep* 58:1146-1149.

Chandrasekaran C, Grimwood K, Redshaw N, Rich FJ, Wood C, Stanley J, Kirman JR, New Zealand Rotavirus Study Group. 2010. Geographical differences in the proportion of human group A rotavirus strains within New Zealand during one epidemic season. *J Med Virol* 82:897-902.

Chen Y, Wen Y, Liu X, Xiong X, Cao Z, Zhao Q, Yu Y, Yin X, Li C, Fan Y. 2008. Full genomic analysis of human rotavirus strain TB-Chen isolated in China. *Virology* 375:361-373.

Chen Y, Zhu W, Sui S, Yin Y, Hu S, Zhang X. 2009. Whole genome sequencing of lamb rotavirus and comparative analysis with other mammalian rotaviruses. *Virus Genes* 38:302-310.

Cilla G, Montes M, Gomariz M, Pineiro L, Perez-Trallero E. 2010. Rotavirus genotypes in children in the Basque Country (northern Spain) over a 13-year period (July 1996-June 2009). *Eur J Clin Microbiol Infect Dis* 29:955-960.

Cook JP, McCrae MA. 2004. Sequence analysis of the guanylyltransferase (VP3) of group A rotaviruses. *J Gen Virol* 85:929-932.

Cortes JE, Curns AT, Tate JE, Parashar UD. 2009. Trends in healthcare utilization for diarrhea and rotavirus disease in privately insured US children <5 years of age, 2001-2006. *Pediatr Infect Dis J* 28:874-878.

Coulson BS, Londrigan SL, Lee DJ. 1997. Rotavirus contains integrin ligand sequences and a disintegrin-like domain that are implicated in virus entry into cells. *Proc Natl Acad Sci U S A* 94:5389-5394.

Cunliffe NA, Booth JA, Elliot C, Lowe SJ, Sopwith W, Kitchin N, Nakagomi O, Nakagomi T, Hart CA, Regan M. 2010. Healthcare-associated viral gastroenteritis among children in a large pediatric hospital, United Kingdom. *Emerg Infect Dis* 16:55-62.

Cunliffe NA, Das BK, Ramachandran M, Bhan MK, Glass RI, Gentsch JR. 1997. Sequence analysis demonstrates that VP6, NSP1 and NSP4 genes of Indian neonatal rotavirus strain 116E are of human origin. *Virus Genes* 15:39-44.

Danchin MH, Bines JE. 2009. Defeating rotavirus? The global recommendation for rotavirus vaccination. *N Engl J Med* 361:1919-1921.

Das BK, Gentsch JR, Cicirello HG, Woods PA, Gupta A, Ramachandran M, Kumar R, Bhan MK, Glass RI. 1994. Characterization of rotavirus strains from newborns in New Delhi, India. *J Clin Microbiol* 32:1820-1822.

Das S, Varghese V, Chaudhuri S, Barman P, Kojima K, Dutta P, Bhattacharya SK, Krishnan T, Kobayashi N, Naik TN. 2004. Genetic variability of human rotavirus strains isolated from Eastern and Northern India. *J Med Virol* 72:156-161.

De Grazia S, Giammanco GM, Potgieter CA, Matthijnsens J, Banyai K, Platia MA, Colomba C, Martella V. 2010. Unusual assortment of segments in 2 rare human rotavirus genomes. *Emerg Infect Dis* 16:859-862.

de Rougemont A, Kaplon J, Pillet S, Mory O, Gagneur A, Minoui-Tran A, Meritet JF, Mollat C, Lorrot M, Foulongne V, Gillet Y, Nguyen-Bourgain C, Alain S, Agius G, Lazrek M, Colimon R, Fontana C, Gendrel D, Pothier P, the French Rotavirus Network. 2010. Molecular and Clinical Characterization of Rotavirus From Diarrheal Infants Admitted to Pediatric Emergency Units in France. *Pediatr Infect Dis J* .

Dennehy PH. 2008. Rotavirus vaccines: an overview. *Clin Microbiol Rev* 21:198-208.

Desai SN, Esposito DB, Shapiro ED, Dennehy PH, Vazquez M. 2010. Effectiveness of rotavirus vaccine in preventing hospitalization due to rotavirus gastroenteritis in young children in Connecticut, USA. *Vaccine* .

Dormitzer PR, Nason EB, Prasad BV, Harrison SC. 2004. Structural rearrangements in the membrane penetration protein of a non-enveloped virus. *Nature* 430:1053-1058.

Dormitzer PR, Sun ZY, Wagner G, Harrison SC. 2002. The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. *EMBO J* 21:885-897.

Dube E, Bradet R, Gilca V, Sauvageau C, Boulianne N, Boucher FD, Bettinger JA, McNeil S, Gemmill I, Lavoie F. 2010. Rotavirus vaccination: Results of a survey on Canadian physicians' knowledge, attitudes and practices. *Can J Infect Dis Med Microbiol* 21:173-173-234.

Ernst H, Duhl JA. 1989. Nucleotide sequence of genomic segment 2 of the human rotavirus Wa. *Nucleic Acids Res* 17:4382.



Esona MD, Steele D, Kerin T, Armah G, Peenze I, Geyer A, Page N, Nyangao J, Agbaya VA, Trabelsi A, Tsion B, Aminu M, Sebuya T, Dewar J, Glass R, Gentsch J. 2010.

Determination of the G and P types of previously nontypeable rotavirus strains from the African Rotavirus Network, 1996-2004: Identification of unusual G types. *J Infect Dis* 202 Suppl:S49-54.

Esteban LE, Rota RP, Gentsch JR, Jiang B, Esona M, Glass RI, Glikmann G, Castello AA. 2010. Molecular epidemiology of group A rotavirus in Buenos Aires, Argentina 2004-2007: reemergence of G2P[4] and emergence of G9P[8] strains. *J Med Virol* 82:1083-1093.

Estes MK, Kapikian AZ. 2007. Rotaviruses. In: Knipe DM, Howley PM, Griffin DE, editors. *Field's Virology*, 5th ed. Lippincott Williams & Wilkins: Philadelphia, PA. p 1917-1917-1974.

Feeney SA, Armstrong VJ, Mitchell SJ, Crawford L, McCaughey C, Coyle PV. 2011. Development and clinical validation of multiplex TaqMan(R) assays for rapid diagnosis of viral gastroenteritis. *J Med Virol* 83:1650-1656.

Fischer TK, Gentsch JR. 2004. Rotavirus typing methods and algorithms. *Rev Med Virol* 14:71-82.

Gentsch JR, Das BK, Jiang B, Bhan MK, Glass RI. 1993. Similarity of the VP4 protein of human rotavirus strain 116E to that of the bovine B223 strain. *Virology* 194:424-430.

Gentsch JR, Glass RI, Woods P, Gouvea V, Gorziglia M, Flores J, Das BK, Bhan MK. 1992. Identification of group A rotavirus gene 4 types by polymerase chain reaction. *J Clin Microbiol* 30:1365-1373.

Ghosh S, Kobayashi N. 2011. Whole-genomic analysis of rotavirus strains: current status and future prospects. *Future Microbiol* 6:1049-1065.

Gilbert JM, Greenberg HB. 1998. Cleavage of rhesus rotavirus VP4 after arginine 247 is essential for rotavirus-like particle-induced fusion from without. *J Virol* 72:5323-5327.

Gomara MI, Cubitt D, Desselberger U, Gray J. 2001. Amino acid substitution within the VP7 protein of G2 rotavirus strains associated with failure to serotype. *J Clin Microbiol* 39:3796-3798.

Gorziglia M, Green K, Nishikawa K, Taniguchi K, Jones R, Kapikian AZ, Chanock RM. 1988. Sequence of the fourth gene of human rotaviruses recovered from asymptomatic or symptomatic infections. *J Virol* 62:2978-2984.

Gouvea V, Glass RI, Woods P, Taniguchi K, Clark HF, Forrester B, Fang ZY. 1990. Polymerase chain reaction amplification and typing of rotavirus nucleic acid from stool specimens. *J Clin Microbiol* 28:276-282.

Graff JW, Mitzel DN, Weisend CM, Flenniken ML, Hardy ME. 2002. Interferon regulatory factor 3 is a cellular partner of rotavirus NSP1. *J Virol* 76:9545-9550.

Greenberg HB, Estes MK. 2009. Rotaviruses: from pathogenesis to vaccination. *Gastroenterology* 136:1939-1951.

Guglielmi KM, McDonald SM, Patton JT. 2010. Mechanism of intraparticle synthesis of the rotavirus double-stranded RNA genome. *J Biol Chem* 285:18123-18128.

Gurgel RQ, Cuevas LE, Vieira SC, Barros VC, Fontes PB, Salustino EF, Nakagomi O, Nakagomi T, Dove W, Cunliffe N, Hart CA. 2007. Predominance of rotavirus P[4]G2 in a vaccinated population, Brazil. *Emerg Infect Dis* 13:1571-1573.

Halaihel N, Lievin V, Ball JM, Estes MK, Alvarado F, Vasseur M. 2000. Direct inhibitory effect of rotavirus NSP4(114-135) peptide on the Na(+)-D-glucose symporter of rabbit intestinal brush border membrane. *J Virol* 74:9464-9470.

Hall BG. 2007. *Phylogenetic Trees Made Easy Third Edition*, 3rd ed. Sinauer Associates, Inc: Sunderland, Massachusetts. 233 p.

Hall BG, Barlow M. 2006. Phylogenetic analysis as a tool in molecular epidemiology of infectious diseases. *Ann Epidemiol* 16:157-169.

Han TH, Kim CH, Chung JY, Park SH, Hwang ES. 2010. Genetic characterization of rotavirus in children in South Korea from 2007 to 2009. *Arch Virol* .

Heiman EM, McDonald SM, Barro M, Taraporewala ZF, Bar-Magen T, Patton JT. 2008. Group A human rotavirus genomics: evidence that gene constellations are influenced by viral protein interactions. *J Virol* 82:11106-11116.

Horie Y, Masamune O, Nakagomi O. 1997. Three major alleles of rotavirus NSP4 proteins identified by sequence analysis. *J Gen Virol* 78 ( Pt 9):2341-2346.

Hua J, Mansell EA, Patton JT. 1993. Comparative analysis of the rotavirus NS53 gene: conservation of basic and cysteine-rich regions in the protein and possible stem-loop structures in the RNA. *Virology* 196:372-378.

Isa P, Gutiérrez M, Arias CF, López S. 2008. Rotavirus cell entry. *Future Virology* 3:135-146.

Isegawa Y, Nakagomi O, Nakagomi T, Ueda S. 1992. A VP4 sequence highly conserved in human rotavirus strain AU-1 and feline rotavirus strain FRV-1. *J Gen Virol* 73 ( Pt 8):1939-1946.

Ito H, Minamoto N, Sasaki I, Goto H, Sugiyama M, Kinjo T, Sugita S. 1995. Sequence analysis of cDNA for the VP6 protein of group A avian rotavirus: a comparison with group A mammalian rotaviruses. *Arch Virol* 140:605-612.

Ito H, Sugiyama M, Masubuchi K, Mori Y, Minamoto N. 2001. Complete nucleotide sequence of a group A avian rotavirus genome and a comparison with its counterparts of mammalian rotaviruses. *Virus Res* 75:123-138.

Iturriza-Gomara M, Anderton E, Kang G, Gallimore C, Phillips W, Desselberger U, Gray J. 2003. Evidence for genetic linkage between the gene segments encoding NSP4 and VP6 proteins in common and reassortant human rotavirus strains. *J Clin Microbiol* 41:3566-3573.

Iturriza-Gomara M, Green J, Brown DW, Desselberger U, Gray JJ. 2000. Diversity within the VP4 gene of rotavirus P[8] strains: implications for reverse transcription-PCR genotyping. *J Clin Microbiol* 38:898-901.

Iturriza-Gomara M, Kang G, Gray J. 2004. Rotavirus genotyping: keeping up with an evolving population of human rotaviruses. *J Clin Virol* 31:259-265.

Jayaram H, Estes MK, Prasad BV. 2004. Emerging themes in rotavirus cell entry, genome organization, transcription and replication. *Virus Res* 101:67-81.

Jayaram H, Taraporewala Z, Patton JT, Prasad BV. 2002. Rotavirus protein involved in genome replication and packaging exhibits a HIT-like fold. *Nature* 417:311-315.

Jothikumar N, Kang G, Hill VR. 2009. Broadly reactive TaqMan assay for real-time RT-PCR detection of rotavirus in clinical and environmental samples. [JIN2@cdc.gov](mailto:JIN2@cdc.gov). *J Virol Methods* 155:126-131.

Kirkwood C, Bogdanovic-Sakran N, Palombo E, Masendycz P, Bugg H, Barnes G, Bishop R. 2003. Genetic and antigenic characterization of rotavirus serotype G9 strains isolated in Australia between 1997 and 2001. *J Clin Microbiol* 41:3649-3654.

Kirkwood C, Masendycz PJ, Coulson BS. 1993. Characteristics and location of cross-reactive and serotype-specific neutralization sites on VP7 of human G type 9 rotaviruses. *Virology* 196:79-88.

Kirkwood CD, Bogdanovic-Sakran N, Cannan D, Bishop RF, Barnes GL. 2006a. National Rotavirus Surveillance Program annual report, 2004-05. *Commun Dis Intell* 30:133-136.

Kirkwood CD, Boniface K, Bishop RF, Barnes GL. 2010. Australian Rotavirus Surveillance Program: annual report, 2009/2010. *Commun Dis Intell* 34:427-434.

Kirkwood CD, Boniface K, Bishop RF, Barnes GL, Australian Rotavirus Surveillance Group. 2009. Australian Rotavirus Surveillance Program annual report, 2008/2009. *Commun Dis Intell* 33:382-388.

Kirkwood CD, Cannan D, Bogdanovic-Sakran N, Bishop RF, Barnes GL, National Rotavirus Surveillance Group. 2007. Australian Rotavirus Surveillance Program: annual report, 2006-07. *Commun Dis Intell* 31:375-379.

Kirkwood CD, Cannan D, Bogdanovic-Sakran N, Bishop RF, Barnes GL, National Rotavirus Surveillance Group. 2006b. National Rotavirus Surveillance Program annual report, 2005-06. *Commun Dis Intell* 30:434-438.

Kirkwood CD, Cannan D, Boniface K, Bishop RF, Barnes GL, Australian Rotavirus Surveillance Group. 2008. Australian Rotavirus Surveillance Program annual report, 2007/08. *Commun Dis Intell* 32:425-429.

Kirkwood CD, Palombo EA. 1997. Genetic characterization of the rotavirus nonstructural protein, NSP4. *Virology* 236:258-265.

Kojima K, Taniguchi K, Kobayashi N. 1996. Species-specific and interspecies relatedness of NSP1 sequences in human, porcine, bovine, feline, and equine rotavirus strains. *Arch Virol* 141:1-12.

Kostouros E, Siu K, Ford-Jones EL, Petric M, Tellier R. 2003. Molecular characterization of rotavirus strains from children in Toronto, Canada. *J Clin Virol* 28:77-84.

Kudo S, Zhou Y, Cao XR, Yamanishi S, Nakata S, Ushijima H. 2001. Molecular characterization in the VP7, VP4 and NSP4 genes of human rotavirus serotype 4 (G4) isolated in Japan and Kenya. *Microbiol Immunol* 45:167-171.

Kumar M, Jayaram H, Vasquez-Del Carpio R, Jiang X, Taraporewala ZF, Jacobson RH, Patton JT, Prasad BV. 2007. Crystallographic and biochemical analysis of rotavirus NSP2 with nucleotides reveals a nucleoside diphosphate kinase-like activity. *J Virol* 81:12272-12284.

Laird AR, Gentsch JR, Nakagomi T, Nakagomi O, Glass RI. 2003. Characterization of serotype G9 rotavirus strains isolated in the United States and India from 1993 to 2001. *J Clin Microbiol* 41:3100-3111.

Le Saux N, Bettinger J, Dery P, Embree J, Vaudry W, Halperin SA, McDermid A, Booth TF, Coyle D. 2011. The Hidden Costs and Characteristics of Childhood Rotavirus Emergency Visits in Canada. *Pediatr Infect Dis J*. 31: 159-163.

Lennon G, Reidy N, Cryan B, Fanning S, O'Shea H. 2008. Changing profile of rotavirus in Ireland: predominance of P[8] and emergence of P[6] and P[9] in mixed infections. *J Med Virol* 80:524-530.

Lopez S, Arias CF. 2006. Early steps in rotavirus cell entry. *Curr Top Microbiol Immunol* 309:39-66.

Madhi SA, Cunliffe NA, Steele D, Witte D, Kirsten M, Louw C, Ngwira B, Victor JC, Gillard PH, Cheuvart BB, Han HH, Neuzil KM. 2010. Effect of human rotavirus vaccine on severe diarrhea in African infants. *N Engl J Med* 362:289-298.

Martella V, Ciarlet M, Banyai K, Lorusso E, Cavalli A, Corrente M, Elia G, Arista S, Camero M, Desario C, Decaro N, Lavazza A, Buonavoglia C. 2006. Identification of a novel VP4 genotype carried by a serotype G5 porcine rotavirus strain. *Virology* 346:301-311.

Mason BB, Dheer SK, Hsiao CL, Zandle G, Kostek B, Rosanoff EI, Hung PP, Davis AR. 1985. Sequence of the serotype-specific glycoprotein of the human rotavirus Wa strain and comparison with other human rotavirus serotypes. *Virus Res* 2:291-299.

Mast TC, Walter EB, Bulotsky M, Khawaja SS, DiStefano DJ, Sandquist MK, Straus WL, Staat MA. 2010. Burden of childhood rotavirus disease on health systems in the United States. *Pediatr Infect Dis J* 29:e19-25.



Matthijnssens J, Bilcke J, Ciarlet M, Martella V, Banyai K, Rahman M, Zeller M, Beutels P, Van Damme P, Van Ranst M. 2009. Rotavirus disease and vaccination: impact on genotype diversity. *Future Microbiol* 4:1303-1316.

Matthijnssens J, Ciarlet M, Heiman E, Arijs I, Delbeke T, McDonald SM, Palombo EA, Iturriza-Gomara M, Maes P, Patton JT, Rahman M, Van Ranst M. 2008b. Full genome-based classification of rotaviruses reveals a common origin between human Wa-Like and porcine rotavirus strains and human DS-1-like and bovine rotavirus strains. *J Virol* 82:3204-3219.

Matthijnssens J, Ciarlet M, McDonald SM, Attoui H, Banyai K, Brister JR, Buesa J, Esona MD, Estes MK, Gentsch JR, Iturriza-Gomara M, Johne R, Kirkwood CD, Martella V, Mertens PP, Nakagomi O, Parreno V, Rahman M, Ruggeri FM, Saif LJ, Santos N, Steyer A, Taniguchi K, Patton JT, Desselberger U, Van Ranst M. 2011. Uniformity of rotavirus strain nomenclature proposed by the Rotavirus Classification Working Group (RCWG). *Arch Virol* 156:1397-1413.

Matthijnssens J, Ciarlet M, Rahman M, Attoui H, Banyai K, Estes MK, Gentsch JR, Iturriza-Gomara M, Kirkwood CD, Martella V, Mertens PP, Nakagomi O, Patton JT, Ruggeri FM, Saif LJ, Santos N, Steyer A, Taniguchi K, Desselberger U, Van Ranst M. 2008a. Recommendations for the classification of group A rotaviruses using all 11 genomic RNA segments. *Arch Virol* 153:1621-1629.

- Matthijnssens J, Heylen E, Zeller M, Rahman M, Lemey P, Van Ranst M. 2010a. Phylogenetic analyses of rotavirus genotypes G9 and G12 underscore their potential for swift global spread. *Mol Biol Evol* 27:2431-2436.
- Matthijnssens J, Joelsson DB, Warakomski DJ, Zhou T, Mathis PK, van Maanen MH, Ranheim TS, Ciarlet M. 2010b. Molecular and biological characterization of the 5 human-bovine rotavirus (WC3)-based reassortant strains of the pentavalent rotavirus vaccine, RotaTeq. *Virology* 403:111-127.
- Matthijnssens J, Rahman M, Van Ranst M. 2008. Two out of the 11 genes of an unusual human G6P[6] rotavirus isolate are of bovine origin. *J Gen Virol* 89:2630-2635.
- McDonald SM, Matthijnssens J, McAllen JK, Hine E, Overton L, Wang S, Lemey P, Zeller M, Van Ranst M, Spiro DJ, Patton JT. 2009. Evolutionary dynamics of human rotaviruses: balancing reassortment with preferred genome constellations. *PLoS Pathog* 5:e1000634.
- Mohan KV, Atreya CD. 2001. Nucleotide sequence analysis of rotavirus gene 11 from two tissue culture-adapted ATCC strains, RRV and Wa. *Virus Genes* 23:321-329.
- Mulherin E, Bryan J, Beltman M, O'Grady L, Pidgeon E, Garon L, Lloyd A, Bainbridge J, O'Shea H, Whyte P, Fanning S. 2008. Molecular characterisation of a bovine-like rotavirus detected from a giraffe. *BMC Vet Res* 4:46.
- Munos MK, Walker CL, Black RE. 2010. The effect of rotavirus vaccine on diarrhoea mortality. *Int J Epidemiol* 39 Suppl 1.

Nakagomi O, Kaga E. 1995. Distinctness of NSP1 gene of human rotavirus AU-1 from NSP1 gene of other human genogroups. *Res Virol* 146:423-428.

National Advisory Committee on Immunization (NACI). 2008a. Statement on the recommended use of pentavalent human-bovine reassortant rotavirus vaccine. An Advisory Committee Statement (ACS). *Can Commun Dis Rep* 34:1-33.

National Advisory Committee on Immunization (NACI). 2008b. Statement on the recommended use of pentavalent human-bovine reassortant rotavirus vaccine. An Advisory Committee Statement (ACS). *Can Commun Dis Rep* 34:1-33.

Newall AT, Beutels P, Macartney K, Wood J, MacIntyre CR. 2007. The cost-effectiveness of rotavirus vaccination in Australia. *Vaccine* 25:8851-8860.

Nordgren J, Bucardo F, Svensson L, Lindgren PE. 2010. Novel light-upon-extension real-time PCR assay for simultaneous detection, quantification, and genogrouping of group A rotavirus. *J Clin Microbiol* 48:1859-1865.

O'Brien J. 2008. 11th Annual Conference on Vaccine Research. *Expert Rev Vaccines* 7:721-723.

Okada Y, Richardson MA, Ikegami N, Nomoto A, Furuichi Y. 1984. Nucleotide sequence of human rotavirus genome segment 10, an RNA encoding a glycosylated virus protein. *J Virol* 51:856-859.

Padilla-Noriega L, Dunn SJ, Lopez S, Greenberg HB, Arias CF. 1995. Identification of two independent neutralization domains on the VP4 trypsin cleavage products VP5\* and VP8\* of human rotavirus ST3. *Virology* 206:148-154.

Pang XL, Lee B, Boroumand N, Leblanc B, Preiksaitis JK, Yu Ip CC. 2004. Increased detection of rotavirus using a real time reverse transcription-polymerase chain reaction (RT-PCR) assay in stool specimens from children with diarrhea. *J Med Virol* 72:496-501.

Patel MM, de Oliveira LH, Bispo AM, Gentsch J, Parashar UD. 2008. Rotavirus P[4]G2 in a vaccinated population, Brazil. *Emerg Infect Dis* 14:863-865.

Patel MM, Steele D, Gentsch JR, Wecker J, Glass RI, Parashar UD. 2011. Real-world impact of rotavirus vaccination. *Pediatr Infect Dis J* 30:S1-5.

Patel NC, Hertel PM, Estes MK, de la Morena M, Petru AM, Noroski LM, Revell PA, Hanson IC, Paul ME, Rosenblatt HM, Abramson SL. 2010. Vaccine-acquired rotavirus in infants with severe combined immunodeficiency. *N Engl J Med* 362:314-319.

Patton JT, Jones MT, Kalbach AN, He YW, Xiaobo J. 1997. Rotavirus RNA polymerase requires the core shell protein to synthesize the double-stranded RNA genome. *J Virol* 71:9618-9626.

Patton JT, Salter-Cid L, Kalbach A, Mansell EA, Kattoura M. 1993. Nucleotide and amino acid sequence analysis of the rotavirus nonstructural RNA-binding protein NS35. *Virology* 192:438-446.

Patton JT, Silvestri LS, Tortorici MA, Vasquez-Del Carpio R, Taraporewala ZF. 2006. Rotavirus genome replication and morphogenesis: role of the viroplasm. *Curr Top Microbiol Immunol* 309:169-187.

Paulke-Korinek M, Kundi M, Rendi-Wagner P, de Martin A, Eder G, Schmidle-Loss B, Vecsei A, Kollaritsch H. 2011. Herd immunity after two years of the universal mass vaccination program against rotavirus gastroenteritis in Austria. *Vaccine* 29:2791-2796.

Payne DC, Szilagyi PG, Staat MA, Edwards KM, Gentsch JR, Weinberg GA, Hall CB, Curns AT, Clayton H, Griffin MR, Fairbrother G, Parashar UD. 2009. Secular variation in United States rotavirus disease rates and serotypes: implications for assessing the rotavirus vaccination program. *Pediatr Infect Dis J* 28:948-953.

Pesavento JB, Crawford SE, Estes MK, Prasad BV. 2006. Rotavirus proteins: structure and assembly. *Curr Top Microbiol Immunol* 309:189-219.

PHAC and CNCI. 2011. **Publicly funded Immunization Programs in Canada - Routine Schedule for Infants and Children including special programs and catch-up programs (as of December 2011)**. 2012:1.

Pietsch C, Liebert UG. 2009. Human infection with G12 rotaviruses, Germany. *Emerg Infect Dis* 15:1512-1515.

Pongsuwanna Y, Guntapong R, Chiwakul M, Tacharoenmuang R, Onvimala N, Wakuda M, Kobayashi N, Taniguchi K. 2002. Detection of a human rotavirus with G12 and P[9] specificity in Thailand. *J Clin Microbiol* 40:1390-1394.

Rao CD, Das M, Ilango P, Lalwani R, Rao BS, Gowda K. 1995. Comparative nucleotide and amino acid sequence analysis of the sequence-specific RNA-binding rotavirus nonstructural protein NSP3. *Virology* 207:327-333.

Richardson V, Hernandez-Pichardo J, Quintanar-Solares M, Esparza-Aguilar M, Johnson B, Gomez-Altamirano CM, Parashar U, Patel M. 2010. Effect of rotavirus vaccination on death from childhood diarrhea in Mexico. *N Engl J Med* 362:299-305.

Riepenhoff-Talty M, Barrett HJ, Spada BA, Ogra PL. 1983. Negative staining and immune electron microscopy as techniques for rapid diagnosis of viral agents. *Ann N Y Acad Sci* 420:391-400.

Ruiz MC, Cohen J, Michelangeli F. 2000. Role of Ca<sup>2+</sup> in the replication and pathogenesis of rotavirus and other viral infections. *Cell Calcium* 28:137-149.

Salvadori M, Le Saux N. 2010. Recommendations for the use of rotavirus vaccines in infants. *Paediatr Child Health* 15:519-519-23.

Santos N, Hoshino Y. 2005. Global distribution of rotavirus serotypes/genotypes and its implication for the development and implementation of an effective rotavirus vaccine. *Rev Med Virol* 15:29-56.

Silvestri LS, Taraporewala ZF, Patton JT. 2004. Rotavirus replication: plus-sense templates for double-stranded RNA synthesis are made in viroplasm. *J Virol* 78:7763-7774.

Simmonds MK, Armah G, Asmah R, Banerjee I, Damanka S, Esona M, Gentsch JR, Gray JJ, Kirkwood C, Page N, Iturriza-Gomara M. 2008. New oligonucleotide primers for P-typing of rotavirus strains: Strategies for typing previously untypeable strains. *J Clin Virol* 42:368-373.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731-2739.

Tate JE, Panozzo CA, Payne DC, Patel MM, Cortese MM, Fowlkes AL, Parashar UD. 2009. Decline and change in seasonality of US rotavirus activity after the introduction of rotavirus vaccine. *Pediatrics* 124:465-471.

Tate JE, Patel MM, Steele AD, Gentsch JR, Payne DC, Cortese MM, Nakagomi O, Cunliffe NA, Jiang B, Neuzil KM, de Oliveira LH, Glass RI, Parashar UD. 2010. Global impact of rotavirus vaccines. *Expert Rev Vaccines* 9:395-407.

Than VT, Le VP, Lim I, Kim W. 2011. Complete genomic characterization of cell culture adapted human G12P[6] rotaviruses isolated from South Korea. *Virus Genes* 42:317-322.

Torres-Vega MA, Gonzalez RA, Duarte M, Poncet D, Lopez S, Arias CF. 2000. The C-terminal domain of rotavirus NSP5 is essential for its multimerization, hyperphosphorylation and interaction with NSP6. *J Gen Virol* 81:821-830.

Tsugawa T, Hoshino Y. 2008. Whole genome sequence and phylogenetic analyses reveal human rotavirus G3P[3] strains Ro1845 and HCR3A are examples of direct virion transmission of canine/feline rotaviruses to humans. *Virology* 380:344-353.

Vasquez-Del Carpio R, Gonzalez-Nilo FD, Riadi G, Taraporewala ZF, Patton JT. 2006. Histidine triad-like motif of the rotavirus NSP2 octamer mediates both RTPase and NTPase activities. *J Mol Biol* 362:539-554.

Vasquez-del Carpio R, Morales JL, Barro M, Ricardo A, Spencer E. 2006. Bioinformatic prediction of polymerase elements in the rotavirus VP1 protein. *Biol Res* 39:649-659.

Ward RL. 2008. Rotavirus vaccines: how they work or don't work. *Expert Rev Mol Med* 10:e5.

Widdowson MA, Meltzer MI, Zhang X, Bresee JS, Parashar UD, Glass RI. 2007. Cost-effectiveness and potential impact of rotavirus vaccination in the United States. *Pediatrics* 119:684-697.

Yee EL, Staat MA, Azimi P, Bernstein DI, Ward RL, Schubert C, Matson DO, Turcios-Ruiz RM, Parashar U, Widdowson MA, Glass RI. 2008. Burden of rotavirus disease among children visiting pediatric emergency departments in Cincinnati, Ohio, and Oakland, California, in 1999-2000. *Pediatrics* 122:971-977.

Yoshinaga M, Phan TG, Nguyen TA, Yan H, Yagyu F, Okitsu S, Muller WE, Ushijima H. 2006. Changing distribution of group A rotavirus G-types and genetic analysis of G9 circulating in Japan. *Arch Virol* 151:183-192.



Yuan L, Honma S, Kim I, Kapikian AZ, Hoshino Y. 2009. Resistance to rotavirus infection in adult volunteers challenged with a virulent G1P1A[8] virus correlated with serum immunoglobulin G antibodies to homotypic viral proteins 7 and 4. *J Infect Dis* 200:1443-1451.

Zeller M, Patton JT, Heylen E, De Coster S, Ciarlet M, Van Ranst M, Matthijnsens J. 2011. Genetic Analyses Reveal Differences in the VP7 and VP4 Antigenic Epitopes between Human Rotaviruses Circulating in Belgium and Rotaviruses in Rotarix<sup>®</sup> and RotaTeq<sup>®</sup>; *J Clin Microbiol* .

Zuridah H, Kirkwood CD, Bogdanovic-Sakran N, Bishop RF, Yap KL. 2010. Circulating human group A rotavirus genotypes in Malaysia. *J Med Virol* 82:707-711.

## 7. Appendix

Table A1: Sample collection sites broken down by site and year.

Rotavirus Season	Paediatric Hospital	Location	Number of Samples
2007	Centre hospitalier universitaire de Québec, Centre hospitalier universitaire de Laval (CHUQ, CHUL)	Quebec City, Quebec	27
2007	IWK Health Centre	Halifax, Nova Scotia	10
2007	Children's Hospital of Eastern Ontario (CHEO)	Ottawa, Ontario	30
2007	Stollery Children's Hospital	Edmonton, Alberta	24
2007	Winnipeg Children's Hospital	Winnipeg, Manitoba	4
2008	CHUQ, CHUL	Quebec City, Quebec	20
2008	CHEO	Ottawa, Ontario	51
2008	IWK Health Centre	Halifax, Nova Scotia	30
2009	CHUQ, CHUL	Quebec City, Quebec	16
2009	IWK Health Centre	Halifax, Nova Scotia	35
2009	CHEO	Ottawa, Ontario	40
2010	CHEO	Ottawa, Ontario	7
2010	IWK Health Centre	Halifax, Nova Scotia	11
2010	Royal University Hospital	Saskatoon, Saskatchewan	11
2010	Hospital for Sick Children	Toronto, Ontario	3
2010	Children's & Women's Health Centre of British Columbia	Vancouver, British Columbia	20
2010	Cadham Provincial Laboratory	Winnipeg, Manitoba	5

Table A2: Primers used for full genome sequencing.

Genome Segment	Primer Name	Primer Sense	Nucleotide Location	SEQUENCE (5'-3')	Mt (°C)	Ta (°C)	Amplicon Length
VP1	VP1-F-04 <sup>1</sup>	+	1-17	GGCTATTAAGCTGTAC	33.4	43	639
	VP1-F-04Rev	-	639-623	TGCCATGTTACTAAGT	36.7		
	VP1F-1AKM2	+	583-604	TATAGATATGAGTACGAGGTGA	41.2	44	727
	VP1R-1AKM2	-	1309-1287	CATTAGCCATATCATCCATTACA	49.0		
	VP1F-2AKM	+	929-952	CAAAAATGATACAAGACTGGTTAG	47.7	45	762
	VP1R-2ALT	-	1690-1666	GCGTCTGTATTACTTTAGCATCAT	49.6		
	VP1F-3AKM	+	1588-1606	TGGGATTCATCTCAACATA	42.6	46	951
	VP1R-3AKM	-	2538-2521	ATTCAATTCGCTTTTTC	41.9		
	VP1F-4AKM	+	2389-2408	TTATCAAGTCAGAAATCAGG	41.6	46	914
	VP1R-4AKM	-	3302-3283	GGTCACATCTAAGCACTCTA	41.3		
	VP14.5F*	+	2105-2127	TTAAATAATGAGAAGCGTGGACAA	54.8	45	593
VP14.5R*	-	2933-2915	GCATCAACTGGCGGAATC	52.2			
VP1-3'ENDF1	+	2751-2770	ACAATGTATAGGGTCAAGAA	41.0	47	535	
VP1-R-04 <sup>1</sup>	-	3302-3286	TCGCGAATCTACACT	47.0			
VP2	VP2F-1AKM	+	1-17	GGCTATTAAGGCTCAA	39.7	43	646
	VP2R-1AKM	-	646-623	TCAACAACTTTACCTGCATCTCTA	50.6		
	VP2F-2AKM	+	600-622	ATGGCTGTAGAAAATAAAAACCTC	46.7	47	920
	VP2R-2AKM	-	1519-1496	TCAATATCCATCAATAACTACTT	44.6		
	VP2F-3AKM	+	1404-1425	AGACTCCGTTTCAATAGCAGA	50.3	47	881
	VP2R-3AKM	-	2284-2265	TTGCGCATAGTCTCCTGTTC	50.4		
	VP2F-4AKM2	+	2131-2150	TTCAGATAAAAATTGCTCAAG	42.4	47	586
	GEN-VP2R <sup>2</sup>	-	2717-2699	GGTCATATCTCCACAGTGG	44.0		
VP3	VP3-F <sup>1</sup>	+	1-20	GGCTATTAAGCAGTACTAG	38.7	44	976
	VP3-R1ALT**	-	976-953	GTTTCAAATGTAATAATCAGAATAC	44.8		
	VP3F-2Alt	+	790-812	AGAGATGGCTAGGTAAGAGTA	46.2	46	960
	VP3R-2Alt	-	1749-1723	TGATCTAGTTTATAATTTTGTTCAGTT	45.8		
	VP3F-3AKM	+	1628-1651	GGATGTTTAGGTATATTTGGATTA	46.7	46	963
	VP3R-5AKM	-	2591-2571	GGTCACATCATGACTAGTGTG	44.4		
VP4	VP4-5'EndF	+	-19-8	T <sub>19</sub> GGCTATAA	50.0	47	259
	VP4-5'EndR1	-	240-216	TAGTTGGTTGATGTAATATTAGGTGG	52.5		
	VP4-1a <sup>1</sup>	+	2-23	GGCTATAAAAAGGCTTCGCCTCA	55.3	50	958
	VP4R-1ALT	-	959-936	TGCAGTTACTTSTTCDCCATCWCK	51.3		
	VP4F-2Alt	+	755-776	GAGAGCAC AAGTTAATGAAGAT	45.0	44	876
	VP4R-2Alt	-	1630-1612	ATTGTRCTTTTATTTCCTG	37.3		
	VP4F-3AKM	+	1545-1562	CAAGAAATAGCAATGCA	37.9	44	843
	VP4R-4ALT	-	2378-2356	TTTGGTCACATCCSARWHGRGC	53.4		
VP6	Gen-VP6F <sup>2</sup>	+	1-20	GGCTTTTAAACGAAGCTTTC	45.6	46	761
	VP6R-1AKM	-	762-746	GCTGAATTAATCACACTTGG	45.1		
	GEN-VP6RFor	+	592-610	GTCGCTGGATTGATTACT	43.9	50	766
Gen-VP6R <sup>2</sup>	-	1356-1340	GGTCACATCTCTCACTAT	36.9			
VP7	Beg9 <sup>1</sup>	+	1-28	GGCTTTAAAGAGAGAATTTCCGTTTGG	60.7	42	1062
	End9 <sup>1</sup>	-	1062-1036	GGTCACATCATACAATTCATCTAAG	49.9		
NSP1	NSP1-F <sup>1</sup>	+	1-22	GGCTTTTTTATGAAAAGTCTT	46.8	45	970
	NSP1-FRev	-	970-945	ATTTGGTTTAATTAATTTAGATACTA	43.8		
	NSP1F-2AKM	+	621-643	CAGAAATGGCATACTACCAGCAAC	50.3	50	999
	NSP1R-2AKM	-	1619-1597	CACATTTTATGCTGCC TAGGCCG	59.7		
NSP2	Gen-NSP2F <sup>2</sup>	+	1-19	GGCTTTTAAAGCGTCTCAG	47.1	50	1059
	Gen-NSP2R <sup>2</sup>	-	1059-1042	GGTCACATAAGCGCTTTC	45.2		
NSP3	GEN-NSP3F <sup>2</sup>	+	1-21	GGCTTTTAAATGCTTTTCAGTG	49.1	50	1105
	NSP3-R <sup>1</sup>	-	1105-1084	GGTCACATAACGCCCTATAGC	54.4		
NSP4	NSP4-1a <sup>1</sup>	+	1-22	GGCTTTTAAAGTTCGTTCCTG	52.0	46	752
	NSP4-R <sup>1</sup>	-	752-729	GGTCACATTAAGACCATTCCCTTCC	54.7		
NSP5	Gen-NSP5F <sup>2</sup>	+	1-19	GGCTTTTAAAGCGCTACAG	47.6	44	823
	GenNSP5R <sup>2</sup>	-	667-651	GGTCACAAAACGGGAGT	44.3		

<sup>1</sup>. Tsugawa, T. *et al.* 2008 <sup>2</sup>. Matthijnssen, J. *et al.* 2008 <sup>3</sup>. Gouvea, V. *et al.* 1990. <sup>4</sup>. Kudo S. *et al.* 2001.

\*If required

\*\*Rarely, some DS-1-like strains fail with this primer. In that case use VP3R1ALDS-1 5'TCGCCATTCTCCAGTC 3' Length 1111, annealing temperature 45°C.

Table A3: Type-strain genotypes for all 11 rotavirus segments, except for vaccine strains, in which VP7 and VP4 segments were relevant and available.

Strain Name	VP7	VP4	VP6	VP1	VP2	VP3	NSP1	NSP2	NSP3	NSP4	NSP5
Wa	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1
Ku	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1
DS-1	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
AU-1	G3	P[9]	I3	R3	C3	M3	A3	N3	T3	E3	H3
B1711	G6	P[6]	I2	R2	C2	M2	A2	N2	T2	E2	H2
Matlab13-03	G12	P[6]	I1	R1	C1	M1	A1	N1	T2	E1	H1
L26	G12	P[4]	I2	R2	C2	M1/M2	A2	N1	T2	E2	H1
RMC321	G9	P[19]	I5	R1	C1	M1	A1	N1	T1	E1	H1
116E	G9	P[11]	I1	R1	C1	M1	A1	N1	T1	E1	H1
69M	G8	P[10]	I2	R2	C2	M2	A2	N2	T2	E2	H2
IAL28	G5	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1
TB-Chen	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2
T152	G12	P[9]	I3	R3	C3	M3	A12	N3	T3	E3	H6
RV176-00	G12	P[6]	I2	R2	C2	M2	A2	N2	T2	E6	H2
P	G3	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1
ST-3	G4	P[6]	I1	R1	C1	M1	A1	N1	T1	E1	H1
WI79-9	G1	P[5]									
SC2-9	G2	P[5]									
WI78-8	G3	P[5]									
BrB-9	G4	P[5]									
WI79-4	G6	P[8]									
89-12	G1	P[8]									
Po-13	G18	P[17]	I4	R4	C4	M4	A4	N4	T4	E4	H4

Table A4: Type-strain accession numbers for all 11 rotavirus segment sequences, except for vaccine strains, in which VP7 and VP4 segments were relevant and available.

Full Strain Name	Strain Name	VP7	VP4	VP6	VP1	VP2	VP3	NSP1	NSP2	NSP3	NSP4	NSP5
RVA/Human-tc/USA/Wa/1974/G1P1A[8]	Wa	ROHVP7A	ROHVP4OCP	RO2SE68	DQ490539	X14942	AY267335	L18943	L04534	X81434	K02032	AF306494
RVA/Human-tc/JPN/KU/1988/G1P1A[8]	Ku	ROHVP71	AB222784	AB022768	AB022765	AB022766	AB022767	AB022769	AB022770	AB022771	AB022772	AB022773
RVA/Human-tc/USA/DS-1/1976/G2P1B[4]	DS-1	EF672581	EF672577	EF583028	EF583025	EF583026	EF583027	EF672578	EF672580	EF672579	EF672582	EF672583
RVA/Human-tc/JPN/AU-1/1982/G3P3[9]	AU-1	D86271	D10970	DQ490538	DQ490533	DQ490536	DQ490537	D45244	DQ490534	DQ490535	D89873	AB008656
RVA/Human-wt/BEL/B1711/2002/G6P[6]	B1711	EF554087	EF554085	EF554086	EF554082	EF554083	EF554084	EF554088	EF554090	EF554091	EF554092	EF554092
RVA/Human-tc/BGD/Matlab13/2003/G12P[6]	Matlab13-03	DQ146676	DQ146674	DQ146675	DQ146671	DQ146672	DQ146673	DQ146677	DQ146678	DQ146679	DQ146680	DQ146681
RVA/Human-tc/PHL/L26/1987/G12P[4]	L26	M58290	EF672591	EF583036	EF583033	EF583034	EF583035	EF672592	EF672594	EF672593	EF672596	DQ146698
RVA/Human-wt/IDN/RMC321/1990/G9P[19]	RMC321	AF501578	AF523677	AF531913	AY601114	AY601115	AY601116	AF506292	AF506293	AF541920	AF541921	AY033396
RVA/Human-tc/IDN/69M/1985/G9P[11]	116E	FJ361209	L07934	FJ361206	FJ361201	FJ361202	FJ361203	FJ361205	FJ361208	FJ361207	GAU78558	FJ361211
RVA/Human-tc/IDN/69M/1980/G8P4[10]	69M	EF672560	EF672556	EF583016	EF576937	EF583014	AY277916	D38151	EF672559	EF672558	EF672561	EF672562
RVA/Human-tc/BRA/IAL28/1992/G5P[8]	IAL28	EF672588	EF672584	EF583032	EF583029	EF583030	EF583031	EF672585	EF672587	EF672586	EF672589	EF672590
RVA/Human-tc/CHN/TB-Chen/1996/G2P[4]	TB-Chen	AY787646	AY787644	AY787645	AY787653	AY787652	AY787654	AY787647	AY787648	AY787649	AY787650	AY787651
RVA/Human-tc/THA/T152/1998/G12P[9]	T152	AB071404	AB077766	DQ146702	DQ146699	DQ146700	DQ146701	AB097459	DQ146703	DQ146704	DQ146705	DQ146706
RVA/Human-wt/BGD/RV176-00/2000/G12P[6]	RV176-00	DQ490556	DQ490554	DQ490555	DQ490551	DQ490552	DQ490553	DQ490557	DQ490558	DQ490559	DQ490560	DQ490561
RVA/Human-tc/USA/P/1974/G3P1A[8]	P	AB118024	AJ540228	EF583040	EF583037	EF583038	EF583039	EF672599	EF672601	EF672600	EF672603	EF672604
RVA/Human-tc/GBR/ST3/1975/G4P2A[6]	ST3	EF672616	L33895	EF583048	EF583045	EF583046	AY277919	EF672613	EF672615	EF672614	U59110	EF672618
RVA/Vaccine/USA/RotaTeq-WI79-9/1992/G1P7[5]	WI79-9	GU565057	GU565055									
RVA/Vaccine/USA/RotaTeq-SC2-9/1992/G2P7[5]	SC2-9	GU565068	GU565066									
RVA/Vaccine/USA/RotaTeq-WI78-8/1992/G3P7[5]	WI78-8	GU565079	GU565077									
RVA/Vaccine/USA/RotaTeq-BrB-9/1996/G4P7[5]	BrB-9	GU565090	GU565088									
RVA/Vaccine/USA/RotaTeq-WI79-4/1992/G6P1A[8]	WI79-4	GU565046	GU565044									
RVA/Vaccine/USA/Rotarix-A41CB052A/1988/G1P1A[8]	89-12	JN849114	JN849113									
RVA/Pigeon-tc/JPN/PO-13/1983/G18P[17]	Po-13	D82979	AB009632	AROVP6	AB009629	AB009630	AB009631	AB009633	AB009625	AB009626	AB009627	AB009628

Table A5: Raw data sorted from West to East by vaccine site. Mixed infections are included as rotavirus-positive samples but not included in genotype counts.

City	Year	G1 P[8]	G2 P[4]	G3 P[8]	G4 P[8]	G9 P[8]	Other	Mixed Infection
<b>Vancouver</b>	2010	9	3	5	0	0	0	0
<b>Edmonton</b>	2007	13	4	0	0	0	0	6
<b>Saskatoon</b>	2010	5	1	2	0	0	1	0
<b>Winnipeg</b>	2007	3	0	0	0	0	0	0
	2010	2	1	1	0	0	0	0
<b>Toronto</b>	2010	3	0	0	0	0	0	0
	2007	25	1	1	1	0	0	0
<b>Ottawa</b>	2008	16	3	4	0	0	0	1
	2009	20	3	0	0	12	0	1
	2010	3	1	2	0	0	0	0
<b>Quebec City</b>	2007	15	1	7	0	0	0	0
	2008	7	1	0	0	0	0	0
	2009	6	2	5	0	0	0	2
	2007	5	1	1	0	1	0	1
<b>Halifax</b>	2008	29	0	0	0	0	0	0
	2009	3	0	17	2	0	0	1
	2010	10	0	0	0	0	1	0
<b>Total</b>		174	22	45	3	13	2	12

Table A6: CanadianSample strain accession numbers and full names on genbank.

Full Strain Name	Strain Name	VP7	VP4	VP6	VP1	VP2	VP3	NSP1	NSP2	NSP3	NSP4	NSP5
RVA/Human-wi/CAN/RT002-07/2007/G3P8	RT002-07	JQ069458	JQ069621	JQ069537	JQ069872	JQ069788	JQ069704	JQ069374	JQ069290	JQ069206	JQ069122	JQ069038
RVA/Human-wi/CAN/RT004-07/2007/G3P8	RT004-07	JQ069459	JQ069622	JQ069538	JQ069873	JQ069789	JQ069705	JQ069375	JQ069291	JQ069207	JQ069123	JQ069039
RVA/Human-wi/CAN/RT005-07/2007/G1P8	RT005-07	JQ069460	JQ069623	JQ069539	JQ069874	JQ069790	JQ069706	JQ069376	JQ069292	JQ069208	JQ069124	JQ069040
RVA/Human-wi/CAN/RT006-07/2007/G1P8	RT006-07	JQ069461	JQ069624	JQ069540	JQ069875	JQ069791	JQ069707	JQ069377	JQ069293	JQ069209	JQ069125	JQ069041
RVA/Human-wi/CAN/RT008-07/2007/G2P4	RT008-07	JQ069462	JQ069625	JQ069541	JQ069876	JQ069792	JQ069708	JQ069378	JQ069294	JQ069210	JQ069126	JQ069042
RVA/Human-wi/CAN/RT012-07/2007/G1P8	RT012-07	JQ069463	JQ069626	JQ069542	JQ069877	JQ069793	JQ069709	JQ069379	JQ069295	JQ069211	JQ069127	JQ069043
RVA/Human-wi/CAN/RT013-07/2007/G3P8	RT013-07	JQ069464	JQ069627	JQ069543	JQ069878	JQ069794	JQ069710	JQ069380	JQ069296	JQ069212	JQ069128	JQ069044
RVA/Human-wi/CAN/RT015-07/2007/G1P8	RT015-07	JQ069465	JQ069628	JQ069544	JQ069879	JQ069795	JQ069711	JQ069381	JQ069297	JQ069213	JQ069129	JQ069045
RVA/Human-wi/CAN/RT016-07/2007/G1P8	RT016-07	JQ069466	JQ069629	JQ069545	JQ069880	JQ069796	JQ069712	JQ069382	JQ069298	JQ069214	JQ069130	JQ069046
RVA/Human-wi/CAN/RT018-07/2007/G1P8	RT018-07	JQ069467	JQ069630	JQ069546	JQ069881	JQ069797	JQ069713	JQ069383	JQ069299	JQ069215	JQ069131	JQ069047
RVA/Human-wi/CAN/RT019-07/2007/G1P8	RT019-07	JQ069468	JQ069631	JQ069547	JQ069882	JQ069798	JQ069714	JQ069384	JQ069300	JQ069216	JQ069132	JQ069048
RVA/Human-wi/CAN/RT025-07/2007/G3P8	RT025-07	JQ069469	JQ069632	JQ069548	JQ069883	JQ069799	JQ069715	JQ069385	JQ069301	JQ069217	JQ069133	JQ069049
RVA/Human-wi/CAN/RT028-07/2007/G1P8	RT028-07	JQ069470	JQ069633	JQ069549	JQ069884	JQ069800	JQ069716	JQ069386	JQ069302	JQ069218	JQ069134	JQ069050
RVA/Human-wi/CAN/RT031-07/2007/G1P8	RT031-07	JQ069471	JQ069634	JQ069550	JQ069885	JQ069801	JQ069717	JQ069387	JQ069303	JQ069219	JQ069135	JQ069051
RVA/Human-wi/CAN/RT032-07/2007/G1P8	RT032-07	JQ069472	JQ069635	JQ069551	JQ069886	JQ069802	JQ069718	JQ069388	JQ069304	JQ069220	JQ069136	JQ069052
RVA/Human-wi/CAN/RT033-07/2007/G1P8	RT033-07	JQ069473	JQ069636	JQ069552	JQ069887	JQ069803	JQ069719	JQ069389	JQ069305	JQ069221	JQ069137	JQ069053
RVA/Human-wi/CAN/RT034-07/2007/G9P8	RT034-07	JF964998	JQ069637	JQ069553	JQ069888	JQ069804	JQ069720	JQ069390	JQ069306	JQ069222	JQ069138	JQ069054
RVA/Human-wi/CAN/RT036-07/2007/G2P4	RT036-07	JQ069474	JQ069638	JQ069554	JQ069889	JQ069805	JQ069721	JQ069391	JQ069307	JQ069223	JQ069139	JQ069055
RVA/Human-wi/CAN/RT037-07/2007/G3P8	RT037-07	JQ069475	JQ069639	JQ069555	JQ069890	JQ069806	JQ069722	JQ069392	JQ069308	JQ069224	JQ069140	JQ069056
RVA/Human-wi/CAN/RT048-07/2007/G4P8	RT048-07	JQ069476	JQ069640	JQ069556	JQ069891	JQ069807	JQ069723	JQ069393	JQ069309	JQ069225	JQ069141	JQ069057
RVA/Human-wi/CAN/RT049-07/2007/G1P8	RT049-07	JQ069477	JQ069641	JQ069557	JQ069892	JQ069808	JQ069724	JQ069394	JQ069310	JQ069226	JQ069142	JQ069058
RVA/Human-wi/CAN/RT054-07/2007/G1P8	RT054-07	JQ069478	JQ069642	JQ069558	JQ069893	JQ069809	JQ069725	JQ069395	JQ069311	JQ069227	JQ069143	JQ069059
RVA/Human-wi/CAN/RT055-07/2007/G1P8	RT055-07	JQ069479	JQ069643	JQ069559	JQ069894	JQ069810	JQ069726	JQ069396	JQ069312	JQ069228	JQ069144	JQ069060
RVA/Human-wi/CAN/RT059-07/2007/G3P8	RT059-07	JQ069480	JQ069644	JQ069560	JQ069895	JQ069811	JQ069727	JQ069397	JQ069313	JQ069229	JQ069145	JQ069061
RVA/Human-wi/CAN/RT061-07/2007/G1P8	RT061-07	JQ069481	JQ069645	JQ069561	JQ069896	JQ069812	JQ069728	JQ069398	JQ069314	JQ069230	JQ069146	JQ069062
RVA/Human-wi/CAN/RT064-07/2007/G1P8	RT064-07	JQ069482	JQ069646	JQ069562	JQ069897	JQ069813	JQ069729	JQ069399	JQ069315	JQ069231	JQ069147	JQ069063
RVA/Human-wi/CAN/RT065-07/2007/G1P8	RT065-07	JQ069483	JQ069647	JQ069563	JQ069898	JQ069814	JQ069730	JQ069400	JQ069316	JQ069232	JQ069148	JQ069064
RVA/Human-wi/CAN/RT066-07/2007/G1P8	RT066-07	JQ069484	JQ069648	JQ069564	JQ069899	JQ069815	JQ069731	JQ069401	JQ069317	JQ069233	JQ069149	JQ069065
RVA/Human-wi/CAN/RT068-07/2007/G2P4	RT068-07	JQ069485	JQ069649	JQ069565	JQ069900	JQ069816	JQ069732	JQ069402	JQ069318	JQ069234	JQ069150	JQ069066
RVA/Human-wi/CAN/RT072-07/2007/G1P8	RT072-07	JQ069486	JQ069650	JQ069566	JQ069901	JQ069817	JQ069733	JQ069403	JQ069319	JQ069235	JQ069151	JQ069067
RVA/Human-wi/CAN/RT074-07/2007/G1P8	RT074-07	JQ069487	JQ069651	JQ069567	JQ069902	JQ069818	JQ069734	JQ069404	JQ069320	JQ069236	JQ069152	JQ069068
RVA/Human-wi/CAN/RT077-07/2007/G1P8	RT077-07	JQ069488	JQ069652	JQ069568	JQ069903	JQ069819	JQ069735	JQ069405	JQ069321	JQ069237	JQ069153	JQ069069
RVA/Human-wi/CAN/RT092-07/2007/G1P8	RT092-07	JQ069489	JQ069653	JQ069569	JQ069904	JQ069820	JQ069736	JQ069406	JQ069322	JQ069238	JQ069154	JQ069070
RVA/Human-wi/CAN/RT094-07/2008/G1P8	RT094-07	JQ069490	JQ069654	JQ069570	JQ069905	JQ069821	JQ069737	JQ069407	JQ069323	JQ069239	JQ069155	JQ069071
RVA/Human-wi/CAN/RT095-07/2008/G1P8	RT095-07	JQ069491	JQ069655	JQ069571	JQ069906	JQ069822	JQ069738	JQ069408	JQ069324	JQ069240	JQ069156	JQ069072
RVA/Human-wi/CAN/RT096-07/2008/G2P4	RT096-07	JQ069492	JQ231221	JQ069572	JQ069907	JQ069823	JQ069739	JQ069409	JQ069325	JQ069241	JQ069157	JQ069073
RVA/Human-wi/CAN/RT097-07/2008/G1P8	RT097-07	JQ069493	JQ069656	JQ069573	JQ069908	JQ069824	JQ069740	JQ069410	JQ069326	JQ069242	JQ069158	JQ069074
RVA/Human-wi/CAN/RT098-07/2008/G1P8	RT098-07	JQ069494	JQ069657	JQ069574	JQ069909	JQ069825	JQ069741	JQ069411	JQ069327	JQ069243	JQ069159	JQ069075
RVA/Human-wi/CAN/RT099-07/2008/G1P8	RT099-07	JQ069495	JQ069658	JQ069575	JQ069910	JQ069826	JQ069742	JQ069412	JQ069328	JQ069244	JQ069160	JQ069076
RVA/Human-wi/CAN/RT101-07/2008/G1P8	RT101-07	JQ069496	JQ069659	JQ069576	JQ069911	JQ069827	JQ069743	JQ069413	JQ069329	JQ069245	JQ069161	JQ069077
RVA/Human-wi/CAN/RT102-07/2008/G1P8	RT102-07	JQ069497	JQ069660	JQ069577	JQ069912	JQ069828	JQ069744	JQ069414	JQ069330	JQ069246	JQ069162	JQ069078
RVA/Human-wi/CAN/RT103-07/2008/G1P8	RT103-07	JQ069498	JQ069661	JQ069578	JQ069913	JQ069829	JQ069745	JQ069415	JQ069331	JQ069247	JQ069163	JQ069079
RVA/Human-wi/CAN/RT118-07/2008/G1P8	RT118-07	JQ069499	JQ069662	JQ069579	JQ069914	JQ069830	JQ069746	JQ069416	JQ069332	JQ069248	JQ069164	JQ069080
RVA/Human-wi/CAN/RT120-07/2008/G1P8	RT120-07	JQ069500	JQ069663	JQ069580	JQ069915	JQ069831	JQ069747	JQ069417	JQ069333	JQ069249	JQ069165	JQ069081
RVA/Human-wi/CAN/RT121-07/2008/G3P8	RT121-07	JQ069501	JQ069664	JQ069581	JQ069916	JQ069832	JQ069748	JQ069418	JQ069334	JQ069250	JQ069166	JQ069082
RVA/Human-wi/CAN/RT122-07/2008/G1P8	RT122-07	JQ069502	JQ069665	JQ069582	JQ069917	JQ069833	JQ069749	JQ069419	JQ069335	JQ069251	JQ069167	JQ069083
RVA/Human-wi/CAN/RT124-07/2008/G1P8	RT124-07	JQ069503	JQ069666	JQ069583	JQ069918	JQ069834	JQ069750	JQ069420	JQ069336	JQ069252	JQ069168	JQ069084
RVA/Human-wi/CAN/RT125-07/2008/G2P4	RT125-07	JQ069504	JQ069667	JQ069584	JQ069919	JQ069835	JQ069751	JQ069421	JQ069337	JQ069253	JQ069169	JQ069085
RVA/Human-wi/CAN/RT128-07/2008/G2P4	RT128-07	JQ069505	JQ069668	JQ069585	JQ069920	JQ069836	JQ069752	JQ069422	JQ069338	JQ069254	JQ069170	JQ069086
RVA/Human-wi/CAN/RT131-07/2008/G3P8	RT131-07	JQ069506	JQ069669	JQ069586	JQ069921	JQ069837	JQ069753	JQ069423	JQ069339	JQ069255	JQ069171	JQ069087
RVA/Human-wi/CAN/RT133-07/2008/G1P8	RT133-07	JQ069507	JQ069670	JQ069587	JQ069922	JQ069838	JQ069754	JQ069424	JQ069340	JQ069256	JQ069172	JQ069088
RVA/Human-wi/CAN/RT134-07/2008/G1P8	RT134-07	JQ069508	JQ069671	JQ069588	JQ069923	JQ069839	JQ069755	JQ069425	JQ069341	JQ069257	JQ069173	JQ069089
RVA/Human-wi/CAN/RT135-07/2008/G1P8	RT135-07	JQ069509	JQ069672	JQ069589	JQ069924	JQ069840	JQ069756	JQ069426	JQ069342	JQ069258	JQ069174	JQ069090
RVA/Human-wi/CAN/RT172-07/2008/G1P8	RT172-07	JQ069510	JQ069673	JQ069590	JQ069925	JQ069841	JQ069757	JQ069427	JQ069343	JQ069259	JQ069175	JQ069093
RVA/Human-wi/CAN/RT177-07/2008/G1P8	RT177-07	JQ069511	JQ069674	JQ069591	JQ069926	JQ069842	JQ069758	JQ069428	JQ069344	JQ069260	JQ069176	JQ069091
RVA/Human-wi/CAN/RT178-07/2008/G1P8	RT178-07	JQ069512	JQ069675	JQ069592	JQ069927	JQ069843	JQ069759	JQ069429	JQ069345	JQ069261	JQ069177	JQ069092
RVA/Human-wi/CAN/RT179-07/2008/G1P8	RT179-07	JQ069513	JQ069676	JQ069593	JQ069928	JQ069844	JQ069760	JQ069430	JQ069346	JQ069262	JQ069178	JQ069094
RVA/Human-wi/CAN/RT185-07/2008/G1P8	RT185-07	JQ069514	JQ069677	JQ069594	JQ069929	JQ069845	JQ069761	JQ069431	JQ069347	JQ069263	JQ069179	JQ069095
RVA/Human-wi/CAN/RT186-07/2008/G1P8	RT186-07	JQ069515	JQ069678	JQ069595	JQ069930	JQ069846	JQ069762	JQ069432	JQ069348	JQ069264	JQ069180	JQ069096
RVA/Human-wi/CAN/RT187-07/2008/G1P8	RT187-07	JQ069516	JQ069679	JQ069596	JQ069931	JQ069847	JQ069763	JQ069433	JQ069349	JQ069265	JQ069181	JQ069097
RVA/Human-wi/CAN/RT188-07/2008/G1P8	RT188-07	JQ069517	JQ069680	JQ069597	JQ069932	JQ069848	JQ069764	JQ069434	JQ069350	JQ069266	JQ069182	JQ069098
RVA/Human-wi/CAN/RT193-07/2008/G1P8	RT193-07	JQ069518	JQ069681	JQ069598	JQ069933	JQ069849	JQ069765	JQ069435	JQ069351	JQ069267	JQ069183	JQ069099
RVA/Human-wi/CAN/RT004-09/2009/G3P8	RT004-09	JQ069519	JQ069682	JQ069599	JQ069934	JQ069850	JQ069766	JQ069436	JQ069352	JQ069268	JQ069184	JQ069100
RVA/Human-wi/CAN/RT005-09/2009/G1P8	RT005-09	JQ069520	JQ069683	JQ069600	JQ069935	JQ069851	JQ069767	JQ069437	JQ069353	JQ069269	JQ069185	JQ069101
RVA/Human-wi/CAN/RT008-09/2009/G2P4	RT008-09	JQ069521	JQ069684	JQ069601	JQ069936	JQ069852	JQ069768	JQ069438	JQ069354	JQ069270	JQ069186	JQ069102
RVA/Human-wi/CAN/RT010-09/2009/G3P8	RT010-09	JQ069522	JQ069685	JQ069602	JQ069937	JQ069853	JQ069769	JQ069439	JQ069355	JQ069271	JQ069187	JQ069103
RVA/Human-wi/CAN/RT017-09/2009/G3P8	RT017-09	JQ069523	JQ069686	JQ069603	JQ069938	JQ069854	JQ069770	JQ069440	JQ069356	JQ069272	JQ069188	JQ069104
RVA/Human-wi/CAN/RT018-09/2009/G3P8	RT018-09	JQ069524	JQ069687	JQ069604	JQ069939	JQ069855	JQ069771	JQ069441	JQ069357	JQ069273	JQ069189	JQ069105
RVA/Human-wi/CAN/RT019-09/2009/G3P8	RT019-09	JQ069525	JQ0696									

Table A7: McNemar's test contingency table to analyze reassortments between genogroups. Reassortment events were non-existent so no significant evidence of regular reassortment between any genome segments was found for genogroup II strains.

	C1	Not C1	Row Total
All Wa-like (excl. VP7)	76	0	76
Not all Wa-like (excl. VP7)	0	8	8
Column Total	76	8	84
Chi-Squared	0, Accept null hypothesis that marginal proportions are not different from one another		

Table A8: McNemar's test contingency table to analyze reassortments between genogroups. Reassortment events were non-existent so no significant evidence of regular reassortment between any genome segments was found for genogroup I strains.

	C2	Not C2	Row Total
All DS-1-like	8	0	8
Not all DS-1-like	0	76	76
Column Total	8	76	84
Chi Squared	0, Accept null hypothesis that marginal proportions are not different from one another		