Early-life Human Cytomegalovirus Infections in Canadians: Implications for a Healthy Development

by

KAREN EVELYN SIMMONS

B.Sc., University of British Columbia, 2014

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies
(Pathology and Laboratory Medicine)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

April 2019

© Karen Evelyn Simmons, 2019
The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, a thesis/dissertation entitled:

Early-life Human Cytomegalovirus Infections in Canadians: Implications for a Healthy Development

submitted by Karen Simmons in partial fulfillment of the requirements for
the degree of Master of Science
in The Faculty of Graduate and Postdoctoral Studies (Pathology and Laboratory Medicine)

Examining Committee:

Soren Gantt
Supervisor

Tobias Kollmann
Supervisory Committee Member

Peter van den Elzen
Department of Pathology Examiner

Bruce Vallance
Additional Examiner

Additional Supervisory Committee Members:

Marc Horwitz
Supervisory Committee Member
Abstract

**Background:** Cytomegalovirus (CMV) is a ubiquitous herpesvirus that has evolved with humans over millions of years and causes lifelong infection. Therefore, CMV is considered part of the “virome”, the viral component of the human microbiome. The bacterial intestinal microbiome is essential for the education of the immune system and appears to be critical for immunologic health, however, much less is known about the importance of the virome on healthy development. CMV has potent immunomodulatory effects and can replicate within the intestines. Thus, CMV might impact the composition of the bacterial intestinal microbiome and its relationship with the developing immune system. CMV infections are occurring later in life in high income countries like Canada, which also have a high prevalence of allergic diseases. Indeed, CMV could be a necessary component of the virome to educate the immune system and prevent the later development of allergic diseases, such as asthma and eczema. *Thus, we hypothesized that infection with CMV in healthy Canadian infants impacts the composition of the intestinal microbiome and protects against the development of allergic diseases.*

**Methods:** Urine samples collected from CHILD Study participants at 3 and 12 months of age were tested for CMV by qPCR. The CHILD Study questionnaire, fecal 16S rRNA sequence and clinical diagnoses data were used to determine risk factors for infection, CMV impact on the microbiome and CMV associations with allergic diseases, respectively.

**Results:** Of the 1,151 Canadian children tested, 19% were infected by 1 year of age. The duration of breastfeeding and parental ethnicity was associated with the risk of infant CMV infection. CMV infection by 3 months of age was associated with a significantly decreased α-diversity and altered β-diversity of the intestinal microbiome. CMV infection was not associated
with a clinical diagnosis of asthma or eczema but was associated with a higher risk of an atopic phenotype at 1 year of age.

**Significance:** This study provides the first evidence that early-life CMV infection influences the diversity of the intestinal microbiome. Additionally, it provides supportive evidence that elimination of CMV infection through vaccination is unlikely to contribute to allergic disease development.
Lay Summary

Cytomegalovirus (CMV) is a herpesvirus that infects most of the world’s population and has strong effects on the immune system. In areas where allergic diseases have been increasing, the age of first CMV infection has also been increasing. The immune system and the gut microbiome (healthy bacteria that live in the gastrointestinal tract) interact and develop together throughout the first years of life. Therefore, early-life CMV infection might impact the composition of the gut microbiome and help prevent the development of allergic diseases.

CMV infection status was compared to bacterial species in the gut and allergic disease outcomes in healthy Canadian children (childstudy.ca). Early-life CMV infection decreased the number of bacteria in the gut and was associated with more allergies.

This study provides the first evidence that early-life CMV infection influences the gut microbiome and that elimination of CMV through vaccination would not promote the development of allergic diseases.
Preface

The contents of this thesis are my own original work, including the images and figures presented. The experiments were designed by me in coordination with my supervisor, Dr. Soren Gantt.

British Columbia Provincial Health Services Authority microbiology laboratory previously designed and verified the primer and probe sequences, the CMV mini-gene for quantification and the pUC19 method of internal control that were used to identify CMV in urine samples by quantitative PCR. Selection of CHILD fecal samples sent for 16S rRNA sequencing was done by our CHILD Study collaborators, Drs. Stuart Turvey and Hind Sbihi prior to the start of this study. The 16S rRNA sequencing was carried out by the company Microbiome Insights based at the University of British Columbia.

Informed consent was obtained from each CHILD Study participant prior to recruitment and sample collection as per the protocols approved by The University of British Columbia Research Ethics Board (H07-03120).

All DNA extractions and qPCR testing, described in detail in Chapter 2, were performed in full by me. The statistical analyses described in Chapters 2 – 4 were performed in full by me with guidance from Dr. Hind Sbihi.
Table of Contents

Abstract ........................................................................................................................................... iii
Lay Summary ..................................................................................................................................... v
Preface ............................................................................................................................................. vi
Table of Contents ........................................................................................................................... vii
List of Tables ................................................................................................................................... ix
List of Figures ................................................................................................................................... x
List of Symbols ................................................................................................................................ xi
List of Abbreviations ........................................................................................................................ xii
Acknowledgments ........................................................................................................................... xiv
Dedication ......................................................................................................................................... xv

Chapter 1. Introduction .................................................................................................................... 1
  1.1. Human cytomegalovirus ......................................................................................................... 1
  1.2. Immune response and CMV immunomodulation .................................................................. 3
  1.3. The systemic influence of CMV ............................................................................................. 6
  1.4 The gut microbiome and immunological development ............................................................ 8
  1.5 Atopic (allergic) diseases and early-life intestinal microbiome dysbiosis ............................... 9
  1.6 Linking CMV infection and allergic diseases through the hygiene hypothesis ...................... 11
  1.7. Hypothesis, objectives and approach ................................................................................... 13

Chapter 2. CMV Prevalence and risk factors .................................................................................... 15
  2.1. Introduction ............................................................................................................................ 15
  2.1.1. Prevalence of CMV in Canada .......................................................................................... 15
  2.1.2. Detecting CMV ................................................................................................................ 15
  2.1.3. CMV risk factors of infection ............................................................................................ 16
  2.1.3. Approach .......................................................................................................................... 17
  2.2. Methods .................................................................................................................................. 17
  2.2.1. The Canadian Healthy Infant Longitudinal Development Study ...................................... 17
  2.2.2. Sample selection .............................................................................................................. 18
  2.2.3. Determination of CMV Status ......................................................................................... 18
  2.2.4. Determination of risk factors for CMV infection .............................................................. 22
  2.3. Results .................................................................................................................................... 22
  2.3.1. Prevalence of CMV in healthy Canadian children ............................................................. 22
  2.3.2. Risk factors for CMV infection by 1-year of age ............................................................... 24
  2.4. Discussion .............................................................................................................................. 26

Chapter 3. CMV and the intestinal bacterial microbiome ................................................................... 30
  3.1. Introduction ............................................................................................................................ 30
  3.1.1. Characterizing the intestinal microbiome .......................................................................... 30
3.1.2. Approach .......................................................................................................................... 31
3.2. Methods ................................................................................................................................. 31
3.2.1. Samples collection and processing .................................................................................. 31
3.2.2. Analysis of 16S rRNA data ............................................................................................ 32
3.3 Results .................................................................................................................................... 34
3.3.1. Sample demographics ....................................................................................................... 34
3.3.2. α-diversity ......................................................................................................................... 35
3.3.3. β-diversity and relative abundance of specific phyla and genera ..................................... 39
3.4 Discussion ............................................................................................................................... 45

Chapter 4. CMV and atopic diseases ......................................................................................... 47
4.1. Introduction ............................................................................................................................ 47
4.2. Methods ................................................................................................................................ 47
4.2.1. Medical assessments ........................................................................................................ 47
4.2.2. Statistical analysis of associations .................................................................................... 48
4.3. Results .................................................................................................................................... 48
4.3.1. Association between CMV infection and allergic diseases .............................................. 48
4.3.2. Association of CMV infection and atopic phenotype ....................................................... 50
4.4 Discussion ............................................................................................................................... 52

Chapter 5. Conclusions .............................................................................................................. 54
References ................................................................................................................................. 57
List of Tables

Table 2.1. Demographics of CHILD Study sub-cohort .................................................................23

Table 2.2. Risk factors of CMV infection by 1 year of age .......................................................24

Table 3.1 Demographics of 16S rRNA-sequenced samples evaluated at 3 months and 1 year by
CMV status determined by qPCR of urine samples at the same time point as sequenced fecal
samples .......................................................................................................................................35

Table 3.2 Linear regression of FPD index using multiple factors that contribute to intestinal
microbiome α-diversity in the first 3-months of life .................................................................39
List of Figures

Figure 3.1.  Comparison of bacterial α-diversity of the intestinal microbiome as calculated using Faith’s phylogenetic diversity (FPD) between all children (CMV infected and uninfected) at 3 months and 1 year of age .................................................................37

Figure 3.2.  Comparison of bacterial α-diversity of the intestinal microbiome as calculated using Faith’s phylogenetic diversity (FPD) between CMV infected and uninfected children .....38

Figure 3.3.  Comparison of bacterial β-diversity of the intestinal microbiome as calculated by Weighted UniFrac distances followed by principal coordinate analysis (PCoA) between children at 3 months and 1 year of age ........................................................................................................41

Figure 3.4.  Comparison of bacterial β-diversity of the intestinal microbiome as calculated by Weighted UniFrac distances followed by principal coordinate analysis (PCoA) between CMV infected and uninfected children at a) 3 months and b) 1 year of age .......................................................42

Figure 3.5.  Comparison of the bacteria present in the intestinal microbiota at the phylum level between CMV infected and uninfected infants at 3 months of age .................................................43

Figure 3.6.  Comparison of the bacteria present in the intestinal microbiota at the genus level between CMV infected and uninfected infants at 3 months of age .................................................44

Figure 4.1.  Associative analysis of a.) asthma diagnosis and b.) eczema diagnosis at 3 years of age by CMV status at 1 year of age .........................................................................................................49

Figure 4.2.  Associative analysis of a.) asthma diagnosis and b.) eczema diagnosis at 3 years of age by CMV status at 3 months of age .........................................................................................................50

Figure 4.3.  Associative analysis of a.) CMV status at 1 year and b.) CMV status at 3 months with atopic phenotype at 1 year of age .........................................................................................................51
List of Symbols

\( \alpha \) - alpha
\( \beta \) - beta
List of Abbreviations

CD – cluster of differentiation
CHILD – Canadian Healthy Infant Longitudinal Development Study
CI – confidence interval
CMV – cytomegalovirus
cCMV – congenital cytomegalovirus
CWL – cesarean section with labour
CWOL – cesarean section without labour
EBV – Epstein-Barr virus
FDR – false discovery rate
FPD – Faith’s phylogenetic diversity
GF – germ free
IFN – interferon
IgE – immunoglobulin E
IgG – immunoglobulin G
IL – interleukin
ISAAC - International Study of Asthma and Allergies in Childhood
MCMV – murine cytomegalovirus
MHC – major histocompatibility complex
NK – natural killer
OR – odds ratio
OTU – operation taxonomic unit
PCoA – principal coordinate analysis
PCR – polymerase chain reaction
PHSA – Provincial Health Services Authority
qPCR – quantitative polymerase chain reaction
SPF – specific-pathogen free
TLR – toll-like receptor
UC – ulcerative colitis
Acknowledgments

I would like to extend my gratitude to Dr. Soren Gantt, for his support, encouragement and patience throughout my graduate studies adventure. Thanks to the Department of Pathology, for nominating and awarding me with the financial support to carry out my research. Dr. Hind Sbihi was a monumental resource for me in completing this thesis. In particular, I want to thank her for her guidance in navigating the CHILD Study and mentoring me in data analysis, but more importantly for introducing me to her yoga studio and becoming a close friend. Dr. Citlali Marquez has been an immense help and a mentor for me throughout my graduate studies, but mostly I want to thank her for her long-lasting friendship. Thank you to my family, for their patience and understanding of my extended absences. Lastly and mostly, my thanks to Derek, for his unwavering love and support, and for bringing home our kitten.
Dedication

For Derek and Artemis, we can now fully enjoy lazy-Sundays again.
Chapter 1. Introduction

1.1. Human cytomegalovirus

Cytomegalovirus (CMV) is a member of the β-herpesvirus family. It is an enveloped virus with a large double-stranded DNA genome, the largest of all known human viral pathogens at approximately 235 kilobase pairs\(^1,2\). This large genome encodes over 160 proteins that function not only to construct more virions but also to manipulate the host environment in which it replicates\(^1-4\). Through this manipulation, and millions of years of co-evolution with its host, CMV infects its host for life, usually without overt consequence. Thus, most CMV infections are asymptomatic, and the infection may never even be discovered. Horizontal CMV infection is acquired at mucosal epithelium, for example, through ingestion of virus shed in breast milk. Primary infection is associated with high levels of lytic replication and shedding of infectious virus from mucosal fluids. During primary infection, however, CMV also establishes latent infection of long-lived myeloid precursors, which allows lifelong infection of the host and can result in periodic lytic viral reactivation. Carriage of CMV throughout life can thus alternate between two different states, either in a latent state with little to no viral replication, or in episodes of lytic replication during which CMV virions are produced and shed at mucosal surfaces and passed onto new hosts\(^5\).

Although CMV infection is usually asymptomatic and seemingly benign, for immunocompromised individuals this is not the case. With impaired immunity, the balance between immune control of CMV and lytic virion production is disrupted which can lead to viremia and severe end-organ disease. Careful consideration of CMV status, of both donor and recipient, is imperative for transplant patients\(^6\). A transplant patient’s own CMV infection can reactivate due to iatrogenic immunosuppression, or CMV from a donated tissue can result in a
primary infection in a recipient. CMV remains the biggest single infectious consideration for transplantation as it can lead to tissue-invasive disease, allograft rejection and death. HIV/AIDS is another immunocompromised state that can be exploited by CMV infection, where CMV retinitis is the more common manifestation of uncontrolled CMV replication.

Infants born with congenital (c)CMV infections can suffer from severe disease and long-term consequences. Approximately 0.5-0.7% of live-birth infants are infected with CMV in-utero, and approximately 10% of infants with cCMV have severe disease manifestations including sensorineural hearing loss, cognitive impairment and retinitis. There is evidence that use of antivirals at birth can protect from some of these long-term consequences of cCMV infection. However, often an infected infant does not have symptoms at birth but presents with them later after damage has already occurred. Strategies to screen for cCMV at birth are an area of ongoing investigation. Most postnatal infections are asymptomatic and have not been associated with long-term disease outcome. However, some studies have shown that low-birth weight preterm infants that become infected early in life may suffer adverse outcomes. The above examples demonstrate that although CMV is a mostly asymptomatic disease, CMV-related complications result in it being a massive public health burden. As such, it was the highest ranked vaccine development priority by the US Institute of Medicine. There is optimism about the prospect of developing an effective CMV vaccine, but if successful it would still be many years away from clinical use.

CMV is prevalent in human populations across the globe, with 45-100% of adults infected, depending on geographical location. Seroprevalence, being the proportion of people with detectable CMV-specific antibodies in the blood, is highest in Africa, South America and Asia, where most children acquire the virus in the first few years of life. In Canada, the USA and
western Europe, the prevalence is lower, and people are infected later in life. In these high-income counties, the seroprevalence is only approximately 4-20% among pre-school aged children\textsuperscript{18–20} and 50 – 60%\textsuperscript{18,21} among adults. Relatively few studies have specifically looked at the prevalence of CMV in healthy children or infants. Two studies in Canada describe the prevalence in children under 3 years of age to be 4-8%\textsuperscript{18,19}. In a study in Bradford, UK the seroprevalence of children under 12 months of age was found to be 9-34% largely varying with ethnicity\textsuperscript{20}. In Croatia, a study found a CMV prevalence of 53% in children from 6 months to 9 years of age\textsuperscript{22}.

Other than age and geographic location, several other risk factors have consistently been found to be associated with CMV infection. Seroprevalence in non-white ethnic groups is consistently found to be 20 – 30% higher than in whites\textsuperscript{17,20,23,24}. A study done in Montreal, Canada characterized having a low or middle-income country of birth to be a risk factor for CMV infection\textsuperscript{21}, and other studies have also shown people of lower socioeconomic status to be 10 – 30% more likely to be infected\textsuperscript{17}. A proven risk factor is breast feeding. CMV is transmitted to infants through breast milk, and therefore breast feeding by CMV-infected mothers is a risk factor for infection in infants\textsuperscript{14,15,25–27}. Nevertheless, because most postnatal infections are asymptomatic the risk of CMV infection does not outweigh the benefits provided by breastfeeding\textsuperscript{14}.

1.2. Immune response and CMV immunomodulation

CMV is thought to be such a successful virus at infecting most people worldwide in part because it has evolved with humans over millions of years\textsuperscript{26,29}. CMV has evolved numerous mechanisms to hide from and manipulate the immune system, enabling it to hide within its host for life.
Upon infection with CMV, healthy people mount an effective immune response involving both the innate and adaptive immune systems. Pattern recognition receptors, such as Toll-like receptors (TLRs), can recognize CMV and trigger the production of pro-inflammatory cytokines within hours after infection. In particular, TLR2 has been found to recognize glycoproteins B and H on the surface of the virus and TLRs 3 and 9 recognize CMV DNA; the recognition of CMV by these TLRs leads to a response that increases the production of interleukin (IL)-6, -8, -12 and interferon (IFN)-β. The upregulation of these pro-inflammatory cytokines help fibroblast, endothelial and epithelial cells to resist infection with CMV by inhibiting viral replication and activating other immune effector cells. One such effector cell considered to be important in CMV infection control in the natural killer (NK) cell. Most information on the role of NK cells in controlling infection has been learned from mouse studies using a murine homologue (MCMV) of human CMV. Mice depleted of NK cells are highly susceptible to MCMV infection, and adoptive transfer of NK cells to NK-deficient mice can confer protection against MCMV infection. Less is known about the role of NK cells in humans, but studies have demonstrated an increased number of NK cells in primary and reactivated CMV infections in transplant patients, and that increased numbers of NK cells are correlated with the ability to recover from CMV infection.

The adaptive immune response plays a large role in combating CMV dissemination, replication and reactivation throughout life. This response largely involves effector T cells, both cluster of differentiation (CD)4+ and CD8+ cells recognize CMV antigens presented on major histocompatibility complex (MHC)-II and MHC-I molecules, respectively. Recognition by effector CD4+ T cells results in the promotion of humoral (antibody) responses, which may help to limit viral dissemination, and CD8+ T cells can directly target and kill CMV-infected cells.
CD8+ T cells recognize and respond to CMV proteins expressed in lytic and latent stages of infection, and continue to be generated throughout life. Indeed, it has been estimated that by late adulthood approximately 9 – 10% of all circulating CD4+ and CD8+ T cells recognize CMV.

Despite the enormous immune response mounted by healthy individuals, CMV continues to replicate and be shed into saliva and urine for extended periods of time (months to years) after a primary infection in young children. Although the immune response launched against CMV limits viral replication and dissemination, usually at least to the point of preventing overt disease, the virus is never eliminated and establishes latency in cells of the myeloid lineage. CMV accomplishes this through a myriad of gene products capable of manipulating the immune system. CMV produces a viral-derived IL-10 that promotes an anti-inflammatory environment, as well as an HLA-E homolog that suppresses NK cell responses by binding to the inhibitory cell surface molecule NKG2A. CMV also has mechanisms to inhibit apoptosis in infected cells, and decrease phagocytosis of infected cells by antigen-presenting cells. Chemokine homologues have been identified that appear to induce chemotaxis of neutrophils and migration of peripheral blood mononuclear cells to the anatomical location of CMV-infected cells, allowing these recruited cells to become infected aiding in viral dissemination. Some CMV gene products affect the ability of cells to express the surface markers MHC-I and -II used by effector T cells to recognize and respond to CMV infection. Through direct or indirect methods, CMV successfully manipulates the immune environment to prevent its detection and elimination. The full repertoire of immune manipulation strategies exploited by CMV is still under investigation, and ultimately, we do not know the extent of the impact a life-long CMV infection has on the human immune system.
1.3. The systemic influence of CMV

There is a vast collection of microbes living on and in the human body, collectively referred to as the “microbiota”, or the “microbiome” when discussing the genomic composition. The bacterial species of the microbiota are most commonly investigated, particularly those living within the intestinal tract. However, more recently, viruses have also been added to the discussion of the microbiome and are collectively referred to as the “virome”\(^{34-36}\). Herpesviruses, including CMV, are of particular interest to the study of the virome due to their prevalence, the life-long nature of infection, wide dispersal throughout the body, and immunomodulatory capabilities. It has been speculated that viruses such as CMV may actually contribute to the healthy regulation of the immune system in a beneficial way\(^{34}\). This is supported by studies that found herpesviruses to have the ability to affect host gene expression in various organs, usually genes in relation to cytokine (interferon) expression\(^{3,34,37}\). In mouse studies, infection with herpesviruses increases resistance to certain bacterial infections and tumor grafts, also suggesting a potential benefit of these life-long viral infections\(^{34,38,39}\). This idea of CMV acting as part of the virome suggests there may be a fine balance between the potentially beneficial and pathogenic nature of this virus.

Another study demonstrated how CMV can have a systemic impact on the cellular composition and responses of the immune system. Brodin et al. found that monozygotic twins that were discordant for CMV had greatly reduced correlations in immune measurements when compared to monozygotic twins that were both CMV uninfected\(^{40}\). They found significant impacts on the correlations between twins in frequency of CD8+ T cells as well as IL-10 and IL-6 cytokine expression, and overall CMV influenced 119 of the 204 immunological parameters that were
measured\textsuperscript{40}. This study illustrates the pervasive impact of CMV on the composition and function of the immune system in healthy individuals\textsuperscript{40}.

Recently, one group looked at the impact of CMV infection on the composition of the gut microbiome in a non-human primate model, and the impact subclinical CMV infection could have on responses to environmental stimuli and vaccines\textsuperscript{41}. They compared specific-pathogen-free (SPF) animals that did not have CMV to non-SPF animals that naturally have subclinical CMV infections. When these animals were housed in an outdoor environment, potentially exposed to many environmental immunological insults/influences, they found significant differences in the microbial communities in the intestines of CMV-infected versus uninfected animals\textsuperscript{41}. In particular, 12 bacterial taxa within the Firmicutes phylum were statistically significantly different between the two groups. Interestingly, they also found higher frequencies of circulating lymphocytes in CMV-infected animals\textsuperscript{41}. Ultimately, this group concluded that subclinical CMV infection may contribute to the heterogeneity of immune responses to environmental exposures and vaccines\textsuperscript{41}.

Although the above examples demonstrate potential positive impacts of CMV infection, it is still an opportunistic pathogen with the capability of causing systemic deleterious effects. CMV has been found in the gut of inflammatory bowel disease (IBD) patients and is suspected to be involved in exacerbating disease in ulcerative colitis (UC) patients. It is estimated that between 10 and 36\% of UC patients have CMV involvement, shown through presence of the virus in the colon, where it normally cannot be detected in healthy tissue\textsuperscript{42,43}. However, the presence of virus alone is not enough to determine causality, as it is known that CMV reactivation and shedding is frequently asymptomatic, even in highly immunosuppressed individuals\textsuperscript{44}. Furthermore, reactivation of CMV can be induced by inflammation as well as cause inflammation. Some small
studies have found that CMV exacerbates symptoms and complicates therapy, while others suggest that it could be an “innocent-bystander” of severe UC disease\textsuperscript{42,43,45–48}. Although these studies provide evidence of CMV involvement in the intestinal tract, as of yet, the impact of CMV on the composition of the bacterial intestinal microbiome in healthy humans, or how it may impact the development of a healthy microbiome in infants, has not been investigated.

1.4 The gut microbiome and immunological development

The impact of a healthy microbiome during the first months of life is currently under investigation, and it has been postulated that there is a critical time period, a “window of opportunity”, for the establishment of a healthy microbiome\textsuperscript{49–51}. At birth the diversity of the microbiota is low and continues to increase and change until a microbiome resembling that of an adult is developed by 2 – 5 years of age\textsuperscript{51}. During the first year of life, the gut is systematically colonized with bacteria from the phyla Proteobacteria, Actinobacteria, Firmicutes and Bacteroides\textsuperscript{49,51,52}. These early colonizers help to protect against invasive pathogens and are responsible for the initial education of the evolving immune system\textsuperscript{49,52–54}. The neonatal immune system has been described as “immature” or incompletely developed in comparison to that of an adult. Indeed, their pro-inflammatory responses are dampened as evidenced by low production of pro-inflammatory cytokines and reduced numbers of antigen-presenting cells\textsuperscript{52}. The permeability of the intestinal wall is also increased in neonates\textsuperscript{52}. These differences in the neonatal immune system promote a tolerogenic environment that favours establishment of commensal microbes\textsuperscript{49}.

The immune system and the intestinal microbiota develop concurrently in the first year of life, and therefore are hypothesized to largely influence one another. To determine the impact of the gut microbiome on healthy immune development, many studies have used germ-free (GF) mice.
GF mice have several marked differences when compared to normally colonized mice which include decreased frequencies of CD4+ and CD8+ T cell subsets, a Th2-skewed immune response, and increased immunoglobulin E (IgE) associated with increased IL-4\(^50\). Many of these perturbations can be reversed following introduction of normal flora, though the timing of colonization is important. For example, the increased IgE and IL-4 production in GF mice can only be normalized if a standard microbiota is introduced by 4 weeks of age\(^50\). These studies support the importance of the concomitant development of the gut microbiota and the immune system, and the education that the microbiota provides during a critical window of opportunity for healthy development early in life\(^49-51\). Dysbiosis, or an abnormal microbiome, has been associated with many diseases that are immune-mediated, such as inflammatory bowel disease, asthma, allergies, type 1 diabetes, and others\(^55\).

1.5 Atopic (allergic) diseases and early-life intestinal microbiome dysbiosis

Atopic diseases encompass the conditions asthma and eczema, also referred to as atopic dermatitis, as well as allergic rhinoconjunctivitis and food allergies. These diseases are all characterized by a predisposition to generate IgE responses to common, innocuous environmental stimuli\(^56,57\). Both asthma and eczema are clinically diagnosed by the presentation of recurrent symptoms. Asthma affects 5 – 16% of people worldwide varying widely by country and is defined as recurring episodes of wheezing and coughing\(^56,58,59\). AD is characterized by itchy, dry, chafed patches of skin anywhere on the body and affects approximately 20% of children and 10% of adults in western countries\(^60\). Understanding of these diseases is complicated by the many different risk factors that have been identified, including family background and genetic links, environmental triggers (caesarean section, use of acetaminophen, tobacco smoke, respiratory viral infections, air pollution and diet), and importantly dysregulation
of the innate and adaptive immune systems\textsuperscript{56,60,61}. Allergic diseases have traditionally been thought of as having a Th2-skewed immune response, and although it is an oversimplification, this paradigm is still a common characteristic of most presentations of allergic diseases. Given the interplay between immunologic development and the microbiome in early life, the gut microbiome has become an area of great interest in the development and prevention of allergic diseases.

Investigators have recently been making progress in characterizing the intestinal microbiome using 16S rRNA sequencing. This technique allows for characterization of many species of bacteria in comparison to methods such as species-specific PCR or bacterial culture. By revealing the composition of most of the bacterial species in the gut microbiome and comparing the relative abundances overtime, associations between small temporal changes in the microbiome of children that go on to develop atopic diseases have been described. Lower overall microbial diversity and reduced abundance of particular bacterial species early in life (1 – 3 months of age) have been associated with the later development of atopy and asthma. One study in 2014 found that children who went on to be diagnosed with asthma by 7 years of age had lower overall microbial diversity in the first month of life, however no differences were found in any particular phyla or genera of bacteria\textsuperscript{62}. Soon after, another study was able to identify specific genera that had reduced abundance at 3-months of age in children that had positive skin prick tests and clinically relevant wheezing at 1-year of age\textsuperscript{63}. Reduced abundance of \textit{Faecalibacterium, Lachnospira, Rothia, Veillonella} and \textit{Peptostreptococcus} were described that interestingly disappeared by 1-year of age\textsuperscript{63}. Since then, the possibility of a gut microbiome biomarker as an indicator for the development of asthma by pre-school age has been described\textsuperscript{64}. A reduction in the genera \textit{Lachnospira} at 3-months of age was significantly associated with a
diagnosis of asthma by 4 years of age \(^{64}\). Another group found that children most at risk for the development of parental-reported, doctor-diagnosed asthma at 4 years of age had a distinct microbiome profile that was characteristically depleted of specific species including *Bifidobacteria, Lactobacillus, Faecalibacterium* and *Akkermanis* \(^{65}\). While yet another group found an increased risk of asthma at 5 years of age with increased abundance of *Veillonella*, and decreased abundance of *Roseburia, Alistipes* and *Flavonifractro* \(^{66}\). These studies provide evidence of a “critical window” in early life, where microbial alterations in the gut are the most impactful as a risk factor for atopic diseases, though the exact timing and particular species that play a role are still inconclusive.

1.6 Linking CMV infection and allergic diseases through the hygiene hypothesis

In 1989, Dr. David Strachan published the observation that the number of children in a household, particularly older children, was inversely related to the development of eczema in the first year of life and later development of hay fever in 11 to 23 year-olds \(^{67}\). From this observation he hypothesized that allergic diseases could be prevented by the acquisition of infections early in childhood, and that younger siblings would contract these infections from unhygienic contact with older siblings \(^{67}\). This led to the idea that the increase in prevalence of allergic disease might be the “cost” of decreased exposure to infectious diseases, which was labelled the “hygiene hypothesis”. This hypothesis fits with the evidence that allergic diseases have increased in prevalence dramatically in the last several decades, particularly in industrialized countries where exposure to infectious disease has most dramatically decreased \(^{68-70}\). However, subsequent studies have shown that personal and household hygiene practices do not associate with allergic diseases, and that childhood infections do not protect against allergic disorders \(^{68,71,72}\).
Over the years the “hygiene hypothesis” was expanded and modified into the “old friends hypothesis”\textsuperscript{73}. This hypothesis specifies that it isn’t exposure to infectious diseases such as flus, colds, measles, etc., but rather to natural environmental organisms and commensal microbiotas that are necessary to prevent allergic diseases\textsuperscript{55,68,73}. This exposure is hypothesized to help guide the development of the microbiome and educate the immune system to prevent it from overreacting to innocuous environmental or self-stimuli\textsuperscript{55}. Thus, it is now thought that exposure to particular species (bacterial, viral, fugal, helminth) at particular times in infancy, guide the development of a healthy microbiome which then educates the immune system towards proper balance and tolerance. Not surprisingly, the most important time of exposure is likely in the first few months of life\textsuperscript{68}.

CMV, as with all herpesviruses, is part of the human virome, and as such could be one of the “old friends” necessary for proper microbiome and immune development. CMV infection in infants has been associated with developing a more robust immune response to certain stimuli (measles and staphylococcal enterotoxin B)\textsuperscript{3}. As asthma prevalence rises in developed countries, the age of primary infection with herpesviruses has increased. One study conducted in Sweden found that infection with another herpesvirus (Epstein-Barr virus; EBV) early in life was associated with lower risk of persistent IgE-sensitization, but if infected after 2 years of age it increased risk\textsuperscript{74}. A related study found that CMV infection after EBV infection further decreased risk of IgE sensitization, but CMV infection alone was not associated with decreased risk\textsuperscript{75}. Therefore, it is possible that CMV infections at the right time of life could contribute to the education of the immune system, either through immune modulation, impact on the establishment of the early life microbiome, or both, and ultimately protect against the development of atopic diseases.
1.7. Hypothesis, objectives and approach

Considering that Canada is an industrialized country with high prevalence of atopic disease and later incidence of CMV infections, a Canadian cohort is ideal for exploring connections between early-life CMV infections and its impact on the microbiome and atopic disease outcomes. The Canadian Healthy Infant Longitudinal Development (CHILD) Study is a longitudinal, prospective, population-based, birth cohort study that set out to investigate the interaction between environmental exposures and genetic predispositions in the development of atopic diseases (childstudy.ca). Since initiation of the project, the goals have broadened to include the study of other chronic non-communicable diseases and ultimately becoming a platform for the investigation of the Developmental Origins of Health and Disease\textsuperscript{76}. From 2008 – 2012, in 4 cities across Canada, Vancouver, Edmonton, Winnipeg and Toronto, the CHILD Study recruited 3455 pregnant women. The children born into the CHILD Study were followed from birth to 5 years of age. At several time points throughout the 5 years (described in detail in Chapter 2), biological samples, medical assessment information and questionnaire data were provided.

By making use of this incredible cohort of healthy Canadian infants, the connections between early-life CMV infection and the development of the intestinal microbiome were investigated for the first time, along with associations between early-life CMV infection and the development of allergic diseases in the first years of life. It was hypothesized that early-life infection with CMV occurs in a minority of healthy Canadian infants, and is associated with acquisition of intestinal microbiome constituents that help to prevent the development of allergic diseases. The specific aims of this study were to:

1. Describe the prevalence and risk factors of CMV infections in the first year of life of healthy Canadians.
2. Investigate the impact of early-life CMV infections on the intestinal microbiomes of healthy infants.

3. Determine if any associations exist between early-life CMV infection and the development of allergic diseases in the first year of life.
Chapter 2. CMV Prevalence and risk factors
2.1. Introduction

2.1.1. Prevalence of CMV in Canada

Few studies have investigated the prevalence of CMV infection in infants, especially in Canada. One study carried out in Halifax, Nova Scotia from 1967-1968, surveyed 550 samples, 100 of which were from cord blood\textsuperscript{18}. Testing for CMV using complement-fixing antibodies in the blood, these authors found a CMV prevalence of 12\% in infants 6 – 12 months of age, and 16\% in children 1-2 years of age\textsuperscript{18}. In 1996, a group investigated the prevalence of CMV in 38 childcare centers in urban Toronto\textsuperscript{19}. By detecting CMV in the urine of 471 children by (shell vial assay), it was found that 16.8\% of children between 4 months and 3.6 years of age were shedding CMV in the urine\textsuperscript{19}.

2.1.2. Detecting CMV

The historical “gold standard” assay for detecting CMV is viral culture, which was adapted in recent decades to a shell vial culture assay that uses antibodies to the CMV immediate early 1 protein making it faster\textsuperscript{77}. However, this is still a cumbersome assay that is not practical for high-throughput determination of CMV status. As such, the use of qPCR in place of shell vial culture as the standard has been widely adopted\textsuperscript{78,79}. Two studies have compared the sensitivity of the shell vial culture assay to various qPCR assays for CMV detection in urine and found that qPCR was 5 – 11\% more sensitive\textsuperscript{78,79}. Thus, in addition to convenience and speed, the superior sensitivity of qPCR has led to its replacement of culture for determination of CMV status\textsuperscript{78,79}.

Serology, testing for anti-CMV antibody in blood serum, is a useful standard practice for determining if a person has ever been infected with CMV; however, the accuracy of serologic testing for CMV infection in infancy is complicated by transplacental transfer of maternal
antibodies. This was demonstrated by Embil et al. when CMV prevalence determined in this way dropped from 28% in infants 0 – 6 months of age to 12% from 6 – 12 months. Testing for CMV in the blood by viral/DNA isolation methods is highly specific for infection, but viremia may not be present during infection or reactivation. Instead, looking for CMV shed into body fluids, such as saliva and urine, is the most sensitive and specific way to test for CMV infection in infants. Adults will only periodically shed CMV in the saliva, making serology most useful for diagnosing infection. Infants, however, shed CMV into their urine for many months to years after a primary infection. Therefore, in infants, testing for virus in the urine by qPCR has become the gold standard for diagnosis of congenital infection. Here, this extended duration of shedding in the urine is used to test Canadian infants for CMV infection at 3 months and 1 year of age.

### 2.1.3. CMV risk factors of infection

Age is the strongest risk factor for CMV infection; CMV prevalence increases steadily with age until approximately 60% of most populations are infected by 50 years of age in highly developed countries. Other common risk factors for CMV infection are race/ethnicity, socioeconomic status, number of other children in the home, and child daycare attendance. Few studies have looked at the risk factors for CMV infection of Canadians. One study conducted in the city of Toronto, Ontario found the risk factors for adult health care workers to be country of birth outside of Canada, presence of children at home <5 years of age and increased household size to be associated with seropositivity. Although children were included in the study, the risk factors for children were not examined. Another study carried out in Montreal, Quebec investigated the risk factors for CMV infection among female day care educators. For this group, risk factors included age, low- or middle-income country of birth and having 2 or more children of their
own\textsuperscript{21}. A pediatric cohort outside of Canada, in Bradford, UK, examined the risk factors for infection of children ages 12 – 24 months\textsuperscript{20}. For these children, race (specifically, Pakistani- vs. white-British) was a risk factor for infection and breastfeeding was associated with earlier infection\textsuperscript{20}. The risk of transmitting CMV through breast milk has been thoroughly examined in several groups\textsuperscript{14,15,25–27}. CMV is shed by healthy immunocompetent seropositive women in breast milk, and mother-to-child transmission of CMV occurs more frequently in breast-fed than bottle-fed children\textsuperscript{25–27}. To my knowledge this is the first pediatric Canadian cohort to assess risk factors of CMV infection.

2.1.3. Approach

The urine of CHILD Study participants was tested by qPCR to determine the prevalence of CMV infection in healthy Canadian infants. These CMV infection status data, combined with extensive information collected through questionnaires (described below), were used to assess the risk factors for CMV infection by 1 year of age.

2.2. Methods

2.1.1. The Canadian Healthy Infant Longitudinal Development Study

Recruitment to the study targeted pregnant women from the general population\textsuperscript{83}. Pregnant women were recruited during their second trimester when attending health centres for ultrasounds, at kiosks at community events, malls and baby shows, and through social media advertising\textsuperscript{83,84}. All interested mothers that were greater than 18 years of age, living within 50 kilometers of a participating hospital and planning to deliver there, willing to donate cord blood, able to read, write, and speak English were initially included\textsuperscript{83}. Any infants that were premature (less than 35 weeks gestational age), conceived by \textit{in vitro} fertilization, born twins (or greater), had any major congenital abnormalities or respiratory distress syndrome were excluded from the
study. Additionally, any children that were anticipated to spend less than 80% of their time at home or were planning on moving away from the recruitment center in the first year of life were excluded.

From all participating families, questionnaires describing environmental and time-activity details were collected 11 times throughout the study: at 18 and 36 weeks of gestation, 3, 6, 12, 18, 24 and 30 months, and 3, 4, and 5 years. Biological samples, including blood, feces and urine were also collected; for the purposes of this study, the biological samples of interest are the child’s feces and urine at 3 months and 1 year of age.

2.2.2. Sample selection

A sub-cohort from the CHILD study was created that represented all children who: 1) already had fecal 16S rRNA sequence data available through Dr. Stuart Turvey’s lab at BC Children’s Hospital Research Institute and through the multi-site research group Symbiota (Synergy in Microbiota Research; http://allergen-nce.ca/research/symbiota/), and 2) had stored urine samples collected at 3 and 12 months of age. In total, 1151 children met these criteria for determination of CMV status.

2.2.3. Determination of CMV Status

2.2.3.1. DNA extraction

Qiagen’s DNeasy® Blood and Tissue kits were used to extract DNA from urine samples. Before extraction, 198 ul of urine was aliquoted into a 1.5 ml microcentrifuge tube to which 2 ul of an internal control plasmid was added. The internal control plasmid was an unmodified pUC19 double-stranded closed circular DNA plasmid isolated from E. coli ordered from ThermoScientific. The plasmid has a total of 2686 base pairs with a molecular weight of 1.74 X 10^6 DA from which the total copy number was calculated using the provided DNA concentration.
of 0.5ug/ul (GenBank accession number L09137). A total of 5 X 10^5 copies of pUC19 plasmid was spiked into each sample before extraction. Once the sample was prepared, the kit manufacturer’s instructions were followed without modification from step #2 on page 26 of the DNeasy® Blood and Tissue Handbook. Briefly, 20 ul of proteinase K and 200 ul of Buffer AL, both provided with the kit, were added and thoroughly vortexed for 30 seconds. The mixture was then incubated at 56°C for 10 minutes in a heating block. After incubation, 200 ul of anhydrous ethanol was added, and after vortexing the whole mixture was transferred to a DNeasy® Mini spin column. Columns were spun for 1 minute at 6000 X g, then 500 ul of wash Buffer AW1 was added and spun again. A second wash, Buffer AW2, was added and spun at 20,000 X g for 3 minutes. Finally, DNA was eluted in 100 ul of Buffer AE, and the eluate was spun through the column a second time to try to maximize yield.

2.2.3.2. CMV detection

Primers and probes sequences, for both CMV amplification and pUC19 internal control amplification were provided by British Columbia Provincial Health Services Authority (PHSA) microbiology laboratory. This in-house assay amplifies a unique section of CMV UL83 gene, which encodes the tegument protein pp65, the most abundant of the tegument particles. CMV protein pp65 inhibits innate immune responses and simultaneously activates immediate-early gene expression to promote viral growth. Importantly, although CMV can replicate in vitro without UL83, this is a necessary protein for successful CMV infection and thus is never lost from the genomes of natural isolates. A FAM-labeled probe that specifically binds within the 75-base pair (bp) region amplified by the UL83 primers allows quantification of viral genome copy number. UL83 primer and probe sequences are as follows: CMV-UL83-F: 5'-
GCGCACGAGCTGTTTG -3’, CMV-UL83-R: 5’- TGGTCACCTATCACCTGCATCT -3’, CMV-UL83-P: 5’-/56-FAM/TCCATGGAG /ZEN/AACACGCGCGC/3IABkFQ/ - 3’.

2.2.3.3. Internal control
To ensure that each qPCR reaction is valid, and that inhibitors were not interfering with amplification, the pUC19 plasmid described above was used as an internal control and amplified along with CMV. The probe fluorophore used was HEX, to allow for duplex qPCR of the virus and pUC19 plasmid within each sample. The pUC19 primer and probes sequences are as follows: IC674R: 5’- GTGATGACGGTGAAAACCTCTGA -3’, IC600F: 5’- TCCCGGCATCCGCTTA -3’, Probe: 5’-/5HEX/ACAAGCTGT/ZEN/GACCGTCTCCGGGAG/3IABkFQ/ -3’. All primers and probes were ordered from Integrated DNA Technologies (IDT).

2.2.3.4. CMV quantitative control
A 415 bp segment of the UL83 gene, containing the same region as the CMV-UL83 primers and probe described above, was inserted into a pUCIDT-AMP plasmid (also ordered from IDT). The concentration (number of copies/ul) of this CMV-minigene was calculated and used to generate a quantitative standard curve from which the CMV copy numbers in CHILD study urine sample DNA were calculated. To prepare the CMV mini-gene for qPCR, 100 ul (2.5 x 10^7 copies) was combined with100 ul of pUC19 plasmid (2.5 x 10^7 copies) and extracted using the same DNeasy® Blood and Tissue kit protocol described above. This resulted in a control mixture used to generate a standard curve in each qPCR plate.

2.2.3.5. qPCR reaction set-up
Primers and probes were resuspended in IDTE buffer and made into concentrated mixes so that only 1 ul of CMV-UL83 primer/probe mixture and 1 ul of pUC19 primer/probe mixture was
needed per qPCR reaction. The final concentration of primers per reaction was 900 nM, and 250 nM of the probes for each CMV-UL83 and pUC19 amplification. These concentrations fell within the recommended values provided by the IDT PrimeTime® Gene Expression Master Mix protocol. PrimeTime® Gene Expression Master Mix is a 2X mix containing antibody-mediated, hot-start DNA polymerase, dNTPs, MgCl₂, enhancers, and stabilizers specifically designed for probe-based qPCR. To minimize reagent use, a 10 ul reaction was used containing: 5 ul of IDT-PT 2X Master Mix, 1 ul of CMV-UL83 primer/probe mix, 1 ul pUC19 primer/probe mix and 4 ul of extracted DNA from the CHILD Study urine samples or extracted control DNA. The maximum volume of DNA possible (4 ul) was chosen to maximize detection of CMV and provide the greatest assay sensitivity.

Each 96-well plate included 40 samples in duplicate, a standard curve using the above described extracted CMV mini-gene plasmid and pUC19 plasmid in 10-fold dilutions from 10⁵ to 10⁰, 2 no-template-control negative control reactions, and 2 pUC19 only reactions using the same pUC19 dilution as added to each sample before extraction.

All qPCR reactions were run in a BIO-RAD C1000 Touch Thermal Cycler Chassis with a CFX96 Optical Reaction Module. Thermal cycling was programmed as recommended by the PrimeTime® Gene Expression Master Mix Protocol as follows: 1 cycle at 95°C for 3 minutes for polymerase activation, then 45 cycles of 95°C for 15 seconds followed by 60°C for 1 minute.

**2.2.3.6. Determination of CMV status from qPCR**

Quantification of CMV and pUC19 was calculated automatically through CFX Manager 3.1 software provided with the machine. Each plate had a standard curve allowing for standard quantification across multiple plate runs. Any samples that had quantifiable amplification of CMV in both replicates were included in the analysis. Samples that had low pUC19
amplification were removed from analysis due to concerns about the presence of inhibitors and false negative CMV results. Any samples that did not have adequate volume (198 ul) of urine for extraction were also excluded for concern of false negative results.

2.2.4. Determination of risk factors for CMV infection

A list of potential risk factors for CMV infection by 1 year of age was determined a priori from the literature. The list of potential risk factors available from CHILD Study questionnaires for investigation were: sex, study site, breast feeding (duration in months), paternal and maternal ethnicity, father’s and mother’s education level and duration, total yearly household income, number of adults in the home, number of children in the home, if the infant had siblings, how many days per week the infant spent outside of the home (e.g., daycare) and, if so, with how many children. Each of these variables were tested individually, first by logistic regression to determine if a statistically significant association (p < 0.05) between the variable and CMV infection was present. If significant in this unadjusted model, the variable was then included in a multivariable model to adjust for correlations between variables. All logistic regression models were run in R Statistical Programming language using the glm function included in the base R package. To determine if a bivariate (individual independent variable) logistic regression model was significant an anova was run on each individual model using the anova function included in the base R package. This test determines if the variable significantly contributes to the variation in the model in comparison to a model that includes only the intercept. Odds ratios were calculated by taking the exponential of the coefficients for each variable.

2.3. Results

2.3.1. Prevalence of CMV in healthy Canadian children
In total, 1151 children were tested for CMV shedding in the urine at 1 year of age, and after exclusion of samples with low volume, low pUC19 amplification (potential inhibitors), or absence of amplification in at least one of two replicates, 1083 qPCR children were remaining for analysis. The demographics of these children are presented in Table 2.1. Children from all 4 study sites were included, though not evenly; 37%, 34%, 17% and 11% were from Winnipeg, Vancouver, Edmonton and Toronto, respectively. Male children made up just over half this sub-cohort (55%), and 96% of infants were breastfed with a mean duration of 7.25 months (Table 2.1).

By 1 year of age, 19% (206) of these children had been infected with CMV. The 206 children found to be CMV positive at 1 year of age were also tested for CMV shedding in the urine at 3 months to estimate the time of infection: before 3 months, or between 3 and 12 months. By 3 months of age, 67% (139) were already CMV infected.

Table 2.1. Demographics of CHILD Study sub-cohort

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>CMV Status at 1 year of age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Negative</td>
</tr>
<tr>
<td>Evaluated number</td>
<td>877 (81)*</td>
</tr>
<tr>
<td>Location</td>
<td></td>
</tr>
<tr>
<td>Vancouver</td>
<td>289</td>
</tr>
<tr>
<td>Edmonton</td>
<td>154</td>
</tr>
<tr>
<td>Winnipeg</td>
<td>329</td>
</tr>
<tr>
<td>Toronto</td>
<td>105</td>
</tr>
<tr>
<td>Male sex</td>
<td>427 (55)</td>
</tr>
<tr>
<td>Breastfeeding</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>37 (4)</td>
</tr>
<tr>
<td>Duration (median in months)</td>
<td>10</td>
</tr>
</tbody>
</table>

* Numbers in brackets indicate percentage
2.3.2. Risk factors for CMV infection by 1 year of age

Potential predictors of CMV infection by 1 year of age are listed in Table 2.2. All potential risk factors were first evaluated individually through binomial logistic regression models (Table 2.2. Unadjusted columns). Risk factors that were statistically significant (p < 0.05) predictors in these unadjusted models were then used to generate a multivariable logistic regression model which adjusts for possible collinearity between variables that could confound the results (Table 2.2., Adjusted columns). Variables included in the multivariable/adjusted model were: breastfeeding duration, paternal ethnicity, maternal ethnicity, total household income in a year and number of adults residing in the home. After adjustment, only paternal and maternal ethnicity, and breastfeeding duration remained statistically significant predictors of CMV infection by 1 year of age. Specifically, fathers that identified as Black or Middle Eastern led to increased risk of CMV infection (odds ratio (OR) [95% confidence interval (CI)]: 6.112 [1.535-24.026] and 10.318 [1.412-67.692], respectively), as well as mothers that identified as South East Asian (4.608 [1.671-12.642]) (Table 2.2). For every month of breastfeeding there was an increased risk of 1.090 [1.034-1.154] of becoming infected with CMV by age 1.

Table 2.2. Risk factors of CMV infection by 1 year of age

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Unadjusted</th>
<th></th>
<th></th>
<th>Adjusted</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR</td>
<td>95% CI</td>
<td>OR</td>
<td>95% CI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breastfeeding duration</td>
<td>1.085</td>
<td>1.040-1.134</td>
<td>1.090</td>
<td>1.034-1.154</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Father’s ethnicity (reference category = Caucasian)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First Nations</td>
<td>1.367</td>
<td>0.606-2.783</td>
<td>1.395</td>
<td>0.485-3.474</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South East Asian</td>
<td>4.596</td>
<td>2.371-8.810</td>
<td>1.314</td>
<td>0.419-3.713</td>
<td></td>
<td></td>
</tr>
<tr>
<td>East Asian</td>
<td>2.313</td>
<td>1.171-4.351</td>
<td>1.860</td>
<td>0.627-5.324</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Asian</td>
<td>2.943</td>
<td>1.338-6.143</td>
<td>1.477</td>
<td>0.341-5.485</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle Eastern</td>
<td>4.013</td>
<td>1.171-12.772</td>
<td>10.318</td>
<td>1.412-87.692</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictors</td>
<td>Unadjusted</td>
<td></td>
<td></td>
<td>Adjusted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>------------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td>OR</td>
<td>95% CI</td>
<td>OR</td>
<td>95% CI</td>
<td>OR</td>
<td>95% CI</td>
</tr>
<tr>
<td>Hispanic</td>
<td>3.511</td>
<td>1.046-1.070</td>
<td>2.401</td>
<td>0.461-9.915</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>1.124</td>
<td>0.058-7.045</td>
<td>2.651</td>
<td>0.119-23.467</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiracial</td>
<td>1.362</td>
<td>0.574-2.879</td>
<td>0.792</td>
<td>0.245-2.107</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mother’s ethnicity (reference category = Caucasian)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First Nations</td>
<td>1.242</td>
<td>0.554-2.508</td>
<td>0.739</td>
<td>0.197-2.163</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South East Asian</td>
<td>7.985</td>
<td>4.192-15.533</td>
<td>4.608</td>
<td>1.671-12.642</td>
<td></td>
<td></td>
</tr>
<tr>
<td>East Asian</td>
<td>2.218</td>
<td>1.253-3.798</td>
<td>1.459</td>
<td>0.549-3.578</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Asian</td>
<td>4.848</td>
<td>2.155-10.751</td>
<td>2.646</td>
<td>0.652-1.086</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>6.787</td>
<td>2.014-23.878</td>
<td>1.393</td>
<td>0.207-8.635</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle Eastern</td>
<td>5.656</td>
<td>1.036-10.866</td>
<td>0.937</td>
<td>0.065-10.747</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>1.305</td>
<td>0.296-4.121</td>
<td>0.795</td>
<td>0.098-3.764</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiracial</td>
<td>0.808</td>
<td>0.273-1.927</td>
<td>0.593</td>
<td>0.164-3.764</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total yearly income</td>
<td>0.888</td>
<td>0.821-0.962</td>
<td>0.914</td>
<td>0.830-1.009</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of adults in the home</td>
<td>1.368</td>
<td>1.110-1.687</td>
<td>1.120</td>
<td>0.856-1.478</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of children in home</td>
<td>0.990</td>
<td>0.815-1.193</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any older siblings</td>
<td>0.887</td>
<td>0.649-1.207</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regular visits outside home</td>
<td>0.887</td>
<td>0.547-1.390</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Days/week visit outside home</td>
<td>1.085</td>
<td>0.832-1.405</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of children at visits</td>
<td>1.291</td>
<td>0.867-1.877</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birth sex (reference = female)</td>
<td>0.956</td>
<td>0.689-1.329</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study center city (reference city = Edmonton)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toronto</td>
<td>0.757</td>
<td>0.386-1.435</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vancouver</td>
<td>1.410</td>
<td>0.900-2.252</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Winnipeg</td>
<td>1.163</td>
<td>0.741-1.860</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mother’s education level (reference category = Some high school)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completed high school</td>
<td>4.125</td>
<td>0.685-76.502</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some college</td>
<td>1.224</td>
<td>0.175-24.610</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completed college</td>
<td>2.151</td>
<td>0.388-40.314</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some university</td>
<td>3.517</td>
<td>0.622-66.351</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completed university</td>
<td>2.499</td>
<td>0.481-45.861</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master’s degree</td>
<td>3.886</td>
<td>0.725-72.078</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PhD</td>
<td>1.818</td>
<td>0.257-36.714</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mother’s education years</td>
<td>1.011</td>
<td>0.956-1.070</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### 2.4. Discussion

Using the specific and sensitive method of qPCR to detect shedding of CMV in the urine, we found that 19% of healthy Canadian children in this sub-cohort were CMV-infected by 1 year of age (Table 2.1.). There have only been two previous studies that estimated CMV prevalence of children in Canada. One of these studies was carried out in Nova Scotia in 1969 where 12% of children were CMV seropositive between 6 – 12 months of age\(^8\). A second study, published in 1996, was carried out in urban Toronto where 471 children between the ages of approximately 4 months to 4 years were tested for CMV by shell viral culture of the urine\(^9\). The overall CMV prevalence in these children was 16.8%, with the highest rate of shedding in the urine in the 1.25 -1.75 years of age group at 21%\(^9\). The range of CMV prevalence in young children in Canada based on these two studies was 12 – 21%, therefore the 19% found in this study aligns with the higher end of previous literature. The study of children in Toronto targeted children in daycare centers which is a known risk factor for infection, and therefore it was expected that the
prevalence of CMV infection in that study would be higher than in this study that did not target a risk group. It is difficult to compare CMV prevalence across studies, however, as there is no unified measurement for determining CMV prevalence. For example, qPCR has been found to be more sensitive than shell vial culture by 5-10%, so it is possible that the Toronto study would have had a higher prevalence if using the same technique for detection as employed in this study \(^7^8,^7^9\).

There have been many studies across the globe that have investigated the many potential risk factors for CMV infection, though fewer studies have targeted risk factors for infection in healthy young children. Of the previously documented risk factors, breastfeeding, paternal and maternal ethnicity, socioeconomic status (measured via household income and parents’ education), number of adults and children in the home, older siblings and daycare attendance (measured via visits outside the home) were available for analysis in this study (Table 2.2.). Of these, the ethnicities of the parents and breastfeeding duration were the only statistically significant risk factors for infection remaining once adjusting for confounding variables (Table 2.2.). In comparison to children whose parents identified as Caucasian, children whose fathers were Black or Middle Eastern or whose mothers were South East Asian were at higher risk of CMV infection in the first year of life (Table 2.2). Several studies have found significant differences across ethnicities with increased prevalence of CMV found in non-white ethnic groups \(^1^7,^2^0,^2^3,^2^4\). In a large literature review that compiled results from English-language publications between the years of 1966 – 2008, it was found that CMV seroprevalence worldwide was 20-30% higher in non-whites than whites \(^1^7\). A study in the Netherlands, that included participants from 6 months – 79 years of age, reported non-Western migrants to have the highest seroprevalence, in comparison to Western migrants and native Dutch participants.
who had the lowest seroprevalence\textsuperscript{24}. In a recent study, 1 to 2 year-old children in Bradford UK were tested for CMV by serology, where they found that Pakistani children were more likely to be infected with CMV at a younger age than White British Children (OR [CI]: 2.53 [1.47-4.33])\textsuperscript{20}. By 2 years of age, the prevalence of CMV in Pakistani-British children was 34%, whereas it was only 9% in white British children\textsuperscript{20}. Thus, our result of children with parents of Black, Middle Eastern or South East Asian origin in comparison to Caucasians is a reasonable finding, with the ethnicities potentially reflecting the groups of populations represented in Canada.

As previously discussed, breastfeeding is also a known risk factor for CMV infection; studies have shown that CMV reactivates in the breast of nearly all immunocompetent lactating women resulting in shedding in milk\textsuperscript{14,15,25-27}. One study found mother-to-child CMV transmission to be 65% in breast-fed children, whereas it was only 28% in bottle-fed infants\textsuperscript{27}. In this study, for each month of breastfeeding there was an increased risk of the infant contracting CMV (OR [CI]: 1.090 [1.034-1.154]; Table 2.2.) Transmission via breastmilk may also explain the high percentage (67%) of CMV infected infants that were already infected by 3 months of age.

It was expected that socioeconomic status (measured via household income and parents’ education) and daycare attendance (measured via visits outside the home) would be risk factors for infection, as reported in other studies\textsuperscript{17,20,21,23}. Potentially, a more sensitive variable for socioeconomic status could be generated that combines the effect of the different factors contributing to socioeconomic status into one variable. Visits outside the home were not exclusive to daycare as parents were asked if the child had regular visits outside the home, how many days per week these visits occurred, and the number of other children present during these visits. This variable would be inclusive of daycare but could also include other social situations.
in which a small number of the same people were being visited (e.g., grandparents), and therefore may not hold the same risk of CMV infection as expected when collecting time spent in daycare as reported in previous studies.

Very few studies look at the prevalence and risk factors of postnatal CMV shedding in healthy infants, and there have been no studies in the last 19 years that have looked at the prevalence of CMV in the healthy Canadian population. As such, this is the largest study in Canada to look at CMV prevalence from multiple sites across Canada in the first year of life of healthy infants. With this, we found a 19% prevalence of CMV infection in 1 year old infants with parental ethnicity and breastfeeding duration as risk factors for infection in the first year of life.
Chapter 3. CMV and the intestinal bacterial microbiome

3.1. Introduction

CMV is an immunomodulatory virus that can infect the intestinal tract of humans. It has been characterized as a potential confounding factor for people with inflammatory bowel disease, but its role in the intestinal microbiome of healthy individuals has not been investigated. It is important to determine if CMV has an impact on the composition of the intestinal microbiome early in life, during the previously described “critical window” of opportunity where the microbiome and immune system are developing contemporaneously.

3.1.1. Characterizing the intestinal microbiome

To investigate a connection between CMV and the intestinal microbiome, 16S rRNA gene sequences were used to characterize the bacterial composition from fecal samples. The 16S rRNA gene encodes the ribosomal RNA that is a component of the small subunit of a prokaryotic ribosome that can be used to profile bacterial communities. To make use of next-generation sequencing platforms, smaller hypervariable regions within the 16S rRNA gene have been identified, such as the V4 hypervariable region that has been shown to most reliably identify bacteria compared to the full gene. From these sequences, and using the previously described CMV status from urine qPCR, differences in bacterial diversity and composition in CMV-infected versus uninfected groups were compared. One measurement used to describe the mean species diversity in a sample is the $\alpha$-diversity. This measurement addresses the question of how many different species are detected in any given sample and provides a measurement for comparison of diversity between samples or groups. A second commonly used measurement for 16S rRNA sequence analysis is the $\beta$-diversity. The $\beta$-diversity is used to determine the diversity
between (at least) two groups and addresses the questions of what bacteria are present, and
different, between those groups.

3.1.2. Approach
To investigate if early-life CMV infection impacted the composition of the intestinal microbiome
in the first year of life, the CMV status, as determined by urine CMV qPCR, was used to
compare the microbiomes of CMV-infected to -uninfected children at 3 months and 1 year of
age.

3.2. Methods
3.2.1. Samples collection and processing
The CHILD Study collected infant fecal samples at 3 months and 1 year of age by providing
Tegaderm™ liners to parents. Up to 3 days prior to an in-home visit, parents placed the liner in a
new diaper and once collected placed the liner in sealed plastic bag to store in the refrigerator. A
CHILD Study representative would then collect the stored sample at an in-home visit, and after
transportation on ice to the lab, samples were aliquoted and stored frozen at -80°C.

The Stuart Turvey lab at BC Children’s Hospital Research Institute selected samples for 16S
rRNA sequencing leading to an atopy-enriched sub-cohort. First, participants that had stool and
blood samples available at both visits (3 months and 1 year) were identified. From these
participants, those with positive skin prick tests and wheezing were requested. Second,
participants with a fecal sample from at least one visit along with positive skin prick test and
wheezing were added. Last, control samples were added by the same biological sample
availability (but with negative skin prick test and wheezing) while trying to match number of
samples by city. These samples were then sent to the company Microbiome Insights at the
University of British Columbia for DNA extraction, PCR amplification and sequencing.
Genomic DNA was extracted from 80 to 200 mg of stool using the QIAamp® DNA Stool Mini Kit. Bacterial 16S rRNA V4 hypervariable region was amplified by PCR using universal primers (V4-515f: V4-806r)\textsuperscript{89,90}. The reverse primers were uniquely barcoded for each sample to enable sequence multiplexing. Barcoded PCR amplicons were pooled and used for paired-end sequencing by Illumina® Miseq™ System.

**3.2.2. Analysis of 16S rRNA data**

**3.2.2.1. QIIME2 processing**

QIIME 2™ (qiime2.org) was used to perform the initial steps of preprocessing, denoising and quality filtering by size. Paired-end sequences were first paired, then denoised using deblur, then trimmed to 248 base pairs using visualization of the quality of paired reads and phed score for guidance. Cleaned sequences were then aligned in QIIME 2™ using Mafft and clustered into operational taxonomic units (OTUs) using FastTree. To define taxonomies, sequences were clustered at 97% similarity to reference genes provided by the Greengenes database. Low frequency reads, below 1000 reads which amount to only ~ 0.001% of total sequencing reads, were removed, along with the genus Halomonas as it is a known contaminant of the filter columns of the QIAamp® DNA Stool Mini Kit. At this point, the sample tree, taxonomy and feature table were exported from QIIME 2™ for further analysis in the R programming language version 3.5.0.

**3.2.2.2. Analysis in R**

Multiple packages were installed and loaded for analysis in R Studio: phyloseq, dplyr, ggplot2, reshape2, btools, data.table, vegan, lme4. The phyloseq package was used to first import objects from QIIME 2™, and to rarefy (ie. draw without replacement from each sample so that all
samples have the same number of total counts\(^1\) to 5000 reads for even sampling depth across samples.

The \(\alpha\)-diversity of the intestinal bacterial microbiome was investigated by comparing CMV infected vs uninfected children based on qPCR results from urine samples of each child at the same time point (i.e., 3-month fecal \(\alpha\)-diversity compared to CMV status at 3 months). The btools package was used to calculate Faith’s phylogenetic diversity (FPD) for all samples, and statistical significance of FPD between CMV status group was analyzed by Pairwise Wilcoxon Rank tests. Linear regression models were used to investigate the impact of possible confounding variables (breast feeding, birth mode and antibiotic use) on associations between CMV status and FPD.

The \(\beta\)-diversity was analyzed by calculating weighted UniFrac distances using the phyloseq package in R. These distances were then analyzed by permutational analysis of variance (PERMANOVA) using the Adonis function provided in the vegan package in R for statistical relevance between groups of interest (e.g., CMV infected vs. uninfected).

To investigate bacterial population composition at phylum, genus and species level, the relative abundance of OTUs within each sample was calculated using the phyloseq package. The phyloseq built-in function “mt” could then be used to perform multiple testing of taxa abundance according to CMV status. The “mt” function calculates wilcoxon rank sum tests, and the Benjamini and Hochberg method to control false discovery rate (fdr) was applied.
3.3 Results

3.3.1. Sample demographics

In total, Dr. Stuart Turvey’s lab provided 16S rRNA sequence data from 941 samples from infants whose CMV status had been determined (described in detail in 2.2.2.).

Of the 3-month samples, 404 were analyzed, of which 12.9% (52) were CMV-infected. For 1-year samples 429 were analyzed with 18.1% (78) CMV positive. The group of children infected at 1 year of age includes any children that were already infected by 3 months. The demographics of interest are presented in Table 3.1, which included birth sex with 56.6% (229) and 56.5% (243) of the 3-month and 1-year groups being male, respectively. Again, breastfeeding rates were high, with only 4.5 and 4.7% of children in the 3-month and 1-year groups having never been breastfed, respectively (Table 3.1). The mean duration of breastfeeding was 2.7 months in the 3-month group and 8.9 in the 1-year group (Table 3.1). Birth mode included three categories: vaginal, Caesarean section with labour (CWL) and C-section without labour (CWOL) (Table 3.1). The majority of children in both groups were delivered vaginally, 76.9% of the 3-month group and 77.9% of the 1-year group (Table 3.1).
Table 3.1. Demographics of 16S rRNA-sequenced samples evaluated at 3 months and 1 year by CMV status determined by qPCR or urine samples at the same time point as sequenced fecal samples.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>CMV Status at 3 months</th>
<th>CMV Status at 1 year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Negative</td>
<td>Positive</td>
</tr>
<tr>
<td>Evaluated</td>
<td>352</td>
<td>52</td>
</tr>
<tr>
<td>Birth sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>199</td>
<td>30</td>
</tr>
<tr>
<td>Female</td>
<td>153</td>
<td>22</td>
</tr>
<tr>
<td>Breastfeeding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>Duration (mean months)</td>
<td>2.7</td>
<td>2.9</td>
</tr>
<tr>
<td>Birth Mode</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vaginal</td>
<td>269</td>
<td>41</td>
</tr>
<tr>
<td>CWL</td>
<td>43</td>
<td>6</td>
</tr>
<tr>
<td>CWOL</td>
<td>35</td>
<td>5</td>
</tr>
</tbody>
</table>

### 3.3.2. α-diversity

The α-diversities of the bacterial intestinal microbiomes were calculated using FPD, in order to determine if there were different average numbers of bacteria between groups. First, the α-diversity of 3 month old children versus 1 year old children was compared, regardless of CMV status (Figure 3.1). A statistically significant difference in FPD of the intestinal microbiome at 3 months of age vs. 1 year of age was found ($p < 0.0001$). To investigate the impact of CMV status on the composition of the intestinal microbiome, 3-month and 1-year samples were analyzed separately. A statistically significant difference was found between CMV infected and uninfected children in the 3 month old group, with CMV infected children having a significantly lower α-diversity (Figure 3.2.a. $p = 0.028$). This difference was no longer present by 1 year of age, as
there was no statistically significant difference between CMV-infected and -uninfected children in this age group (Figure 3.2.b, p = 0.8).

Many factors in the first year of life impact the intestinal microbiome, of which breastfeeding duration in months, birth mode and any antibiotics given to the infant were available for analysis as possible contributing factors to differences in $\alpha$-diversity. These factors, including CMV status, were each analyzed individually in a linear regression model using FPD as the dependent outcome (Table 3.2). Independently, CMV status ($p = 0.016$), breastfeeding duration ($p = 0.003$) and the birth by CWL ($p = 0.010$) (using vaginal delivery as reference category) were all found to be significant contributors to the $\alpha$-diversity variation as calculated by FPD (Table 3.2). None of these factors had statistically significant interactions with each other, and after multivariable regression, all were still statistically significant contributors. The final linear regression model indicates that CMV infection, breastfeeding and CWL delivery all decrease the $\alpha$-diversity of the intestinal microbiome at 3-months of age (Table 3.2).
Figure 3.1. Comparison of bacterial α-diversity of the intestinal microbiome as calculated using Faith’s phylogenetic diversity (FPD) between all children (CMV infected and uninfected) at 3 months and 1 year of age. α-diversity was statistically higher in 1 year old than 3 month old infants as calculated by the Wilcoxon rank sum test (p<0.0001).
Figure 3.2. Comparison of bacterial α-diversity of the intestinal microbiome as calculated using Faith’s phylogenetic diversity (FPD) between CMV infected and uninfected children. CMV status and microbiome data from a) 3 months and b) 1 year of age were analyzed. a) At 3 months of age CMV-infected infants had a significantly lower α-diversity than uninfected children as calculated by a Wilcoxon rank sum test (p = 0.028). b) α-diversity was not statistically different between CMV infected and uninfected children at 1 year of age (p=0.8).
Table 3.2. Linear regression of FPD index using multiple factors that contribute to intestinal microbiome α-diversity in the first 3-months of life. Final model: FPD measurement = 4.770 – 0.453*(CMV positive) – 0.234*(each month of breastfeeding) – 0.547*(CWL delivery).

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Unadjusted</th>
<th>Adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>p-value</td>
</tr>
<tr>
<td>CMV status</td>
<td>-0.497</td>
<td>0.016</td>
</tr>
<tr>
<td>Duration of breastfeeding</td>
<td>-0.268</td>
<td>0.003</td>
</tr>
<tr>
<td>Birth mode</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CWL</td>
<td>-0.554</td>
<td>0.010</td>
</tr>
<tr>
<td>CWOL</td>
<td>-0.181</td>
<td>0.437</td>
</tr>
<tr>
<td>Antibiotics</td>
<td>-0.246</td>
<td>0.098</td>
</tr>
</tbody>
</table>

3.3.3. β-diversity and relative abundance of specific phyla and genera

The bacterial β-diversities of the intestinal microbiomes were analyzed to determine if there were differences in bacterial composition between CMV infected and uninfected children. Weighted UniFrac analysis was first calculated, which uses the branch lengths of the phylogenetic tree that are shared or unique to each sample and weights these lengths by the relative abundance of each sample\(^92,93\). Once the distances are calculated using weighted UniFrac, a permutational multivariate analysis of variance (PERMANOVA) was performed using these distances to determine the statistical significance of the compositional differences between groups of interest. The β-diversity between 3 month and 1 year old children, regardless of CMV status was first analyzed, and found to be significantly different (Figure 3.3.; p = 0.001). Next, β-diversity by CMV status was analyzed for the 3-month and 1-year old groups separately (Figure 3.4.). In 3 month old infants, the bacterial compositions were statistically different between those that were
CMV-infected versus -uninfected (Figure 3.4.a.; p = 0.01). This difference was no longer present by 1 year of age (Figure 3.4.b.; p = 0.179).

Bacteria were analyzed at all levels from phylum down to individual OTUs (phylum, class, order, family, genus, species) for differences in composition between CMV infected and uninfected infants at 3 months of age. There were no statistical differences between CMV-infected and -uninfected groups at any level as calculated using multiple testing of Wilcoxon rank sum test with the false discovery rate correction applied. Visually, there are differences in the number of bacteria in each phylum by relative abundance as seen in Figure 3.5. There was an increase in Proteobacteria and Bacteroidetes, a decrease in Firmicutes and Actinobacteria and no change in Verrucomicrobia in CMV-infected samples compared to -uninfected, again, however, none of these differences were statistically significant. Increases and decreases in the number of bacteria are also visually apparent, though not statistically significant, at the genus level and by far the most genera represented are within the firmicutes phylum (Figure 3.6.)
**Figure 3.3.** Comparison of bacterial β-diversity of the intestinal microbiome as calculated by Weighted UniFrac distances followed by principal coordinate analysis (PCoA) between children at 3 months and 1 year of age. Ellipses denote 95% confidence intervals. PERMANOVA was performed using weighted UniFrac distances to determine statistical differences between groups, p = 0.001.
Figure 3.4. Comparison of bacterial $\beta$-diversity of the intestinal microbiome as calculated by Weighted UniFrac distances followed by principal coordinate analysis (PCoA) between CMV infected and uninfected children at a) 3 months and b) 1 year of age. Ellipses denote 95% confidence intervals. PERMANOVA was performed using weighted UniFrac distances to determine statistical differences between groups; a) $p = 0.01$, b) $p = 0.179$. 
**Figure 3.5.** Comparison of the bacteria present in the intestinal microbiota at the phylum level between CMV infected and uninfected infants at 3 months of age. No statistical differences were found between any phyla as calculated by multiple testing using the Wilcoxon rank sum test with false discovery rate calculations applied using the Benjamini and Hochberg method.
Figure 3.6. Comparison of the bacteria present in the intestinal microbiota at the genus level between CMV infected and uninfected infants at 3 months of age. CMV status: “+” denotes infected and “-” denotes uninfected children. No statistical differences were found between any genera as calculated by multiple testing using the Wilcoxon rank sum test with false discovery rate calculations applied using the Benjamini and Hochberg method.
3.4 Discussion

The intestinal microbiome is constantly changing and diversifying during the first couple of years of life until an adult-like microbiota is established\textsuperscript{49,94}. Many environmental factors can impact this establishment, and here we see that CMV infection could be one of them. As expected, the $\alpha$-diversity was significantly higher and the $\beta$-diversity significantly different in 1 year old children than in 3 months old infants, regardless of CMV status, since the microbiome is expected to increasingly diversify during that time period\textsuperscript{49,94}. Interestingly, CMV infection by 3 months of age significantly decreased the $\alpha$-diversity and altered the composition ($\beta$-diversity) of the intestinal microbiome compared to uninfected infants. This difference was temporal, as it was no longer evident by 1 year of age. In line with this finding, a CHILD Study publication comparing children with atopy and wheezing to control (healthy) children documented a temporal reduction of specific bacteria at 3 months of age that also disappeared at 1 year of age\textsuperscript{63}. In this study, a reduced abundance of \textit{Faecalibacterium, Lachnospira, Rothia, Veillonella} and \textit{Peptostreptococcus} at 3 months of age was associated with children that had positive skin prick tests and clinically relevant wheezing at 1 year of age that was no longer detected at 1 year of age\textsuperscript{63}.

Other known factors, breastfeeding, delivery mode and antibiotic use, that affect the composition of the intestinal microbiome were considered and adjusted for using a linear regression model with $\alpha$-diversity as the predicted outcome. Previous studies show a marked increase in diversity of the intestinal microbiome after the introduction of solid foods, and one study described that it is actually the cessation of breastfeeding, not the introduction of foods, specifically that results in an increase in diversity in the first year of life\textsuperscript{51,95}. Breastfeeding was a significant contributor to the variability in $\alpha$-diversity, and decreased diversity with each month of continued
breastfeeding. Infants delivered by C-section have been found to have lower bacterial diversity and altered microbial compositions\textsuperscript{49}. It was interesting that C-section with labour contributed to decreased $\alpha$-diversity, but that C-section without labour did not. One surprising result was that antibiotic use in the first 3 months of life was not a significant contributor to the variability in $\alpha$-diversity, in contrast to previous studies\textsuperscript{49,94}. CMV infection continued to contribute to the decreased $\alpha$-diversity even after adjusting for breastfeeding and delivery mode.

It is not possible to distinguish here whether the changes in bacterial abundance and diversity observed here are deleterious, advantageous or neutral regarding health outcomes, particularly without being able to tease apart which bacteria are driving the differences. Importantly, however, this is the first indication that CMV infection can impact the composition of the intestinal microbiome in healthy infants.
Chapter 4. CMV and atopic diseases

4.1. Introduction

CMV is highly prevalent in the human population and has evolved with humans through the ages\textsuperscript{28,31}. CMV is part of the human virome, and infection is life-long. Thus, CMV may be one of the “old friends” hypothesized to be beneficial to healthy development\textsuperscript{3,73}. Indeed, CMV infection, after EBV infection, in the first 2 years of life has been found to decrease risk of IgE sensitization\textsuperscript{75}. CMV has many immunomodulatory functions that may contribute to the Th1/Th2 T cell balance/imbalance that is involved in immune-mediated diseases\textsuperscript{28,33,56,60,61}. Additionally, as demonstrated in the previous chapter, CMV impacts the diversity of the intestinal microbiome during the early-life critical “window of opportunity” where the microbiome and immune systems are concurrently developing\textsuperscript{49,52–54}. Through these mechanisms, CMV could be involved in healthy immune programming that prevents the later development of allergic diseases. Therefore, using the CHILD study we aimed to determine if early-life CMV infection is associated with development of eczema and asthma in Canadian children.

4.2. Methods

4.2.1. Clinical assessments

During 5 years of follow-up, the CHILD Study examined each participant for evidence of asthma and atopic dermatitis at 1, 3, and 5 years old using standardized clinical assessments\textsuperscript{84}. As the 5-year data were still relatively new at the time of this study, and not yet as refined as the 3-year data, 3-year diagnoses of asthma or atopic dermatitis are used here. A child was diagnosed with asthma by two possible methods: 1) at the 3-year clinical assessment by a CHILD physician, or 2) they had recurrent wheeze between 2 and 3 years of age along with either parental history of
asthma or physician-diagnosed eczema\textsuperscript{63}. To document evidence of wheezing parents completed the International Study of Asthma and Allergies in Childhood (ISAAC) questions\textsuperscript{84}.

A child was considered to have eczema if diagnosed by a CHILD or non-CHILD clinician as reported by the parents in the 3-month or 1-year health questionnaire\textsuperscript{63}. Using the results of skin prick tests, a more specific and sensitive variable of atopic phenotype, or more simply atopy, was developed. Children were tested with 10 different allergens at 1 year of age through a skin prick test, and a child was considered to have atopy if a greater than 2 mm wheal was measured in response to any allergen\textsuperscript{63}. The wheal measurement was corrected against a negative control (glycerin) and for location (arm vs back). From these measurements a binary variable was created: a “0” was assigned if there was no wheal greater than 2 mm to any allergen, and “1” assigned if there was a reaction to at least one of the allergens.

4.2.2. Statistical analysis of associations

The associations between atopic disease outcomes at 3 years of age and CMV status at either 3 months or 1 year of age were evaluated using Fisher’s Exact tests. Fisher’s Exact tests were also used to determine associations between atopic phenotype at 1 year of age and CMV status at either 3 months or 1 year of age. All analyses were carried out in the R statistical programming language, and a p-value < 0.05 was considered statistically significant.

4.3. Results

4.3.1. Association between CMV infection and allergic diseases

CMV infection by 1 year of age was not found to be statistically associated with asthma diagnosis (Figure 4.1.a.; p= 0.380) or eczema diagnosis (Figure 4.1.b.; p= 0.447) by 3 years of age. Since there was a difference in the microbiome $\alpha$-diversity at 3 months of age that disappeared at 1 year, and early life intestinal microbial perturbations have been previously
linked to the development of atopic diseases, asthma and eczema status were also investigated for an association with CMV infection by 3 months of age. Again, there was no statistical association with CMV status at 3 months of age and asthma (Figure 4.2.a. p = 0.380) or eczema (Figure 4.2.b. p = 0.306) diagnosis by 3 years of age.

**Figure 4.1.** Associative analysis of a.) asthma diagnosis and b.) eczema diagnosis at 3 years of age by CMV status at 1 year of age. In a.) 916 children were evaluated for asthma; 745 were CMV negative and 171 CMV positive. Only 36 were diagnosed as asthmatics, 27(4%) within the CMV negative group and 9(5%) in the CMV positive group. In b.) 944 children were evaluated for eczema; 766 were CMV negative and 178 CMV positive. Of the 116 that were diagnosed with eczema, 91 (12%) were within the CMV negative group and 25(14%) in the CMV positive group. There were no statistically significance associations between CMV status at 1 year and asthma or eczema diagnosis, as calculated by Fisher’s Exact test.
Figure 4.2. Associative analysis of a.) asthma diagnosis and b.) eczema diagnosis at 3 years of age by CMV status at 3 months of age. In a.) 901 children were evaluated for asthma; 782 were CMV-negative and 119 CMV-positive. Only 35 were diagnosed as asthmatics, 30 (4%) within the CMV negative group and 5 (4%) in the CMV positive group. In b) 925 children were evaluated for eczema; 800 were CMV-negative and 125 CMV-positive. Of the 114 that were diagnosed with eczema, 95 (12%) were within the CMV-negative group and 19 (15%) in the CMV positive group. There were no statistically significance associations between CMV status at 3 months and asthma or eczema diagnosis, as calculated by Fisher’s Exact test.

4.3.2. Association of CMV infection and atopic phenotype

The diagnosis of asthma and atopic dermatitis, although meaningful, can be subjective and difficult to diagnose consistently, whereas an atopic phenotype can be characterized very definitively using skin-prick tests. For this reason, a more specific and sensitive variable was
created describing atopic phenotype as a binary outcome, as described in methods. Atopic
phenotype and CMV infection by 1-year of age were statistically associated (Figure 4.3.a.: p =
0.005, OR [95% CI] = 2.030 [1.215-3.356]). This association was also present if already infected
by 3 months of age, and actually increased risk of positive atopic phenotype by 1 year of age
(Figure 4.3.b.: p = 0.0006, OR [95% CI] = 2.614 [1.473-4.595]).

**Figure 4.3.**  Associative analysis of a.) CMV status at 1 year and b.) CMV status at 3 months
with atopic phenotype at 1 year of age. In a.) 513 children were evaluated; 415 were CMV-
negative and 98 CMV-positive. Atopy was characterized in 120 children, 86 (21%) within the
CMV-negative group and 34 (35%) in the CMV-positive group. Association between atopy and
CMV status at 1 year was calculated using Fisher’s Exact Test, with a p-value = 0.005. In b) 514
children were evaluated; 445 were CMV-negative and 69-CMV positive. Of the 120 that were
atopic, 92 (21%) were within the CMV-negative group and 28 (41%) in the CMV-positive
group. Association between atopy and CMV status at 3 months was significant with p < 0.001.
4.4 Discussion

Early-life CMV infection, by 1 year or 3 months of age, was not associated with the clinical outcomes of asthma or eczema as defined by the CHILD Study (Figures 4.1. and 4.2.). There has not been any definitive evidence that CMV infection would protect or promote the development of allergic diseases, although an interplay between CMV and EBV infection and protection against IgE sensitization has been suggested. As previously mentioned, one study found that CMV infection enhanced the protective effect of EBV infection against IgE sensitization, however CMV infection alone was not correlated with protection against IgE sensitization\textsuperscript{75}. Contrarily, another group found no association between CMV seropositivity and asthma or eczema but found that CMV infection in the absence of EBV infection was associated with the presence of IgE antibodies to airborne and food allergens\textsuperscript{96}. Potential correlations between CMV serostatus and numbers of cytokine producing cells in the peripheral blood has also been explored\textsuperscript{97}. CMV-infected children had higher numbers of IFN-gamma and lower numbers of IL-4 producing cells than children without CMV, which is an immune profile that would be expected to be associated with protection against IgE sensitization; however, CMV seropositivity was not associated with decreased IgE sensitivity\textsuperscript{97}.

Allergic diseases develop in a progressive manner from eczema in infants to asthma in children, and this progression is termed the “atopic march”. Eczema may increase the risk of developing further allergic diseases by allowing sensitization to allergens through a compromised and dysfunctional skin barrier\textsuperscript{57,98}. This is hypothesized to initiate the atopic march by allowing allergens to be captured and processed by antigen-presenting cells that then migrate to lymph nodes to interact with T cells leading to the Th2 immune skew and systemic allergic diseases\textsuperscript{57,98}. In support of this hypothesis, one study found early onset eczema to be a main risk factor for
progression to asthma later in childhood\cite{98}. Here, the atopic phenotype defined using the skin-prick-tests was used as an early indication of possible development, or a predisposition to allergic diseases. This is a more sensitive and specific measurement in comparison to clinical outcomes of asthma and eczema. Early-life CMV infection, by either 1 year or 3 months of age, was positively associated with an atopic phenotype at 1 year of age (Figure 4.3.). This outcome is similar to the above describe study that found no associations between CMV seropositivity and asthma or eczema, but an increase in IgE antibodies to food and airborne allergens\cite{96}. Contrary to our hypothesis that early-life CMV infection may help to prevent later development of allergic diseases, these results suggest that it may actually contribute to a predisposition towards allergic diseases.

The etiology of allergic diseases is complicated, and many risk factors have been associated with asthma and eczema\cite{56}. Although there is no definitive evidence of CMV infection as a risk factor for allergic diseases, the association made here between early-life CMV infection and atopic phenotype provides support for further investigation of CMV being an environmental factor that influences health and disease outcomes throughout life.
Chapter 5. Conclusions

This study measured the prevalence of CMV in healthy Canadian infants for the first time in 19 years, and for the first time ever across multiple cities\textsuperscript{18,19}. The prevalence of CMV infection in infants in Canada now (19\%) appears quite similar to what was found in the late 60s and 90s in isolated areas, suggesting that CMV transmission has not changed much over this time period\textsuperscript{18,19}.

The risk factors of CMV infection in the first year of life among healthy infants were identified as breastfeeding and parental ethnicities. The vast database of the CHILD Study provided many potential risk factors of CMV infection to investigate. However, the data were not collected with this study/purpose in mind. In particular, the variable that was used here as an indication of daycare attendance was not specific to daycare visits but included any regular visits outside of the home. This may have diminished our ability to identify daycare as a risk factor for CMV infection, as daycare attendees (who are typically exposed to other young children shedding CMV in urine and saliva) were grouped together with children that regularly visited other family members (which likely does not confer the same risk of CMV exposure).

This study again highlights that breastmilk is a source of CMV transmission. The benefits of breastmilk are well-established and outweigh any potentially detrimental effects of postnatal CMV infection. It has been suggested that short-term heat pasteurization (5 sec at 62\(^{\circ}\)C) of CMV-infected breastmilk can reduce the transmission of CMV, which could be a potential strategy to prevent infection in low-birth-weight pre-mature infants that are at risk of disease from post-natal CMV infection\textsuperscript{25}. More practically, preventing CMV transmission in the population through vaccination could prevent disease in congenital and post-natal infections. To
fully understand the contribution of breastfeeding on CMV transmission in Canada, the CMV status of individuals in the household would be helpful and optimally breastmilk would be tested for the presence of CMV.

For the first time, we describe that early-life CMV infection can temporarily alter the intestinal microbiome. CMV infection significantly reduced the α-diversity and altered the β-diversity of the intestinal microbiomes of infants at 3 months of age, and these differences were undetectable at 1 year of age. Unfortunately, which organisms were driving the alterations could not be determined from this data. The 16S rRNA samples available for analysis represented only approximately half of the samples that have been tested for CMV status in the urine. In the future, the analyses described in this work will be repeated on data provided by the SyMBIOTA group for the remainder of the samples that have been tested for CMV status. This will help confirm the diversity differences identified here and ideally increase the ability to detect bacterial species driving the alterations in diversity at 3 months of age. Ultimately, this is the first indication that early-life CMV infection impacts the diversity of the developing microbiome, which is critical to the education of the developing immune system occurring at this same “critical window” in early life.

CMV infection by 3 months or 1 year of age was not associated with a diagnosis of eczema or asthma by 3 years of age in this study. However, CMV infection by either age was positively associated with an atopic phenotype (positive skin-prick test) by 1 year of age. Diagnosing allergic diseases is difficult to unify and maintains a level of subjectivity from the diagnosing clinicians and parents that are completing the surveys. The CHILD Study made use of the ISAAC questionnaires which were developed to provide simple and consistent methods for measuring asthma and eczema for international comparisons. However, these questionnaires
were developed for the purpose of surveying 6-7 year old and 13-14 year old children\textsuperscript{99}. The definitions of asthma and eczema for the CHILD Study still relied on consistent use of the questionnaires by parents and consistent diagnoses by multiple clinicians, and discrepancies in the diagnoses could mask associations.

It is possible that there is a very specific small window of time in which CMV can cause negative impacts on the microbiome and promote allergic disease outcomes. Here, only 3-month and 1-year samples were analyzed, but the infant microbiome and immune system is rapidly changing throughout the first couple of years of life. The order in which the microbiome and immune systems are exposed to different viruses and bacteria may alter the long-term effects of these infections\textsuperscript{3}. This has been demonstrated previously where CMV in the absence of EBV was found to be associated with IgE sensitization to airborne and food allergens\textsuperscript{96}. Our study adds support to the evidence that protecting against CMV infections (i.e., through vaccination) would not shift children towards a more allergic profile. The data collected here allows for the addition of CMV to the analysis of risk factors for later development of allergic diseases. Indeed, the CHILD Study has continued to collect data from these children and the identification of risk factors for allergic diseases is an ongoing project by another CHILD Study investigator group to whom the CMV data has been provided.

Finally, although this demonstrates yet another way in which CMV causes detrimental effects, it also provides evidence that if we eliminate CMV from the population through vaccination, we are not eliminating an “old friend” that may help prevent allergic diseases. Considering the enormous burden of disease that CMV infection has on congenitally-infants and immunocompromised people, eradicating this infection could only be considered an invaluable accomplishment for humanity.
References


25. Hamprecht K, Goelz R. Postnatal Cytomegalovirus Infection Through Human Milk in


40. Brodin P, Jojic V, Gao T, et al. Variation in the human immune system is largely driven


82. Cannon MJ, Stowell JD, Clark R, et al. Repeated measures study of weekly and daily cytomegalovirus shedding patterns in saliva and urine of healthy cytomegalovirus-


