Genetic adaptation contributing to increased gonococcal fitness during vaginal infection of CEACAM-humanized mice

by

Jessica Lam

A thesis submitted in conformity with the requirements
for the degree of Master of Science

Department of Molecular Genetics
University of Toronto

Genetic adaptation contributing to increased gonococcal fitness during vaginal infection of CEACAM-humanized mice

Jessica Lam

Master of Science

Department of Molecular Genetics
University of Toronto

2017

Abstract

Neisseria gonorrhoeae (Gc), the obligate human pathogen responsible for gonorrhea, has drastic consequences on reproductive health, if left untreated. Although, vaginal infection in wild-type mice can simulate asymptomatic colonization, important host-specific factors essential for human infection are overlooked. I utilize mice expressing human CEACAM3, CEACAM5, and CEACAM6, to reveal factors promoting mucosal colonization. By serial passaging Gc in the lower genital tract of these mice, I demonstrate a higher proportion of mice colonized in subsequent infections, suggesting selection for bacteria better capable of colonization. No changes in susceptibility of bacteria to low pH, complement-mediated killing or antimicrobial peptides, mCRAMP and mBD-1, were observed. Finally, a genome-wide analysis identified mutations acquired by *in vivo* adapted isolates. One mutation has previously been investigated as a promising vaccine target, validating the infection model and screening process. Further examination of other identified changes may provide insight into the Gc lifestyle within the genital mucosa.

Acknowledgements

Firstly, I wanted to thank my supervisor, Scott Gray-Owen. You are one of the kindest, most knowledgeable and understanding supervisors I could ask for. Thank you taking the risk of accepting a student with little previous background knowledge. I will be forever grateful for the opportunity to learn and grow as a scientist in your lab. I would not have accomplished all that I have without your continuous support, patience and guidance.

To my committee members, Alex Ensminger and John Parkinson, thank you for your scientific advice and for providing me with insight into programs to help achieve my scientific goals.

Over the past 2 years, there are numerous people that have helped me grow both personally and academically. Thanks to the past and present members of the Gray-Owen lab that facilitated my academic growth. Carolyn, Anna, Ryan, Epshita, and Stacey, you have all been exemplary role models, demonstrating what it means to be a good scientist. Your combination of knowledge, scientific merit and confidence in your own abilities have been constant sources of inspiration. I wanted to especially thank Epshita for teaching me everything I know about mouse handling.

Don't worry Cynthia Guo, I did not forget about you! Like the others mentioned above, you have been an outstanding scientific role model. Your self-confidence, critical thinking and hard-working nature has made me aspire to be even half of the scientist you already are. I will never be able to express how much your friendship, humour and constant support has shaped me into the person I am currently. All our late evenings in the lab, InteractHome and Prenup Pub have made me love my graduate school experience and I can only hope that it doesn't change! Honorary Gray-Owen member, Vivian Cheung, you have been an amazing source of support, both academically and emotionally! Thank you for being my sounding board and talking sense into me when I doubted my own capabilities. I am very glad to have met you and to have you in my life. My sanity would not remain intact, if not for my graduate school best friend, Erin Wong! Thank you for clinging onto me and forcibly removing me from inside my shell so early on in our graduate school careers. Your caring nature and ability to see the positive in everything and everyone has contrasted my pessimism perfectly, making me a better person because of it. I also wanted to thank all the other friends I have made in MSB over the past 2 years, Sabrina Stanley, you're also amazing! I also wanted to thank my boyfriend, Vicky Patel. You have been a constant crutch throughout all aspects of my life. I really appreciate your patience, warm-hearted nature, always available ear to listen to my complaints, ability to make me forget all my problems, and

yes… occasionally, even your stubbornness. You have had a profound impact on who I am today and I am grateful for everything you have done for me.

Although I don't always express it, I am extremely appreciative for the support of my family. To my mom, thank you for everything you have done. Early on in life, you have always motivated me to be the best version of myself and provided me with all the necessary tools you could to ensure my success. David Lam, thank you for teaching me patience and providing me a constant source of laughter and entertainment through the years. Linda Lam, though I will never say it verbally, you have been the most important role model in my life. You have always been there to support all my life choices, selflessly given me anything you could without question, listened attentively to all my issues, and I appreciate your brutally honest opinion. I will never forget all that you have done to help me get to this point of my life.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AMP – Antimicrobial peptide

BHI – Brain heart infusion

C4BP – Complement 4b binding protein

CDC – Centre for Disease Control

CEACAM – Carcinoembryonic antigen-related cell adhesion molecule

CMP-NANA – Cytidine-5'-monophospho-N-acetylneuraminic acid

DNA – Deoxyribonucleic acid

fH – Factor H

Gc – *Neisseria gonorrhoeae*

ID – Infectious dose

IL-6 – Interleukin-6

IMG – Integrated microbial genomes

IP – Intraperitoneal

ISO – Isovitalex

KC – Keratinocyte-derived chemokine

KDO – 3-Deoxy-D-manno-oct-2-ulosonic

Lgt – Lipooligosaccharide glycosyl transferase

LOS – Lipooligosaccharide

LPS – Lipopolysaccharide

MAC – Membrane attack complex

mBD1 – mouse β-defensin 1

mCRAMP – mouse cathelicidin related antimicrobial peptide

MIP-2 – Macrophage inflammatory protein 2

NCBI – National Center for Biotechnology Information

NGS – Next-generation sequencing

OD – Optical density

Opa – Opacity-associated

P1 – Passage 1

P2 – Passage 2

P3 – Passage 3

PAMP – Pathogen-associated molecular pattern

PBS – Phosphate buffered saline

PID – Pelvic inflammatory disease

PMN – Polymorphonuclear neutrophil

PRR – Pattern recognition receptor

STI – Sexually transmitted infection

TNFα – Tumor necrosis factor alpha

VCNT – Vancomycin, colistin, nystatin, trimethoprim

WGS – Whole-genome sequencing

Chapter 1 – Introduction

1.1 – Disease

Neisseria gonorrhoeae (Gc) is a Gram-negative, diplococcus bacteria that is the causative agent of the sexually transmitted infection (STI) gonorrhea. With a worldwide incidence rate of 106 million cases reported annually, gonorrhea is the second most prevalent bacterial STI[1]. Although infections occur in both sexes, there are several gender-specific differences in terms of disease manifestation, immunological response and bacteria-host interactions[2].

The initial site of Gc infection is typically mucosal tissue of the male urethra or female endocervix. In men, the onset of symptoms typically arise within 2-14 days post-exposure[3]. The characteristic symptoms include pain during urination and a purulent urethral exudate containing a large population of gonococci associated with polymorphonuclear neutrophils (PMNs). Without treatment, bacterial infections can result in epididymitis and sterility. Conversely, 80% of females are asymptomatic[4]. Humans are the sole natural host of Gc; therefore, asymptomatic colonization in females is advantageous for maintaining bacterial populations that can be transmitted to other individuals. Moreover, in the absence of symptoms, most women do not seek treatment, leading to prolonged infections and, often, ascension of Gc into the upper genital tract. The resultant complications include pelvic inflammatory disease (PID), ectopic pregnancy and sterility as the inflammatory response to Gc inadvertently damages the infected tissue

Other frequent sites of natural infection include the human nasopharynx and rectum. Infection of the oral cavity has been implicated in increasing antibiotic resistance. *Neisseria sp.* are naturally competent for genetic transmission by neisserial DNA and some strains actively export their own DNA by the Type IV secretion system[5,6]. Since the nasopharynx contains various commensal *Neisseria* species and, occasionally *Neisseria meningitidis* nasopharyngeal colonization provides an opportunity for the Gc to acquire beneficial genetic material such as genes encoding drug resistance.

Exposure of Gc into the conjunctiva of babies during exit of the birth canal can result in blindness, if left untreated. Additionally, Gc is capable of causing disseminated gonococcal infection if it migrates to other regions of the body resulting in septic arthritis, meningitis and endocarditis.

1.2 – Treatment options

Treatment for gonorrheal infections during the past century has entailed the use of antibiotics. However, throughout the decades, Gc has developed drug resistance to most major classes of antibiotics[7]. Most Gc isolates remain susceptible to azithromycin, tetracycline and the third-

generation cephalosporin class of antibiotics, which include cefixime and ceftriaxone. Nevertheless, in recent years, there has been a rise in the number of clinical isolates demonstrating increasing levels of antibiotic resistance to one or more of the antibiotics mentioned above. Cephalosporin resistance (both cefixime and ceftriaxone) has been documented in many countries including Japan[8], France[9] and Sweden[10]. Thus, in 2012, Centers for Disease Control and Prevention (CDC), has suggested dual treatment regimens consisting of administering azithromycin in combination with either cefixime or ceftriaxone[11]. The implementation of this regimen is to ensure the complete clearance of Gc, in an effort to reduce the spread of antibiotic resistance worldwide. However, given that Gc strains-resistant to either one or the other of these drugs has already emerged, this is a stop-gap measure. With the rate at which Gc has developed resistance to all other prescribed antibiotics; the emergence of untreatable gonorrhea in upcoming years is an urgent global health concern. Since there is presently no vaccine available for the prevention of gonorrhea, there is an urgent necessity to identify novel vaccine and drug targets.

1.3 – Bacteria-host interactions

As previously mentioned, Gc is a human-restricted pathogen with no other natural reservoir. Consequently, Gc has become extremely adept at survival within the urogenital tract of humans. The bacteria-host interactions associated with infection are complex, such that Gc is capable of evading the host immune response yet, in other instances, it is also capable of triggering a massive inflammatory response (**Figure 1**). These differences can be further exemplified in the progression of disease manifestation in males and females.

1.3.1 – Virulence factors

Through evolution alongside humans, Gc has developed mechanisms with the sole purpose of promoting survival in their mucosal niche, including some that directly contribute to disease. For Gc, virulence factors aid in bacterial pathogenesis by enhancing bacterial adhesion to host epithelial cells, manipulation of the host immune system, facilitating nutrient acquisition within the tissues, and masking of bacteria specific surface structures to prohibit identification. Since these are known to be important for infection, I will briefly touch on each here.

Bacterial adherence to epithelial cells is the essential first step in establishing infection. During infection, Gc utilizes a Type IV pilus, a long filamentous extension, that functions to anchor the bacteria onto epithelial cells. Subsequent retraction of the pilus allows for the bacteria to interact with the host cell surface. After initial attachment, the Opacity-associated proteins (Opa) facilitate

intimate bacteria-cell interactions. Opa adhesins aid in the invasion of Gc into epithelial cells, allowing for the transmigration of the bacteria into the submucosal space[12,13]. The submucosa is a nutrient-rich environment with limited access for bacteria. Thus, once in the submucosa, Gc is able to proliferate without competing for nutrients and space with other commensal bacteria. The ability to evade immune processes and access host nutrient stores within this niche are key attributes of the Gc infectious lifestyle.

Gonococcal isolates retrieved from male clinical samples are predominantly Opa^+ in expression, suggesting Opa plays a crucial role in the infection process[14]. These findings can be recapitulated in male urethral challenges, where men infected with a predominantly Opa^- (transparent) inoculum become colonized by Gc that express one or more Opa proteins[14]. For women, Opa expression varies depending on the stages of the menstrual cycle. With the exception of during menses and shortly afterwards, most Gc isolates are Opa^+ throughout the menstrual cycle[15]. In terms of pilus expression, clinical samples are typically piliated[16], however, male urethral challenges with a *pilE* mutant result in asymptomatic colonization[17]. These data suggest that pilus is not required for initially establishing infection but may facilitate the development of symptomatic infection.

Another role of virulence factors is to enable immune evasion. The immune response to invading pathogens comprises of an innate and adaptive immune response. The innate immune response is a non-specific, immediate response to the presence of pathogens. One component of the innate immune response is the complement system. Activation of the complement pathway results in a signalling cascade leading to the formation of membrane attack complexes (MAC)[18]. These MACs form pores in the bacterial cell membrane, disrupting membrane integrity and ultimately results in cell death. It has been suggested that to combat this, Gc express porin protein B1A and B1B. Porin B proteins bind to human serum complement 4b binding protein (C4BP) and factor H (fH), which inhibit the classical and lectin or alternative complement cascades, respectively[19,20]. Without the ability to avoid complement-dependent killing, Gc would succumb to the effects of the complement cascade immediately upon entering the subepithelial space.

If pathogens are unable to be detected by the host immune response, this is advantageous to bacterial pathogenesis. Neisserial antigenic variation, described in the next section, allows Gc to evade recognition of specific antigens. Another mechanism by which Gc evades detection is through the modification of lipooligosaccharides (LOS) decorating the outer membrane of the

bacteria. Typically, the LOS structure in most bacteria have immunostimulatory effects and are often targeted by pattern recognition receptors (PRR) of the host. However, the LOS structure of Gc can quickly vary sugar moiety composition. If the LOS variant expressed terminates with a lacto-N-neotetraose sugar group, Gc is capable of sialylating the LOS structure utilizing host-derived sialic acid, in the form of cytidine-5'-monophospho-N-acetylneuraminic acid (CMP-NANA)[21]. Sialylation is an essential part in masking stimulatory epitopes on the LOS. Consequently, the bacterial surface mimics sialic acid residues present on mammalian host cells, which prevents the host immune system from being able to differentiate sialylated bacteria from self.

During infection, acquisition of nutrients, such as iron, is a key factor restricting Gc proliferation. Iron is fundamental for bacterial growth, but iron is severely limited within all mammalian tissues due to its effective sequestration within intracellular stores (such as associated with ferritin or hemoglobin) and through the high-affinity iron binding proteins transferrin and lactoferrin found in blood and mucosal secretions, respectively. To access iron, Gc express proteins capable of binding to extracellular hemoglobin, transferrin and lactoferrin, and to retrieve their iron[22].

The development of the above mentioned virulence factors, aid in promoting bacterial pathogenesis through all stages of infection. Pilus and Opa proteins facilitate initial attachment and entry into the subepithelial space. Iron-binding proteins assist in nutrient acquisition, whereas sialylation of the LOS structure and binding of C4BP and fH inhibit complement-mediated killing by the host innate immune response. In combination, all these factors play an important role in bacterial survival.

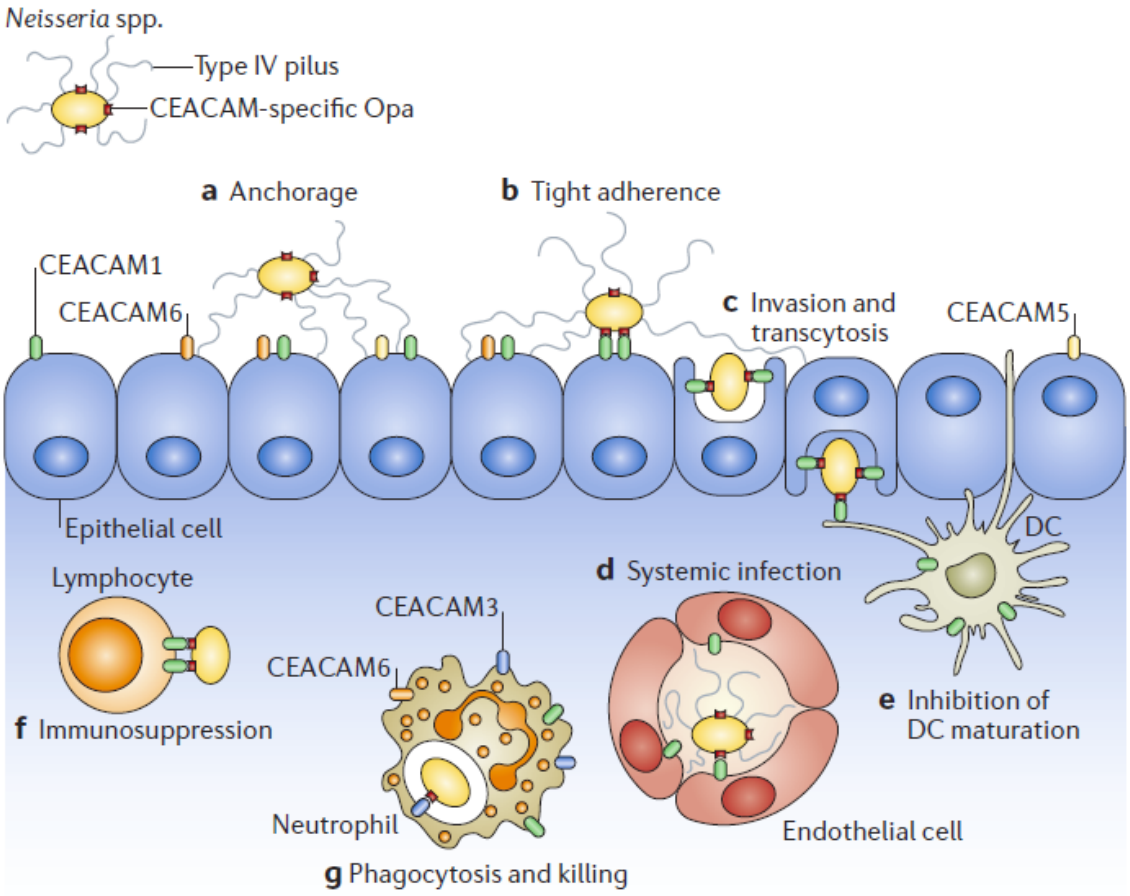


Figure 1. Gc-host interactions throughout various stages of infection[23]. *Neisseria* pilus enables initial anchorage to host epithelial cells (**a**), while Opa interaction with CEACAM1, -5 and -6 allow for tight adherence and transcytosis of bacteria in the submucosa (**b** and **c**). Presence of CEACAM1 on endothelial cells and host immune cells promote systemic infection (**d**) and immunosuppression and inhibition (**e** and **f**). Conversely, CEACAM3 expression on neutrophils results in phagocytosis and bacterial killing (**g**).

1.3.2 – Antigenic and phase variation

The ability of Gc to be able to avoid detection by the host adaptive immune system is largely due to the ongoing modification of surface-exposed antigens. This ongoing diversification maintains a population of bacteria that can persist as the immune response emerges, and prevents individuals from being protected upon future exposure to even the same bacterial strain.

Well-characterized mechanisms associated with host immune evasion by Gc include antigenic and phase variation. Antigenic variation refers to the ability to alter the sequence of expressed proteins through genetic recombination. For instance, the major subunit in pilus is encoded by *pilE*, which contains a promoter allowing for its expression[5]. Gc also maintains a repertoire of 19 silent *pilS* genes, all lacking a promoter[24]. Since the variable surface-exposed sequences are

interspersed with conserved sequences that provide the structural framework of pilin protein, portions of any *pilS* gene can recombine with *pilE*, resulting in expression of a "new" pilus phenotype (**Figure 2A**)[25].

Phase variation is another mechanism that can contribute to the diverse array of bacteria within the population. In *Neisseria*, phase variation occurs at a genomic level due to repeat sequences dispersed throughout the genome. During DNA replication, slip-strand mispairing frequently results in the insertion or deletion of repeat units[26,27]. Depending on the location of the sequence, protein expression levels can vary. For example, the leader peptide of Opa proteins, is encoded by a string of pentameric repeats (CTCTT), where changes in the number of repeat units can result in the protein coding sequence being shifted in or out of reading frame, altering levels of expression (**Figure 2B**)[28]. A single gonococcus encodes ~11 Opa proteins, each of which can independently phase on or off, and each of which alter both epitope expression on the bacterial surface and the host cell receptor binding capacity of these bacteria[29].

Not all repeats are several nucleotides long. In many instances, repeated tracts of a single nucleotide present within the coding sequence of genes can be responsible for changes that alter protein sequence or introduce a premature stop codon. For example, glycosyl transferases are enzymes responsible for the formation of glycosidic bonds between sugar moieties of the LOS structure. In the case of Gc, *lipooligosaccharide glycosyl transferase –A, –C* and *–D* (*lgtA, lgtC* and *lgtD*, respectively) are phase variable. All three contain a homopolymeric guanine tract in the coding sequence, where addition and subtraction of guanine leads to a frameshift and the introduction of a premature stop codon (**Figure 2C**)[30]. An experimental male urethral challenge demonstrated that the infectious dose (ID_{50}) of a Gc strain expressing *lgtA*, referred to as in-frame, only required 250 bacteria[31,32]. On the contrary, the ID_{50} for strains with an out-of-frame *lgtA* was 10^5. These findings further emphasize the importance of alterations to the outer membrane surface structures on gonococcal infectivity. With the usage of whole-genome analysis, Snyder and colleagues have identified over 100 possible phase-variable genes in Gc[33]; collectively, these genes greatly contribute to the vast diversity present within the bacterial population.

Both phase and antigenic variation contribute to the constant switching of expressed intracellular and surface exposed proteins, enabling bacteria to avoid detection by the host immune system. Due to the continuous modifications to surface structures, antigenic and phase variation provides a significant barrier for vaccine development due to the lack of conserved target epitopes.

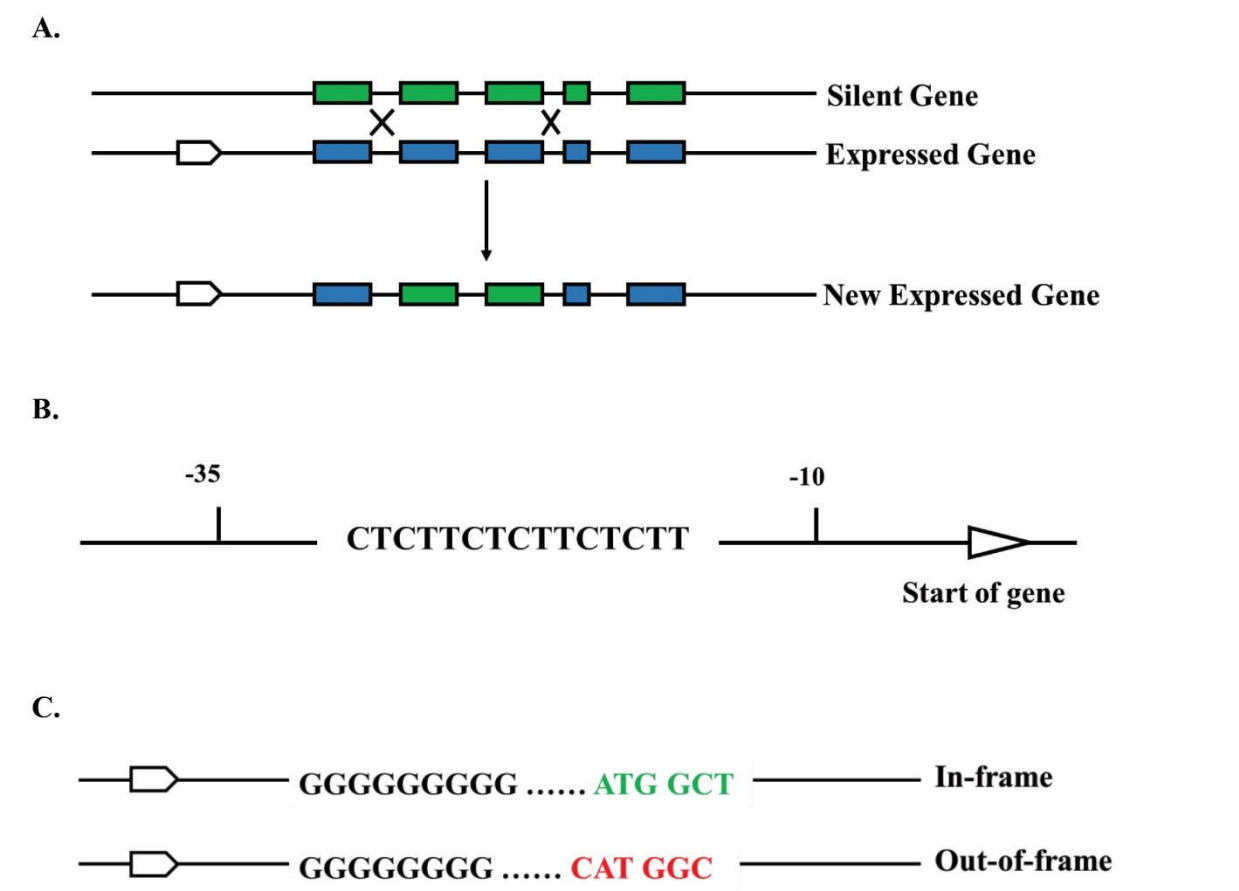


Figure 2. Methods of antigenic and phase variation in *Neisseria gonorrhoeae*. A) Antigenic variation of pilus phenotype. Silent gene (green) substitution for expressed gene (blue) via recombination, resulting in new phenotype. **B)** Presence of pentameric repeat sequence in promoter region of opa contributing to alteration in level of gene expression. **C)** Homopolymeric-G tract in the coding sequence of *lgtA, C* and *D*, where subtraction of one guanine results in introduction of premature stop codon (red)

1.4 – Animal models for studying Gc infection *in vivo*

Humans are the only known natural host of Gc, hence, studying bacteria-host interactions are greatly hindered by the lack of a representative animal model. Although human urethral challenges in males are permitted, this only provides insight into the first 10 days of infection, at which point antibiotics are employed to clear the infection. In contrast, due to the severity of complications that may arise in females, it is deemed unethical to conduct gonococcal challenges in women. Due to gender-specific differences in males and females, a direct comparison between advantageous genes identified in male urethral studies and genes essential for female infections cannot be extrapolated. Essentially, this leaves a large proportion of the population, who suffer from the highest morbidity, left unstudied.

Jerse and colleagues have demonstrated that Gc can consistently colonize the lower genital tract of laboratory mice that have been provided exogenous estradiol and antibiotics for an average of 6-10 days[34-36]. The administration of antibiotics is used to decrease commensal microflora to better promote the initial establishment of infection, providing an environment more amenable to Gc colonization.

Various morphological and physiological similarities exist between the genital tract of female mice and women, which permits mice to be an acceptable model for Gc infection. In terms of morphology, both consist of a lower genital tract with the cervix as a distinct junction separating the upper genital tract (**Figure 3**). The lower genital tract is a non-sterile environment, consisting of the vagina and cervix. In contrast, the upper genital tract is comparatively sterile and contains the uterus, fallopian tubes and ovaries. In mice, however, instead of a single uterine body, there are two uterine horns. From a physiological standpoint, the vaginal pH of mice treated with estradiol is 6.6 (ranging between 5.8-7.2)[37], whilst the average human vaginal pH is 3.5-4.5[38]. However, the mouse vaginal pH is more similar to the human female cervix, which is the common site of infection in women. The human cervical pH during the proliferative and secretory stages of the menstrual cycle, average pH is 6.8 and 6.1, respectively[38].

1.4.1 – Limitations of the mouse infection model

Although the reproductive tract of female mice and humans are quite similar and are both capable of being colonized by Gc, mice are not the natural host for Gc. Once the effects of the exogenous estradiol subside and mice proceed through their natural estrus cycle and enter diestrus, the rapid clearance of bacteria ensues. Furthermore, though mice may have homologues for various human receptors that facilitate Gc interactions, many of these are not recognized by Gc virulence factors. Thus, factors known to play a critical role in establishing infection must be taken into account when trying to elucidate essential components required for the infection process. As previously stated, Gc has develop many human-specific interactions. Therefore, the usage of a transgenic mouse line expressing receptors that facilitate essential Gc-host interactions provide better insight into factors necessary for establishing and maintaining Gc infection.

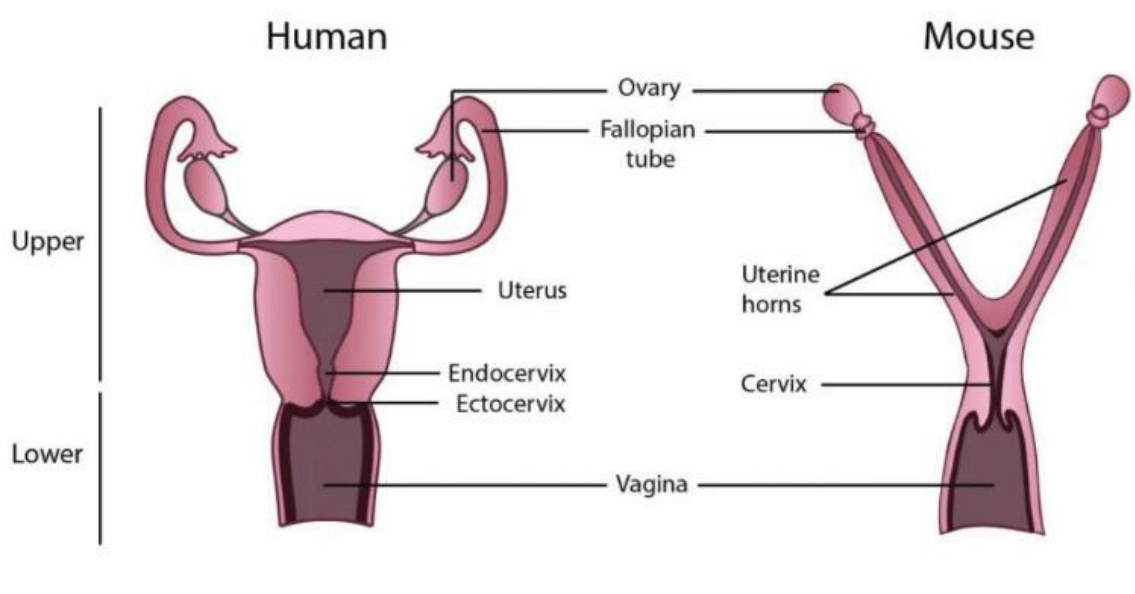


Figure 3. Anatomical similarities and differences between human and mouse female genital tract [39]. Genital tract of mouse and human consist of lower genital tract, where the cervix is a junction separating the upper genital tract. One major difference between human and mouse genital tract is the presence of a single uterine body and two uterine horns, respectively.

1.5 – CEACAM-Opa interactions

A key factor facilitating Gc-host interactions is the expression of Opa proteins. Adherence of Opa+ Gc to epithelial cells is mediated by their binding to carcinoembryonic antigen-related cell adhesion molecule (CEACAM) receptors[40,41]. CEACAMs are present on the mammalian cell surface, and are involved in intercellular adhesion and interactions between cells. In terms of Gc infection, CEACAM1, -3, -5 and -6 have been implicated in bacteria-host interactions (**Figure 4**)[42]. All leukocytes can express CEACAM1. CEACAM1, -5, and/or -6 can be co-expressed on epithelial cells, whereas CEACAM1, -3 and -6 are present on neutrophils[23]. Binding of Opa adhesins to CEACAM1 and -5 provides a bacterial advantage, as binding allows for the anchorage of Gc to the epithelial cell surface and CEACAM1 binding activates phosphatases that are immunosuppressive[43,44]. Alternatively, binding to CEACAM3 on neutrophils, activates bacterial engulfment, killing by pro-inflammatory cytokine expression (**Figure 1**). Despite Opa proteins being phase-variable, our group has observed that a vast majority, 93%, of low-passaged clinical isolates bind CEACAM1 and/or -5, respectively [45]. Conversely, only 27% of the same isolates bind CEACAM3 [45]. Consequently, these findings suggest there is positive selection for bacterial expression of Opa protein variants that bind CEACAM1 and -5, but selection against those that bind CEACAM3.

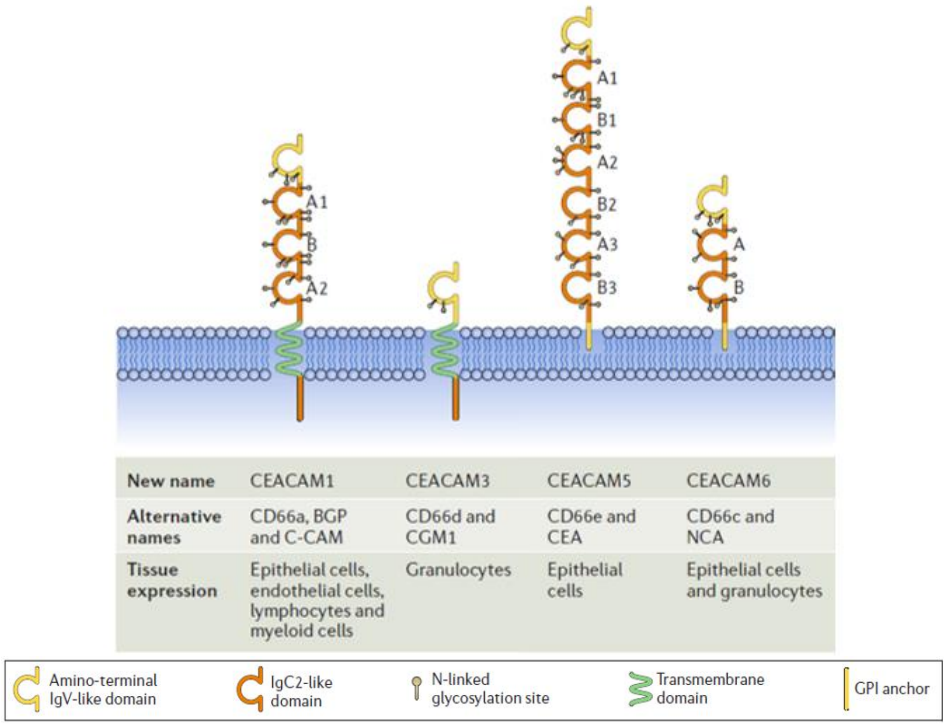


Figure 4. CEACAM structure and expression in humans known to interact with Gc (Modified from 23). Structure consists of none, one or more immunoglobulin constant-region-type-2-like domains and an end immunoglobulin variable-region-like domain. CEACAM5 and 6 do not contain transmembrane domains, while CEACAM1 and -3 have a cytoplasmic tail.

1.6 – Thesis Objective

Gc-host interactions are intricate, where differential expression of proteins is temporally and spatially dependent. In addition, the absence of a representative animal model that contains host-specific receptors to Gc greatly hinders the conclusions that can be made and applied to humans. Furthermore, no vaccine is available for treatment of gonorrhea and with the rise of multi-drug resistant Gc, there is an urgent need to identify conserved factors that can targeted with new vaccines or drugs.

Colonization is the critical first step for establishing long lasting infection, and is a prerequisite both for disease and for the transmission of bacteria from one individual to another. Therefore, I aim to reveal factors that are required for colonization in hopes that they can be targeted with prophylactic and/or therapeutic intervention. Due to the plastic nature of Gc, I postulate that serial passaging of *in vitro*-passaged bacteria within the lower genital tract of female mice will select for bacteria with essential genes required for colonization.

In Chapter 2, I demonstrate that serial passaging in the genital tract of human CEACAM-expressing CEABAC mice using a lab-adapted Gc stain results in the selection for a bacterial phenotype capable of infecting a higher proportion of mice in subsequent infection. These *in vivo* passaged isolates were phenotypically distinct from the parental strain with respect to Opa and pilus expression, and form bacterial aggregates. In Chapter 3, I assessed possible host immune factors potentially driving bacterial selection. My findings show that bacterial survival in the presence of low pH is normal, as is their susceptibility to antimicrobial peptides and the complement system, suggesting that these factors do not explain the increased infection by *in vivo* passaged isolates. Lastly, Chapter 4 describes my work to take advantage of next-generation sequencing to identify genomic mutations that arose throughout *in vivo* passaging, which can be implicated in the increased bacterial fitness observed. The genome-wide screen allowed for the identification of one single gain-of-function mutation present in all *in vivo* passaged isolates that was completely absent in parental and *in vitro* passaged isolates. This mutation is associated with a highly conserved surface epitope shared among clinical samples. Satisfyingly, Gulati and colleagues have recently shown vaccination against this epitope reduces time required for clearance of infection in female wildtype mice[46]. These findings support the validity of my genomic screen and infection model.

Collectively, this work identifies various genomic mutations that can be attributed to the increased proportion of mice initially colonized. With one of our hits currently being tested as a possible vaccine target, there is the potential for other possible vaccine and drug targets present within our screen.

Chapter 2 - Selection of gonococcal isolates with increased bacterial fitness in the lower genital tract of female CEABAC mice

2.1 – Abstract

Female mice administered exogenous hormone and antibiotics are capable of being colonized by Gc for a short period of time. With the availability of a transgenic mouse line expressing human CEACAM3, -5 and -6, I aimed to better understand bacteria-host interactions required to initially establish infection. I hypothesize that serial passaging of a lab-adapted Gc in the lower genital tract of female transgenic mice would select for bacterial phenotypes that provide a survival advantage. Infection with a Gc isolate that had previously been mouse passaged resulted in a higher proportion of mice being colonized, but did not extend the duration of colonization. Differences in growth *in vitro*, protein expression and aggregation were also characterized. My data suggests that *in vivo* passaging does select for bacteria which are phenotypically distinct from the original infecting strain, with the adaptations advantageous for initially establishing infection.

2.2 – Introduction

As previously described, the severity of Gc complications prohibits the use of female human experimental challenge to study infection. Consequently, little is known about how Gc interacts with the host upon the initial stages of infection in women. Due to the many similarities between human and mouse reproductive tracts highlighted above, mice provide a good surrogate for examining such interactions. Female mice undergo cyclic hormonal changes, referred to as the estrus cycle, which is comparable to the female human menstrual cycle. In humans, the menstrual cycle lasts ~28 days, while the mice estrus cycle is 4-5 days in duration[47]. The estrus cycle can be divided into 4 stages: proestrus, estrus, metestrus and diestrus. Each stage can be differentiated by cell morphology and the recruitment of neutrophils, which can be identified by compound light microscopy. Proestrus is characterized by the presence of nucleated, round epithelial cells; estrus by stratified squamous epithelial cells; metestrus by a combination of epithelial cells and few neutrophils; and diestrus signified by the massive influx of neutrophils[47,48]. For Gc infection, estradiol is administered to promote mice entry into the estrus stage, which is more amenable to infection than other stages. While the reason for this increased susceptibility is not certain, it is presumed to occur because the substantial reduction in PMNs within the lower genital tract provides an optimal environment for initial establishment of Gc infection and proliferation[49].

2.3 – Results

2.3.1 – *In vivo* passaging of Gc increases proportion of mice colonized in subsequent infections

Gc populations consist of a panmictic assortment of bacteria expressing different cytoplasmic proteins and surface-exposed antigens. Therefore, I hypothesize that innate immune factors in the lower genital tract of female CEABAC mice will select for bacteria that have a survival advantage. Serial passaging bacteria intravaginally allows of the isolation of phenotypes that are consistently selected for overtime, which will provide insight into what is important for establishing infection. To distinguish advantageous traits specifically involved in increasing infectivity, Gc was also passaged *in vitro*.

Vaginal lavages were conducted on 5-7 week old female CEABAC mice every day for one week leading up to infection to track stages of estrus cycle by cytology. Once in diestrus, mice underwent a hormone and antibiotic regimen which commenced two days prior to infection (Day

-2). Subcutaneous injections of β-estradiol were administered on Day -2, 0 and +2 to prolong the estrus stage of mice. Antibiotics consisted of daily intraperitoneal (IP) injections of vancomycin and streptomycin, with the exception of two injections on Day -1. Moreover, trimethoprim was provided orally in the drinking water. Antibiotics were administered to combat the influx of commensal flora brought on by the β-estradiol injections, ultimately limiting competition for nutrients and space.

For passage 1 (P1), 12 mice were each inoculated with 10^7 bacteria of the parental strain, a laboratory-passaged isolate referred to as N2009 (derived from the clinical isolate MS11), and vaginal washes were performed every two days to recover viable bacteria. On Day 2, only 2 of the 12 mice remained colonized by the bacteria; these were denoted as M1 and M2 (**Figure 5A**). In passage 2 (P2), mice were split into 2 cohorts of 4 and each was infected with either the M1 or M2-derived isolates retrieved from P1. Upon Day 2 of infection, all mice inoculated with the M2 isolate had cleared infection, however, 3 of the 4 mice infected with the M1 isolate had viable bacteria (**Figure 5B**). One mouse infected with the M1 isolate was colonized for a prolonged period of 19 days. Therefore, for passage 3 (P3), a cohort of 8 mice were infected with bacterial isolate M1.05 (12). Notation of the isolate denotes the origin of bacterial lineage, where M1 depicts P1, 05 signifies mouse ear tag from which P2 originated from, and (12) refers to the day the bacterial isolate was recovered. During the P3 infection, 7 out of 8 infected mice had viable bacteria after the second day (**Figure 5C**). Due to the increase in proportion of mice colonized with bacteria in subsequent infections, these findings suggest that bacteria passaged *in vivo* are better adapted for survival during the initial stages of infection.

In order to help distinguish what mutations emerged that contributed to bacterial survival *in vivo*, as opposed to those adaptations that simply increase the rate of growth, the parental strain was also passaged *in vitro*. These bacteria were inoculated into 5 separate culture tubes, each containing nutrient-rich media. Cultures were sub-cultured every 10-12 hours to prevent bacteria from reaching stationary phase, with passaging continuing for 5 consecutive days. Comparison between the *in vivo* vs *in vitro* passaged bacteria allow for the identification of adaptations that confer a fitness advantage solely during the infection process.

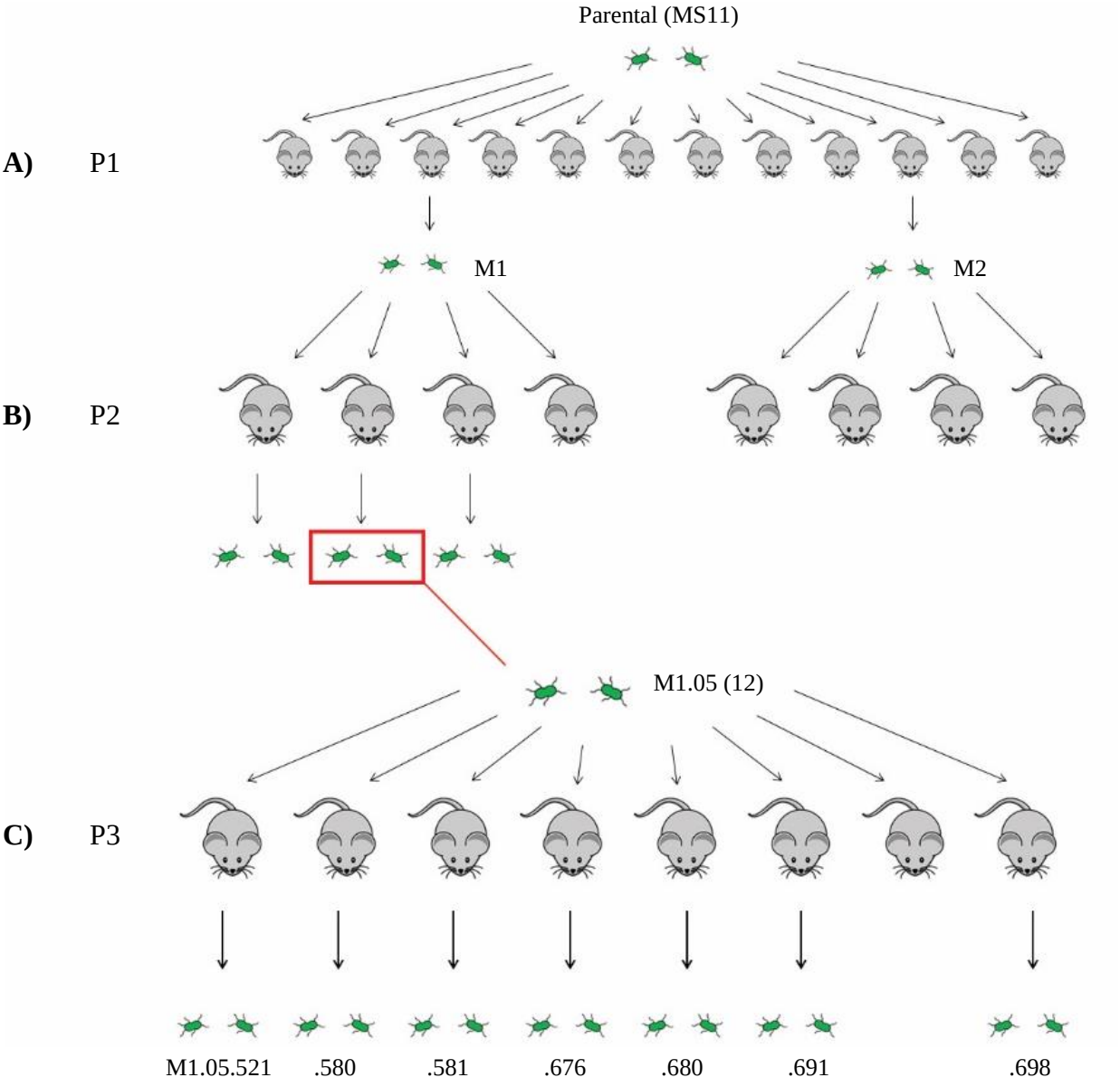


Figure 5. Proportion of mice colonized with each subsequent passage, 2 days post infection. Mice were infected with 10^7 bacteria with the denoted bacterial isolate. **A)** P1; mice infected with lab-adapted Gc strain MS11, where only 2/12 mice infected. **B)** P2; mice were infected with either M1 or M2 isolates from P1. Mice infected with M2 cleared infection, while 3/4 mice inoculated with M1 had viable bacteria. **C)** P3; mice were infected with a P2 isolate, where (12) denotes bacterial isolates recovered from Day 12 of infection. Proportion of mice with recoverable bacterial was 7/8.

2.3.2 – Confirmation of Gc isolation from mice

The lower genital tract of mice is a non-sterile environment where commensal microbes thrive, therefore numerous precautions were taken to ensure that bacteria acquired from *in vivo* passaging was indeed Gc. To combat this, injections of antibiotics were administered to decrease levels of commensal flora and vaginal lavages were plated onto selective plates containing vancomycin,

colistin, nystatin and trimethoprim. Additionally, to confirm the bacteria isolated was Gc, Gram-staining was performed as a preliminary test for bacteria type and morphology. Isolates found to be Gram-negative (pink) and diplococcus (**Figure 6A**) were then subject to colony PCR, with primers specific for the Gc *porA* pseudogene (**Figure 6B**).

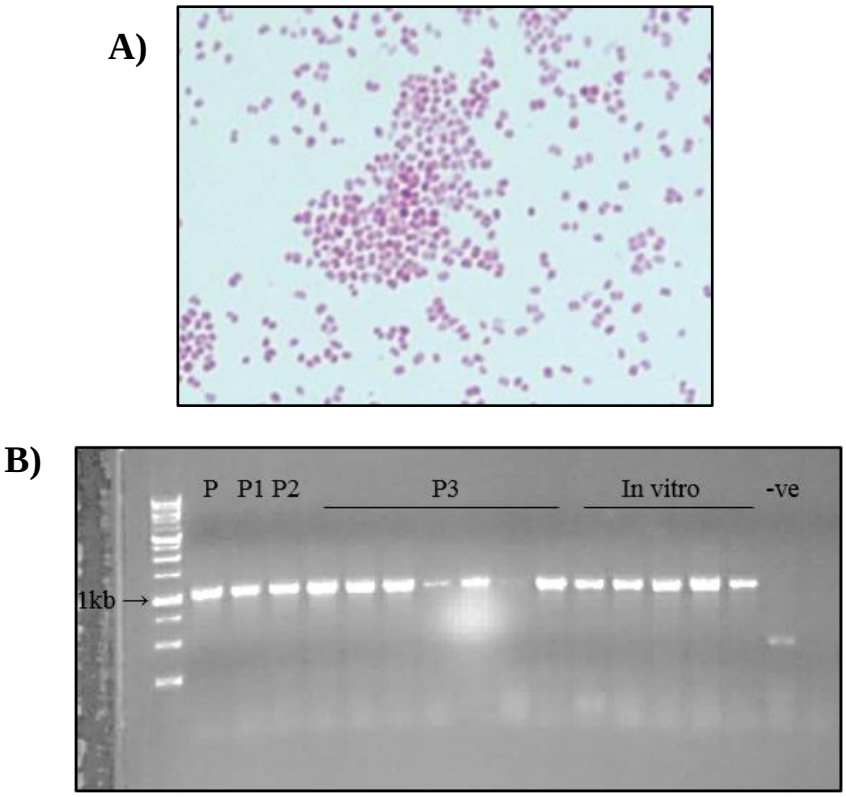


Figure 6. Verification of Gc from *in vivo* bacterial isolates. A) Gram-staining of bacteria recovered on selective plate from vaginal lavage. Bacteria is pink (Gram-negative) and diplococcus in shape consistent with Gc. **B)** Colony PCR of bacterial isolates. Band at ~1kb indicates colonies positive for Gc, where negative control is *E. coli.*

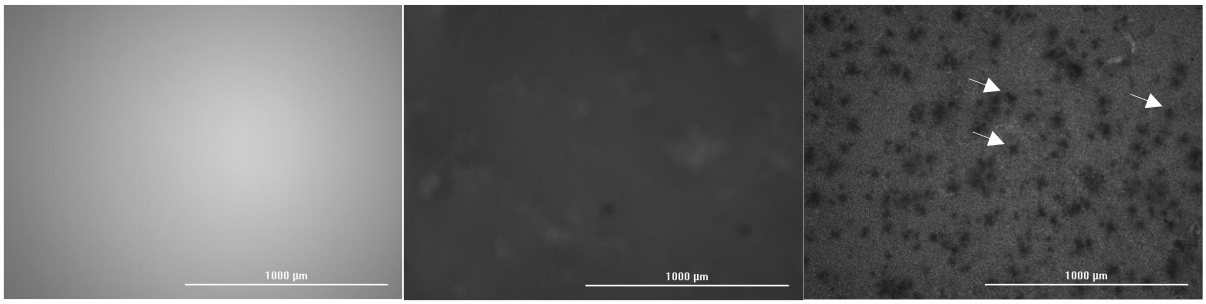
2.3.3 – Functional variation between parental, *in vivo* and *in vitro* passaged isolates

In addition to the drastic increase in the proportion of mice colonized upon infection with previously passaged Gc isolates, one phenotypic difference was observed early in the passaging process. Compared to the parental strain, *in vivo* passaged isolates had the tendency to aggregate, while growth of parental was evenly distributed (**Figure 7A**). Differences in aggregation phenotype observed suggest that *in vivo* passaging selects for Gc with increased intercellular interactions. Consequently, I wanted to further investigate whether other obvious phenotypic differences were present between parental and passaged isolates.

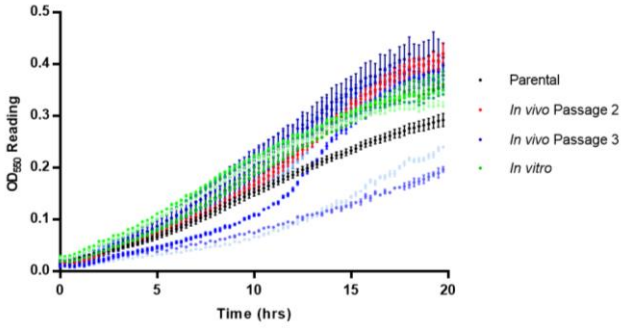
Initially, I sought to determine the presence of any obvious growth differences *in vitro*. Parental, *in vivo*, and *in vitro* passaged isolates were inoculated into brain heart infusion (BHI) supplemented with 1% Isovitalex and allowed to grow at 37˚C over a 20 hour timespan. Optical density (OD_{550}) readings were taken every 15 mins. Consistent with our premise that repeated passage in liquid media would select for adapted variants of the parental strain (which has typically been cultured on solid media), *in vitro* passaged isolates grew slightly better than the original parental strain (**Figure 7B**). Somewhat unexpectedly, the *in vivo* P2 isolate and 5 of 7 P3 isolates grew better than both the parental and *in vitro* passaged isolates. Interestingly, the other two P3 isolates showed defects in growth compared to the parental strain, suggesting that growth enhanced rate alone is not necessary for increased efficiency in mouse infection.

The expression of pilus and Opa are known to be essential during different phases of infection in humans[14,16]. Therefore, I sought to characterize pilus and Opa expression in the passaged isolates. Immunoblot analysis was performed on parental, P2, all isolates recovered from P3, and *in vitro* isolates (**Figure 7C**). The parental strain expressed both pilus and Opa proteins. Fascinatingly, the P2 isolate gained expression of a second Opa variant, while all *in vitro* isolates lost Opa expression. For the *in vivo* re-passaged (P3) isolates, 6 of 7 were positive for the lower molecular weight Opa variant. Conversely, all *in vitro* isolates did not express Opa. In terms of pilus, all *in vivo* isolates appear to have lost pilin expression while, all *in vitro* isolates seem to express a higher molecular weight form.

A)



B)



C)

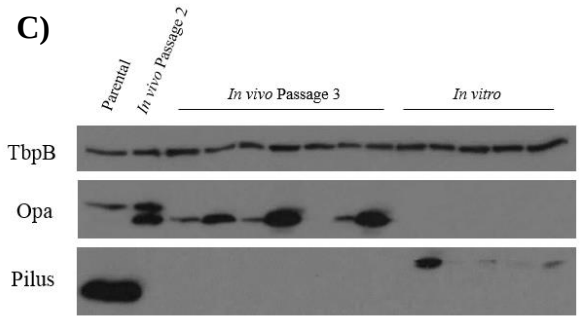


Figure 7. Differences present between parental, *in vivo* and *in vitro* passaged Gc isolates. A) Bacterial aggregation observed after 16 hours of growth in BHI+ISO media (Left = Media only, Middle = parental, right = P3 isolate). Parental growth evenly distributed while P3 growth results in aggregate formation indicated by arrows **B)** Growth curves over a 20 hour time course in BHI+ISO with OD_{550} readings measured every 15 mins **C)** Changes in Opa and Pilus expression of parental, *in vivo* and *in vitro* passaged isolates. Bars indicate mean +/- standard deviation.

2.4 – Discussion

Upon initial infection with the parent strain, clearance of Gc was very rapid in comparison to the studies conducted by Jerse and colleagues. Jerse found wild type BALB/c mice were colonized for an average of 6-10 days. The decrease in bacterial colonization observed could be attributed to the use of different mouse lines. The CEABAC mouse line expresses human CEACAM3, -5 and -6. Of those, CEACAM5 and -6 could aid in bacterial adherence, however the presence of CEACAM3 and -6 on neutrophils can be detrimental to the bacteria because of their ability to increase Gc clearance by neutrophils. Furthermore, the mouse background has been shown to influence susceptibility to Gc infection. Drastic differences have been found when comparing duration of infection and immunological responses between C57BL/6, BALB/c and C3H/HeN mouse backgrounds: mice of the C3H/HeN background are resistant to Gc infection, with no detectable inflammatory response, while C57BL/6 and BALB/c mice are colonized for an average of ~6 days[36]. From an immunological standpoint, infected BALB/c mice had increased levels of

pro-inflammatory cytokines and chemokines, including tumor necrosis factor alpha (TNFα), interleukin-6 (IL-6), macrophage inflammatory protein 2 (MIP-2) and keratinocyte-derived chemokine (KC) in vaginal secretions[36]. These elevations were not observed in C57BL/6 despite the similarities in duration of infection, suggesting that successful infection and this immune response are not necessarily linked. In the scope of my passaging experiments, the CEABAC transgene is on the FVB mouse background. Our group has observed that wild type FVB are less susceptible to Gc than are BALB-c strain, but that the CEABAC transgenic mice display more Gc association with the mucosal epithelia and have a more robust inflammatory response to infection than for the non-transgenic FVB animals (E. Islam, manuscript in revision). While we are convinced that human CEACAM expression increases the relevance of our infection model, it remains difficult to attribute the differences in duration of colonization to any one of these (mouse background, cytokine, CEACAM) factors. Further complicating a comparison between my work and the earlier infection studies done by other groups, it is important to recognize that the genetic variability among Gc and the phase variable nature of their virulence factor expression makes it likely that there are numerous differences between my isolates and that used in the early published work of others, both phylogenetically and due to ongoing passage in laboratory conditions.

Infection with previously passaged bacteria resulted in an increased proportion of mice colonized, however, duration of colonization was only increased in a few mice. In P2, one mouse was infected for 20 days, at which point the mouse was euthanized. Separation of the upper and lower genital tract tissue, followed by homogenization and plating for bacteria suggests that Gc had ascended into the upper genital tract. For the P3 infection, I decided to use bacteria recovered from Day 12 of the P2 infection. I speculated that this isolate was better adapted to establishing infection in the mouse genital tract. Therefore, I would expect mice in P3 to have both a higher frequency and longer duration of colonization. Interestingly, a higher proportion of mice in P3 were infected on the second day, but most mice, 6 of 7, cleared bacteria by Day 4. Consequently, this indicates that bacteria previously passaged increases the proportion of mice infected but does not impact the duration of colonization. One possibility is length of colonization requires other human-specific host factors that are missing in our mouse model, such as human transferrin, lactoferrin, C4BP and fH. These findings suggest a selection for bacteria that express genes which are beneficial for the initial establishment of infection in the lower genital tract of hormone treated mice.

2.5 – Methods

All animal procedures conducted in this study were approved by the Local Animal Care Committee (LACC) at the University of Toronto (Protocol #20011775), which is in compliance with ethical and legal requirements under Ontario's Animals for Research Act and the federal Canadian Council on Animal Care.

Mouse strains

6-8 week old female transgenic FVB mice containing a human chromosome-derived BAC expressing CEACAM3, -5 and -6, were used for all infection experiments unless otherwise specified. Animals were housed in cages with sterile rodent chow, sterile water, enrichment material, and 12 hour light-dark cycles.

Estrus cycle staging

Stages of the estrus cycle were verified by cytology. The vagina of mice between 5-7 weeks old were washed with 30 μL of phosphate-buffered saline (PBS, Life Technologies, Burlington, Canada) and deposited into a 96-well plate. Vaginal smears were examined under a Leica DMIL compound light microscope (Leica Microsystems, Wetzler, Germany) at 10X objective to identify characteristic cell types of the estrus cycle.

Antibiotic and hormone treatment

Mice in the diestrus stage were started on an antibiotic and hormone regimen prior to infection 0.6 mg vancomycin hydrochloride (Bioshop, Burlington, ON, Canada) and 2.4 mg streptomycin sulfate (Sigma-Aldrich, Oakville, Canada) were resuspended in 200 μL of PBS and administered by intraperitoneal (IP) injection daily from 2 days pre-infection (-2) to 5 days post infection (+5), with the exception of two doses provided on Day -1. Trimethoprim (Sigma-Aldrich, Oakville, Canada) was present in the drinking water at 40mg/mL for oral consumption. β-estradiol (Sigma-Aldrich, Oakville, Canada) was prepared as 0.5mg/200μL of PBS and introduced by subcutaneous injection on Days -2, 0 and +2.

Bacterial preparation, infection and recovery

A well-characterized gonococcal strain (MS11) was used for the initial infection; this is referred to as the parental strain. In subsequent infections, bacteria that persisted for the longest duration was used as stated in Section 1.3. Growth conditions consisted of incubation at 37˚C with 5% CO_2 for 16-20 hours. To acclimate bacteria to the antibiotics used for selection during infection,

bacteria was streaked out from a frozen stock onto GC agar (Becton, Dickinson and Company, Sparks, USA) plates supplemented with 1% Isovitalex (ISO, Becton, Dickinson and Company, Sparks, USA) and a vancomycin, colistin, nystatin, and trimethoprim antibiotic mix (VCNT, Becton, Dickinson and Company, Sparks, USA) two days pre-infection. Bacteria was re-streaked onto GC+ISO+VCNT plates on Day -1. On Day 0, bacteria were resuspended in Dulbecco's phosphate-buffered saline (DPBS) with calcium and magnesium (PBS^{++}, Life Technologies, Burlington, Canada). An inoculum of 10^7 bacteria in 2.5 µL PBS^{++} per mouse was prepared and plated to verify accuracy of bacterial preparation. Prior to infection, vaginal washes were conducted using 30 µL of PBS^{++} three times. Excess fluid was removed and 2.5 µL of Gc preparation was deposited into the vagina. Lavages were checked by microscopy to confirm mice were in the estrus stage. Vaginal lavages were performed every two days to recover viable bacteria. 10 µL of PBS^{++} was used to recover bacteria from the vagina, serially diluted, plated onto GC+ISO+VCNT plates and incubated overnight. Bacterial enumeration was conducted after ~20 hrs and bacterial stocks were made in 50% brain heart infusion (BHI, Becton, Dickinson and Company, Sparks, USA) and 25% glycerol (EMD Millipore, Darmstadt, Germany) and stored at -80˚C. Clearance of infection was characterized by 2 consecutive vaginal lavages that tested negative for Gc – at which point mice were humanely euthanized by CO_2 inhalation and serum, and the upper and lower genital tract tissue were collected. Genital tract tissue was homogenized and plated on GC+ISO+VCNT plates to verify the clearance of Gc.

***In vitro* passaging**

Five 15 mL culture tubes containing 5 mL of BHI supplemented with 1% ISO was inoculated with the parental strain with a starting optical density at 550 nm (OD_{550}) of 0.05. Bacterial cultures were placed in a 37˚C shaker at 180 rpm. Every 12 hrs, bacterial stocks were frozen down and also sub-cultured for 5 consecutive days.

Confirmation of Gc

To confirm bacteria recovered from vaginal lavages was Gc, Gram-staining was conducted in combination with PCR analysis. For Gram-staining, a sterile pipette tip was used to select a few colonies retrieved from each mouse and resuspended in 10 µL sterile water on a microscope slide. Slides were air-dried, bacteria were heat killed and stained according to a BD Gram-stain kit (Becton, Dickinson and Company, Sparks, USA). Slides were visualized using Zeiss AxioObserver.Z1 inverted microscope (Carl Zeiss AG, Jena, Germany) at 100X objective with oil immersion for bacterial morphology and colour. Isolates that were Gram-negative and

diplococcus in shape were heat killed for colony PCR. Primers for colony PCR were specific to Gc, targeting the *porA* pseudogene, where presence of a ~1 kb band signified positive Gc colonies.

Comparative growth assay

Gc isolates were inoculated into BHI+1% ISO and the OD_{550} was taken using GENESYS^{TM} 10S UV-Vis Spectrophotometer (ThermoFisher Scientific, Waltham, USA). Isolates were diluted to an OD_{550} of 0.1 and 100 µL of culture was loaded onto a sterile 96 well plate and covered with a plate seal. The 96-well plate was incubated in a TECAN Infinite® 200 multimode reader (TECAN, Männedorf, Switzerland) at 37˚C with OD_{550} readings recorded every 15 mins for 20 hours. Three technical replicates were performed per plate and repeated for at least 3 biological replicates.

Western blot

For each isolate, 10^8 bacteria were resuspended in 100 µL of 2% sodium dodecyl sulfate (SDS) + 10% β-mercaptoethanol (Sigma-Aldrich, Oakville, Canada) and incubated at 95˚C for 5 mins. Samples were loaded onto a 12% polyacrylamide gel and run at 120 V. The gel was transferred onto a polyvinylidene fluoride membrane (GE Healthcare Life Sciences, Mississauga, Canada) and immunoblotted with appropriate primary and secondary antibodies.

Chapter 3 – Effect of selected host immune factors with the potential to driving bacterial selection

3.1 – Abstract

Gc encounters numerous innate immune factors upon inoculation into the lower genital tract. Factors such as low pH, presence of antimicrobial peptides and the complement system have bactericidal activities that could promote the rapid clearance of invading pathogens. In an attempt to elucidate the mechanism driving bacterial selection of *in vivo* passaged isolates, parental, P2 and a pooled P3 bacterial cultures were compared. Bacterial growth at pH 6.0-7.5 were tested and no differences were observed between all isolates. A comparison of bacterial survival in the presence of various concentrations of antimicrobial peptides, LL-37, mCRAMP and mBD1, and 20% naïve mouse serum determined no differences in survival between *in vivo* passaged isolates. Consequently, these findings suggest that low pH, antimicrobial peptides, nor the complement system are responsible for driving bacterial selection.

3.2 – Introduction

Throughout the course of an infection, various bacterial and host factors contribute to the duration of infection and the severity of pathogenesis. From the perspective of the host, the innate immune system is the initial defense to invading pathogens. The innate immune system is non-specific, germ-line encoded, and utilizes a number of mechanisms to combat microbes. In the case of Gc, the typical site of infection in women is the endocervix, where common immune factors include low environmental pH, presence of antimicrobial peptides, host complement system, and shedding of the epithelial cell lining. Despite these typically effective defenses, Gc is capable of thriving in such conditions through various mechanisms that prevent detection and cause a skewing of the immune response so that it does not clear infection.

3.2.1 – Host antimicrobial factors

The innate immune response is crucial for preventing the establishment of a long lasting infection. One mechanism for fending off pathogens is the presence of physical barriers inhibiting host accessibility. For example, the epithelial cell lining limits access of pathogens to deeper, nutrient rich tissue in the host. Additionally, during different stages of the menstrual cycle, constant shedding of the epithelial cell lining eliminates pathogens that have not penetrated into tissue.

Upon initial infection, one major factor that Gc must overcome, is the low pH of the endocervical mucus. The lower genital tract of estradiol-treated female mice is an acidic environment where the average pH is 6.6[36]. In *Salmonella*, the bacterial LPS structure is susceptibility to acid hydrolysis [50,51]. Additionally, presence of acids such as lactic acid and hydrochloric acid increase permeability of the outer membrane and disrupt membrane integrity of bacteria[52].

Another key factor associated with the clearance of invading pathogens in the urogenital tract is the presence of antimicrobial peptides (AMP). AMPs are cationic peptides secreted by epithelial and immune cells and have bactericidal activity. In mammals, AMPs can be divided into 3 classes: defensins, cathelicidins and histatins. The amphipathic nature of defensins and cathelicidins enables their insertion into the membrane of bacteria, thereby disrupting membrane integrity and enabling their lysis[53,54]. In the mouse genital tract, the presence of various AMPs have been identified, including mouse cathelicidin related antimicrobial peptide (mCRAMP)[55] and mouse β-defensin 1 (mBD1)[56] – both capable of contributing to bactericidal activity in the genital tract.

Aside from the directly bactericidal nature of AMPs, the complement system is a major, catalytically-driven component of the innate immune system that is instrumental in pathogen clearance. The complement system consists of three branches: the classical, alternative and lectin pathways. Activation of the complement system occurs through various mechanisms, where one method is the presence of pathogen associated molecular patterns (PAMPs). Upon activation, a cascade of proteolytic reactions is triggered, resulting in numerous negative consequences on bacterial survival. The downstream effects include the release of anaphylatoxins essential in recruiting neutrophils and monocytes, opsonisation of pathogens for quick identification and clearance by phagocytic cells, and the formation of membrane attack complexes (MACs), which disrupt membrane integrity of the bacteria[57]. Due to the potentially devastating consequences associated with inappropriate complement activation, this system is under tight regulation. Factor H and complement component 4b binding protein (C4BP) play major roles in inhibiting downstream activation of the alternative and classical pathways, respectively[18]. The complement system components are present in serum, therefore testing sensitivity of bacteria to serum provides a good indication as to whether complement contributes to the clearance of pathogens. Screening of 29 Gc strains for serum sensitivity found 21 serum resistant strains, where 16 of the 21 were capable of binding C4BP[19]. The remaining 8 strains were serum-sensitive and were unable to bind to C4BP[19]. These findings suggest that Gc is capable of binding C4BP, ultimately inhibiting opsonisation of the bacterial cell surface with C4b.

Alternatively, the clearance of invading pathogens can be further attributed to the recruitment of pro-inflammatory cytokines. Upon detection of pathogens, epithelial cells and other immune cells can secrete pro-inflammatory cytokines and chemokines which recruit neutrophils and leukocytes to the site of infection, thus, aiding in bacterial clearance.

All the factors mentioned above can have drastic consequences on bacterial survival *in vivo*. Therefore, to elucidate why *in vivo* passaged isolates have increased initial fitness, the effects of various innate immune factors were investigated, as detailed below.

3.2.2 – Gonococcal LOS structure

Gram-negative bacteria are enclosed in an outer membrane which is anchored to a thin peptidoglycan layer by lipoproteins. The outer membrane comprises of lipopolysaccharides (LPS), various surface antigens, and glycoproteins where the host immune system targets various components to recognize host cells from pathogen. Most Gram-negative bacteria are decorated in

LPS, which have been associated with immunostimulation, immunomodulation, and preventing hydrophobic compounds from disrupting the inner membrane. Mutants with non-functional LPS in *N. meningitidis* have demonstrated decreased virulence and pathogenicity, further emphasizing the importance of the LPS structure in bacteria[58]. The LPS structure consists of Lipid A, an inner core, outer core, and O-antigen region. Lipid A functions in anchoring LPS to the outer membrane of Gram-negative bacteria and is essential in stimulating the host immune response.

In the case of Gc, the outer membrane structure consists of the anchoring Lipid A, an inner core comprising of 3-Deoxy-D-manno-oct-2-ulosonic acid (KDO) and heptose sugars, and an outer core consisting of various linked sugar moieties (**Figure 13**)[30,59]. The structure is truncated relative to LPS seen in prototypical species such as *E. coli*, since it lacks the terminal repetitive O-antigen and is therefore referred to as a lipooligosaccharide (LOS). As previously mentioned, surface antigens have a crucial role in stimulation and evasion of the host immune response. Consequently, in Gc, *lipooligosaccharide glycosyl transferase A, C, D* and *G* are capable of undergoing phase variation, and therefore, able to manipulate the LOS structure drastically. It has been documented that alterations in the LOS structure of Gc can have dramatic effects on the host immune response. For example, in the absence of a functional LgtD, the enzyme responsible for attachment of the βGalNAc to the end of the α-chain, this allows for Gc to use CMP-NANA in the urogenital tract to sialylate the LOS structure[60-62]. Sialyation of mammalian cells allows immune cells to differentiate host and pathogen. Therefore, with the surface of Gc sialylated, this allows for Gc to evade detection by host immune cells.

3.3 – Results

3.3.1 – Growth of parental, P2 and P3 isolates are impaired in low pH conditions

Parental, P2 and pooled P3 bacterial survival was tested under various conditions that may be encountered during the infection process to identify mechanisms influencing bacterial selection. All seven P3 isolates were streaked out, bacteria enumerated, pooled at equal concentrations, and stocks made to ensure consistency over bacterial survival assays. To simulate low pH conditions faced in the genital tract, BHI was titrated with HCl to pH 7.0, 6.6 and 6.0, and OD_{550} readings of inoculated cultures was taken every 15 mins over a 14 hour time course. No significant growth differences were observed between parental and *in vivo* passaged isolates at normal BHI conditions (pH 7.5, **Figure 8A**), pH 7.0 (**Figure 8B**), and the physiologically relevant pH 6.6 (**Figure 8C**). At the lowest pH tested, pH 6.0, drastic growth defects were observed in all three

isolates compared (**Figure 8D**). These findings suggest that pH is not the determinant that is driving bacterial selection *in vivo*.

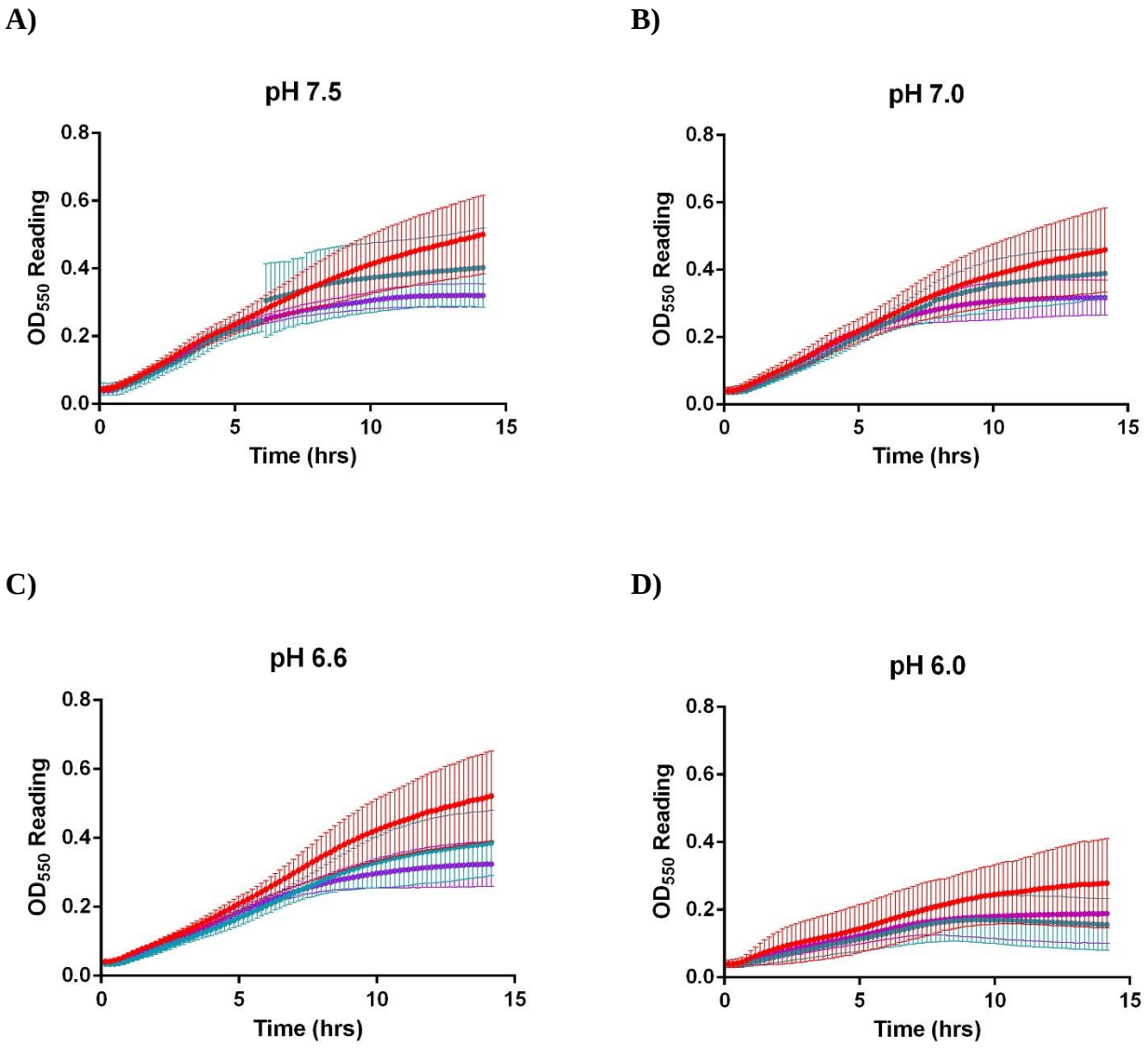


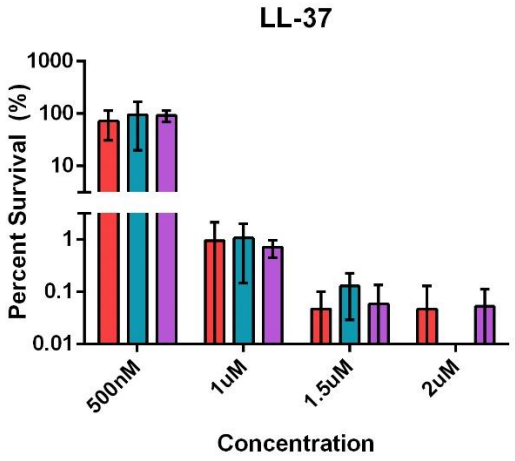
Figure 8. Growth of bacterial isolates in slightly acidic conditions. Parental (red), P2 (green) and pooled P3 (purple) were grown in BHI+ISO titrated to pH 7.5 (**A**), pH 7.0 (**B**), pH 6.6 (**C**) and pH 6.0 (**D**). OD_{550} readings were taken every 15 mins over 14 hours. Bars depict mean +/- standard deviation.

3.3.2. – Presence of antimicrobial peptides LL-37, mCRAMP and mBD1 do not differentially impact bacterial survival

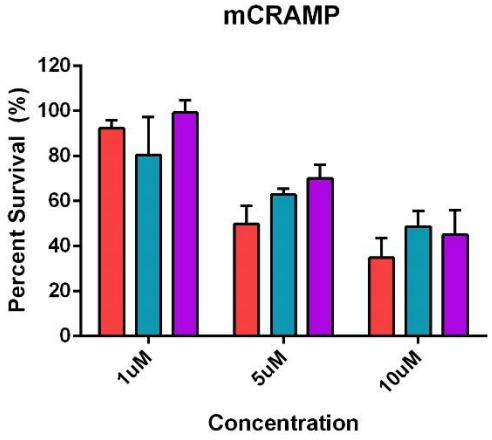
Antimicrobial peptides are directly bactericidal to most bacteria due to their amphipathic nature, which allows them to insert into the bacterial membrane. Comparison of bacterial survival of

parental and *in vivo* passaged isolates after incubation in the human cathelicidin (LL-37), mouse cathelicidin-related antimicrobial peptide (mCRAMP), and the mouse β-defensin 1 (mBD1) were tested at various concentrations. At the lowest tested concentration for LL-37 (500nM), no reduction in bacterial survival was observed, while increasing concentrations of LL-37 resulted in killing of all three isolates tested (**Figure 9A**). Similarly, survival assays were conducted in the presence of mCRAMP, where the lowest concentration test demonstrated no difference in bacterial survival (**Figure 9B**). At 5 and 10 μM concentrations, P2 and P3 appeared to confer a survival advantage over the parental, however these findings were not significant. Upon incubation in 1 and 5 μM mBD1, no variation between parental and passaged isolates were observed (**Figure 9C**). Interestingly, at the highest concentration, bacterial viability of P2 was greatly diminished compared to parental and the pooled P3 isolate. This latter result is opposite to what would be expected to enhance infection of the female genital tract and is not shared by the subsequently passaged strains, so presumably does not explain the increased infection by this isolate.

A)



B)



C)

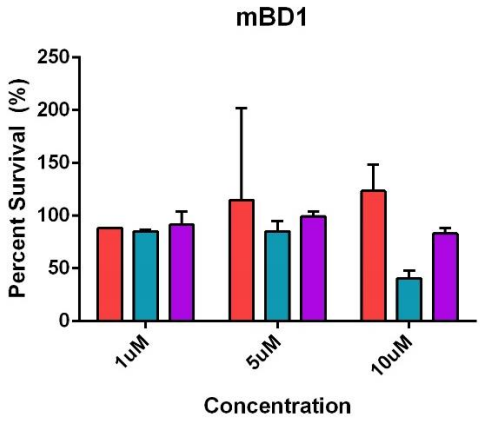


Figure 9. Comparison of survival of bacterial isolates in antimicrobial peptides. Parental (red), P2 (green), and pooled P3 (purple) isolates were incubated in the presence of LL-37 (**A**), mouse cathelicidin-related antimicrobial peptide (**B**) and mouse β-defensin 1 (**C**) at indicated concentrations for 1 hour. Bacteria was enumerated by dilution and colony counts on GC+ISO plates. Percent survival was calculated based on CFU enumeration of untreated bacteria. Bars represent mean +/- standard deviation.

3.3.3. – Parental and *in vivo* passaged isolate have similar rates of survival in the presence of naïve mouse serum

The complement system is an active heat-liable component in mouse serum, where inactivation can be achieved by incubation at 56˚C for 30 minutes. Hence, this allows one to compare bacterial survival in the presence of an active or inactive mouse complement system. $2x10^3$ bacteria were incubated in 20% naïve or heat inactivated mouse serum for 1 hour, at which point Gc was plated

to determine bacterial survival. There was no difference between the survival of parental, P2 and P3 isolate in mouse serum (**Figure 10**). Curiously, the isolates tended to be more susceptible to heat inactivated serum, with this difference becoming significant only in the P3 strains. Given that the complement cascade should not be functional after heat treatment, it is unclear what could be the cause of this bactericidal activity

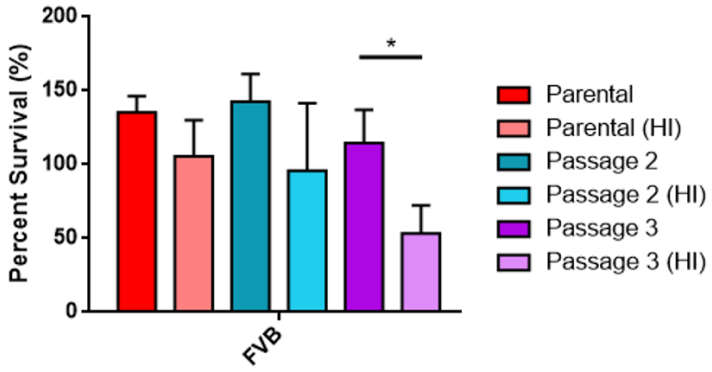


Figure 10. Comparison of survival of bacterial isolates in naïve mouse serum. Parental (red), P2 (blue) and pooled P3 (purple) were incubated in 20% naïve FVB mouse serum for 1 hour. Bacteria was plated on GC+ISO plates, enumerated and percent survival was calculated compared to untreated bacteria. Difference in bacterial survival of P3 serum conditions were observed. Bars signify mean +/- standard deviation, where * indicates $p<0.05$ by two-tail paired t-test. Statistical analysis was performed by Graphpad PRISM 6 software.

Cumulatively, these findings suggest that low pH and the presence of mCRAMP and mBD1 are not driving factors in selecting for the bacterial phenotype observed initially. Although the increased susceptibility of P3 to heat inactivated serum was unexpected, this does not provide insight into innate immune factors responsible for the increased infectivity observed by *in vivo* passaged isolates.

3.4 – Discussion

Simulation of low pH conditions was accomplished by introducing HCl to BHI media prior to sterilization, where pH was monitored before and after autoclaving. Disodium phosphate, one of the components in BHI, is used to buffer the media, and reduce drastic changes in pH. Gc is a facultative anaerobe, therefore in the presence of O_2, Gc generates CO_2 as waste products. Excess CO_2 levels can react with H_2O and produce carbonic acid, thereby contributing to decreased pH levels. After the 14 hour time course, the approximate pH was estimated using pH test strips, where pH levels were observed to decrease slightly. Droge and colleagues demonstrated that the LPS structure of *Salmonella* was sensitive to acid hydrolysis[51]. One of the resultant cleavages was

the ketosidic bond between Lipid A and KDO[51]. In the case of *Neisseria meningitidis*, it has been previously shown that mutants unable to express LOS are still viable, however have reduced growth rates and longer double times[63]. These findings are reflective of the defective growth capacity of isolates grown in media at pH 6.0.

Antimicrobial peptides can be synthesized by epithelial cells constitutively or in the presence of an environmental stimuli, such as a pathogen. Estimated concentrations of peptide in these cells range from 10-100 μg/mL[64]. Conversely, in human neutrophils, antimicrobial peptides are synthesized during leukocyte maturation and is stored in granules[65]. Interestingly, mouse neutrophils do not express defensins[66] however, they do contain mCRAMP[67]. When neutrophils encounter bacteria, they engulf the bacteria into a phagosome that fuses with granules containing high concentrations of antimicrobial peptide. The confined space that peptides are released into allows exposure of bacteria to highly concentrated peptide without the production and secretion of excessive quantities. The concentration in granules are estimated to be >10 mg/mL[68]. In the case of our study, the highest concentration tested was ~40 μg/mL, which is reflective of the median concentration of antimicrobial peptides in epithelial cells. These results suggest that there is no difference in susceptibility to the AMPs tested.

The complement system is composed of heat-liable proteins that when activated, can trigger a cascade resulting in the formation of the MAC. The initial steps of the complement pathway result in the formation of C3 convertases, enzymes responsible for cleavage and activation of C3 into C3a and C3b, the latter of which can covalently link to the bacterial surface[18]. An accumulation of C3b results in the formation of C5 convertases, which are the components that cleave C5 into C5a and C5b[18]. C5b is essential in the formation of MACs. Interestingly, in different mouse backgrounds, some components of the complement system may be absent. For example, the FVB mouse background contains a loss-of-function mutation in C5, which suggests that they are defective in both C5a-mediated inflammation and MAC-dependent killing, while the C57BL/6 background express C5[69]. I have tested the effect of serum from both of these mouse lines and observed no differences in survival. The formation of MACs is one of the possible outcomes of activation of the complement system. Alternatively, the complement system contributes to recruitment of neutrophils and monocytes through the release of C5a and C3a and is predominantly known for its function in assisting in the identification of foreign pathogens through deposition of C3b. Hence, these potential avenues for bacterial selection also have to be further characterized. The unexpected susceptibility of Gc isolates to heat inactivated serum

compared to naïve mouse serum, warrants further investigation to identify the causative bactericidal agent.

Collectively, my findings in this section suggests that selection of the more infectious passaged isolates does not occur due to their increased resistance to acidic pH, the AMPs mCRAMP and mBD1 or the complement system.

3.5 – Methods

Pooling of P3 isolates

The OD_{550} readings of all 7 P3 isolates were measured and diluted in BHI until an OD_{550} of 0.1 was reached. Isolates were diluted, plated, and bacteria enumerated to ensure even distribution of all P3 isolates. Equal volumes of P3 isolates were combined in a single tube and glycerol stock made as per Chapter 2.

Growth in low pH

Titration of BHI to pH 7.5, 7.0, 6.6 and 6.0 was achieved by the addition of HCl (VWR, West Chester, USA) and subsequently autoclaved. The pH was confirmed again before use to confirm that it had not changed during sterilization. Parental, P2 and pooled P3 isolates were prepared as per the growth assay in Chapter 2. OD_{550} readings were taken for 14 hrs on a Cytation 5 cell imaging multi-mode reader (BioTek, Winooski, USA).

Bacterial survival in antimicrobial peptides

Human LL-37, mCRAMP (AnaSpec, Fremont, USA) and mBD1 (Abcam, Cambridge, UK) were diluted to indicated concentrations. $5x10^3$ bacteria was incubated in antimicrobial peptides at various concentrations or only in PBS^{++} for 1 hr, plated and enumerated. Percent survival in antimicrobial peptide was calculated relative to bacteria incubated in PBS^{++}.

Bacterial survival in naïve mouse serum

Female 6-8 week old FVB and C57BL/6 mice were euthanized by CO_2 and cardiac punctures performed to collect blood. Blood was left at room temperature for 1 hr to allow coagulation and centrifuged at 10,000 rpm at 4˚C for 10 mins. Serum was collected, centrifuged at 13,000 rpm for 10 mins, transferred to a new centrifuge tube, and stored on ice. Heat inactivation was

accomplished by incubation of serum at 56˚C for 30 mins. $2x10^3$ bacteria were incubated in 20% naïve mouse, heat inactivated serum or a no serum control for 1 hr at 37˚C. Bacteria were plated, enumerated, and percent survival calculated relative the no serum control.

Chapter 4 – Identification of mutations potentially contributing to increase bacterial fitness

4.1 – Abstract

Affordability and accessibility of sequencing technologies has increased the ease for scientists to sequence and study the genome of various organisms. To elucidate potential mutations responsible for the increased bacterial fitness of *in vivo* passaged isolates, a genome-wide analysis was performed. Standard draft genomes were generated by assembly to a previously annotated and published Gc genome. Draft genomes were multi-aligned, mutations were documented and variants prioritized. A single mutation in the gene *lgtG* was present in 100% of *in vivo* passaged isolates and completely absent in parental and *in vitro* passaged isolates. The mutation lead to a reading frame shift, resulting in the complete transcription of *lgtG* for *in vivo* isolates. Functional LgtG is correlated with presence of an epitope known as 2C7. In clinical samples, the 2C7 epitope is highly conserved. Due to its conservation, it is a currently being tested as a vaccine target. Other groups have demonstrated mice immunized with a peptide mimic recognizing the 2C7 epitope reduces time required for bacterial clearance in mice. Collectively, these findings validate our infection model and screening process, and imply that the other positively-selected antigens may be suitable targets for new vaccines and/or drug therapeutic agents.

4.2 – Introduction

In 1995, the entire genome of *Haemophilus influenzae* was sequenced, assembled and annotated[70]. This event signified a major advancement in sequencing technology as it revealed the first complete bacterial genome. At the time, the sequencing strategy consisted of: sample preparation, Sanger sequencing, assembly, gap closure and annotation. Twenty years later, the steps remain similar with alterations that have decreased the timeline for genome sequencing from years to hours.

Genomic sequencing offers scientists an avenue to broaden their understanding of genes present in the genome and the proteins they encode. However, the low throughput, time-consuming natural, and expensive costs associated with Sanger sequencing made it impractical for whole genome sequencing in basic research laboratories. With the development of next-generation sequencing (NGS) platforms, such as Illumina, millions of DNA fragments can be sequenced at once. The decrease in time required for each run is a result of pooling multiple samples and sequencing smaller DNA fragments, where Illumina sequencers generate millions of different reads of between 36-300 bases[71]. Comparatively, Sanger sequencing has read lengths ~500 bases, but each reaction reveals a single sequence[72]. Therefore, with the massive influx of raw sequence reads, there is a heavy reliance on better assembly programs. Although NGS is an incredibly useful tool, a disadvantage of smaller sized fragments is the generation of numerous disjointed contigs within a single genome during *de novo* assembly[73]. Unfortunately, the time and expense associated with gap closure leaves most genomes as draft sequences[74]. Therefore, third-generation sequencing platforms are aimed at sequencing longer fragments (>1,000 bases) to minimize the number of gaps between contigs[75]. Moreover, a single sequencing run generates a vast amount of data. Hence, another major focus of the field is on the development of better computer hardware and software capable of handling data storage, processing and analysis.

At the moment, there are 5157 complete, and over 37,000 permanent draft bacterial genome sequences available on the integrated microbial genomes (IMG) database[76]. Advances in sequencing technology in the past decade have provided scientists with an improved and more affordable platform for answering questions that could otherwise take years. For example, with whole-genome sequencing, scientists can gain insight into genes responsible for the pathogenic or non-pathogenic nature of different bacterial strains of the same species. This allows for the identification of possible genes essential in pathogenesis. In this chapter, standard draft sequences

of the parental, *in vivo*-passaged, and *in vitro*-passaged isolates were generated to identify genomic mutations that can be attributed to the increased bacterial colonization rate observed in Chapter 2.

4.3 – Results

4.3.1 – Identification of mutations present in passaged isolates

Genomic DNA was obtained from the parental, P2, P3 and all *in vitro* passaged isolates and prepared for whole-genome sequencing by Illumina MiSeq (2x150 nt). Upon retrieval of raw sequence reads, quality of reads were checked by FastQC[77]. The average number of total reads per strain was 846 525. The quality score ranged from 37-38 and GC content was 51-53% (**Table 1**). All strains contained failure notices for the per base sequence content parameter, where nucleotide biases were observed within the first 10 bases (**Figure 11A**). These results are consistent with fragmentation biases associated with Nextera XT library preparation. Additionally, all sequenced isolates contained warnings in Kmer content within the first 9 bases (**Figure 11B**), also reflective of biases during the tagmentation process of library generation.

Table 1. Bacterial strains submitted for Illumina sequencing with FastQC basic statistics

Type of passage	Strain name	Total sequences	Average quality score	%GC Content
Parental	N2009	422 995	38	51
In vivo passage 2	M1.05(12)	771 410	38	52
In vivo passage 3	M1.05.521	969 741	37	52
	M1.05.580	950 679	38	52
	M1.05.581	774 954	38	52
	M1.05.676	475 944	38	52
	M1.05.680	816 897	38	52
	M1.05.691	921 338	37	52
	M1.05.698	914 834	37	53
In vitro passage	T1D5	827 750	37	52
	T2D5	915 066	37	52
	T3D5	852 147	37	52
	T4D5	1 317 419	37	52
	T5D5	920 187	37	53

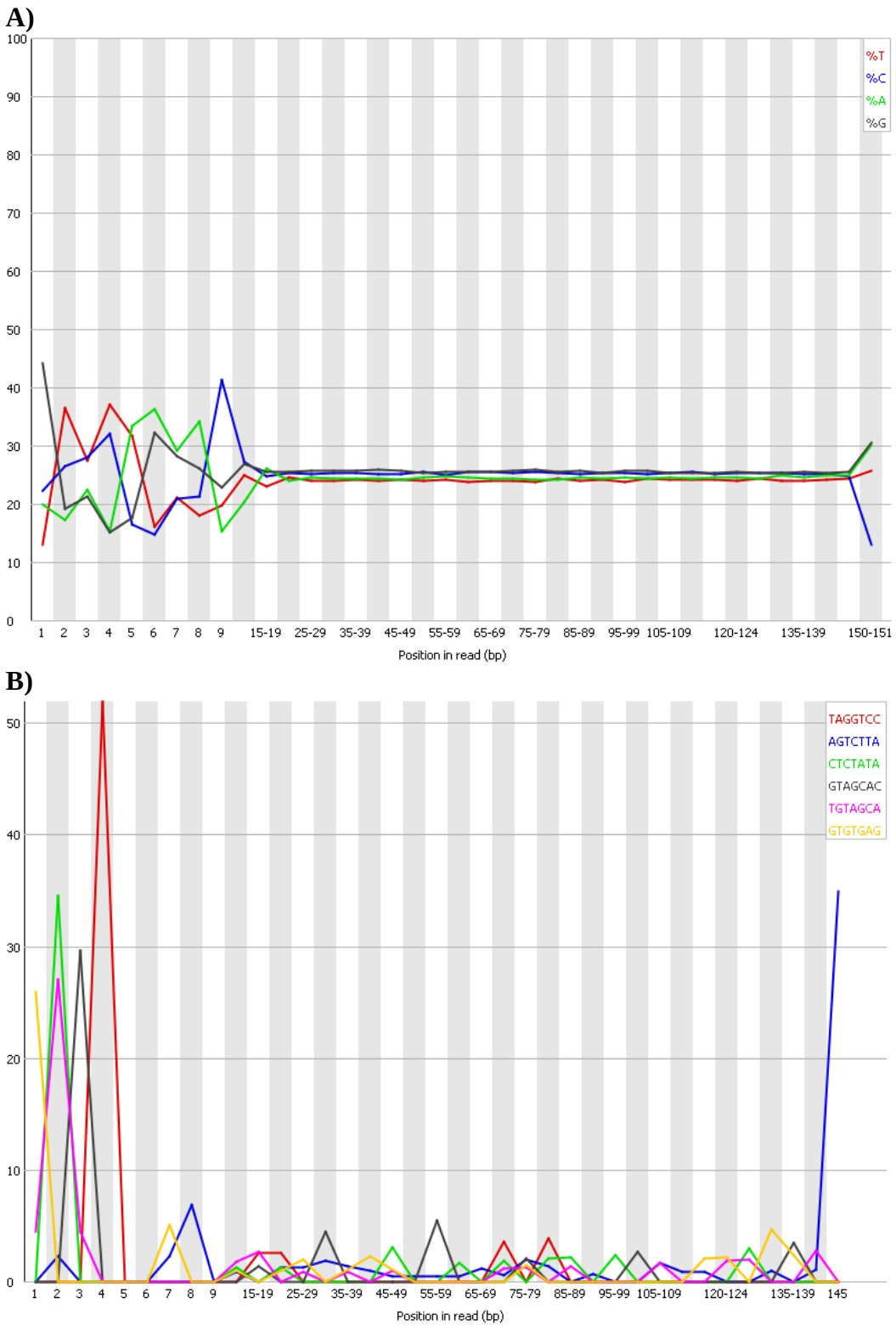


Figure 11. Failures and warnings from FastQC quality check. Raw sequence reads were imported into FastQC to check for quality. All sequences contained failures in base sequence content **(A)** and warnings in Kmer content **(B)**, due to biases present in the first 10 bases of reads, which is due to inherent fragmentation biases in library preparation.

Geneious 9.0.2 was used to trim low quality sequences, pair reads, assemble and multi-align genomic sequences. Parental and passaged isolate genomes were assembled to the published and annotated Gc MS11 genome from NCBI using a Bowtie2 plugin. Multi-genome alignment was performed by Geneious alignment, MUSCLE and ClustalW. A list of mutations present in three or more passaged isolates but completely absent in the parental was generated using the Geneious alignment and presence of mutations verified on MUSCLE and ClustalW alignments programs (**Figure 12**).

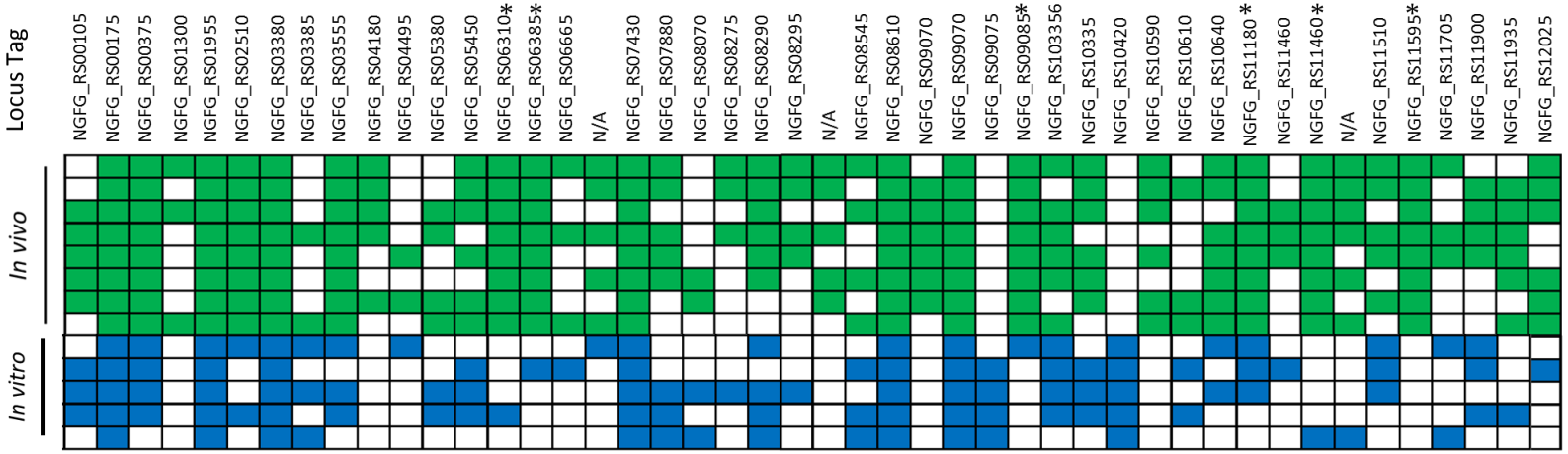


Figure 12. Identification of genomic mutations present in passaged isolates. Whole-genome sequencing of Parental, P2, P3 and *in vitro* passaged isolates was conducted by Illumina MiSeq. Draft genomes were generated by assembly to the reference MS11 Gc genome available from NCBI. Genomes were aligned and mutations present in >3 passaged isolates but absent in parental were documented. Boxes in blue (*in vitro*) and green (*in vivo*) contain a mutation not present in parental at indicated locus. Mutations of interest are denoted with *.

4.3.2 – Prioritization of mutations of interest

The purpose of this study was to identify mutations that were selected for *in vivo*, under the premise that these could be contributing to the increased bacterial infectivity. Therefore, while analyzing the list of mutations (**Figure 12**), those present in most *in vivo* isolates but absent or of low prevalence *in vitro* were further assessed for potential downstream effects. Six loci (NGFG_RS02510, NGFG_RS06310, NGFG_RS06385, NGFG_RS09085, NGFG_RS11460, and NGFG_RS11595) contained mutations in all *in vivo* passaged isolates with a low prevalence *in vitro*, and were therefore deemed high in priority. At NGFG_RS02510, NGFG_RS06385, and NGFG_RS09085, the mutation present consisted of a mixed consensus call, suggesting the presence of two populations of bacteria. For example, at NGFG_RS06385, the consensus base call of *in vivo* passaged isolates is A and C, while the base call of the parental strain is A. This could suggest a mixed bacterial population, where one population contained a single nucleotide polymorphism (SNP) encoding for a single amino acid variant. Changes in amino acid sequence could result in a change in protein folding and consequently alterations to the binding site. Therefore protein sequences were imported in PredictProtein, an open online resource, capable of predicting binding motifs. Interestingly, the SNP in NGFG_RS06385 was present near a predicted binding motif of the protein TrkA. The sequencing data reported a consensus call of a mixed population, where all *in vivo* passaged isolates expressed either threonine (same as parental) or a proline. Threonine is a polar amino acid, whereas proline is hydrophobic, therefore it is possible this amino acid change can have drastic consequences on protein function.

Mutations NGFG_RS06310 and NGFG_RS11460 were consistent with small deletions in the nucleotide sequence. These deletions could potentially be attributed to phenotypically-selected errors introduced during DNA replication or *in silico* errors in genome alignment. Validation of mutations by Sanger sequencing would be needed to validate the presence of these mutations.

Interestingly, a mutation at NGFG_RS11595 was present in all *in vivo* passaged isolates but completely absent in all *in vitro* passaged isolates. The mutation was associated with the addition of a cytidine in a homopolymeric tract in the coding sequence. The parental and *in vitro* passaged isolates all maintained a consensus call of 10 Cs, however, all *in vivo* passaged isolates contained a consensus call of 11 Cs. The insertion resulted in a frameshift, allowing for the complete translation of the gene. Conversely, the parental and *in vitro* isolates had a premature stop codon. The sequence coverage of the insertion region was at least 24x coverage and analysis of the raw sequence reads indicated that ~45-70% of reads present in *in vivo* isolates contained 11 Cs.

BLAST search of the genomic sequence corresponded with the *lipooligosaccharide glycosyl transferase G* (*lgtG*) gene. To validate the nucleotide insertion, genomic DNA was extracted from individual colonies of parental or *in vivo* passaged isolates and the region was submitted for Sanger sequencing. Sanger sequencing results were consistent with whole-genome sequencing results, where 11 Cs were present *in vivo* passaged and 10 Cs in parental. Moreover, immunoblot analysis for the 2C7 epitope, which is generated in the presence of functional LgtG, further verified sequencing results (**Figure 13**).

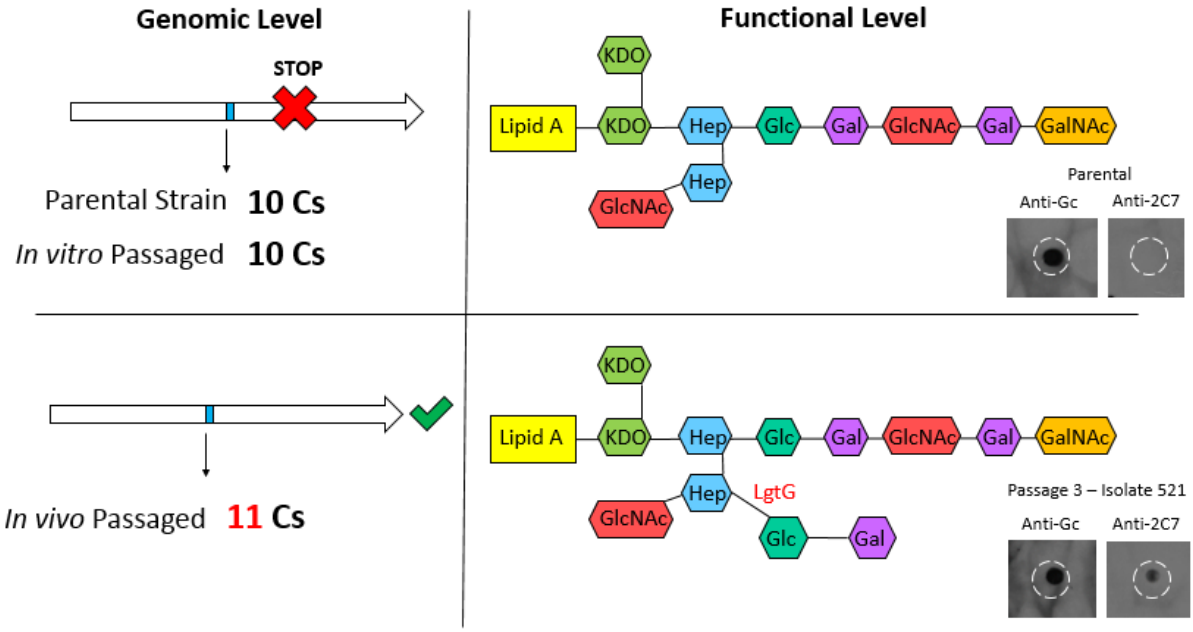


Figure 13. Genomic and functional implications of mutation at NGFG_RS11595 (*lgtG*). *In vivo* passaged isolates encode additional cytidine resulting in functional LgtG and presence of Glc-Gal (β-chain) attachment to HepII of Gc LOS structure. β-chain attachment was verified by an immunoblot assay for the 2C7 epitope.

4.4 – Discussion

In this section, I took advantage of high throughput sequencing to reveal and compare mutations acquired during *in* vivo and *in vitro* adaptations of Gc. Phase and antigenic variation contributes to the plasticity of Gc, enabling the bacteria to undergo constant adaptation so that the population can survive during environmental selective pressures. Consequently, one single colony would not be representative of the entire bacterial population. Therefore, for the infection and sequencing processes, whole bacterial populations were used as opposed to single colonies. As such, raw sequence reads were analysed when consensus calls were variable. This allows for the

identification of a mixture of genotypes present in the population and indicates that further validation is required for each mutation loci. To validate mutations of interest, bacterial isolates were streaked out onto GC+ISO+VCNT plates and single colonies re-streaked, stored and submitted for Sanger sequencing for each locus of interest. If Sanger sequencing results corresponded to WGS results, mutations were considered real and further analysis warranted.

Sequence quality analysis using FastQC resulted in a number of warnings in base sequence calls and Kmer content within the first 10 bases of the raw reads. This result is expected when using Nextera XT library preparation kits. During library preparation, DNA fragments are sheared into smaller fragments and adapters are attached to both ends of each fragment through a process known as tagmentation. Nextera XT library preparation kit utilizes a transposase that cleaves at a preferred sequence, ultimately introducing low-level biases[78-80]. Therefore, reads have a base sequence bias found in the first 10 nucleotides as indicated in **Figure 11**.

One method of prioritizing mutations of interests for single amino acid variants was through the usage of protein prediction software to approximate the location of protein binding motifs. Although there has been improvements to the accuracy of protein prediction resources in recent years, accuracy of predictions is dependent on sequence similarity between query and template sequences[81]. Therefore, usage of PredictProtein was limited to narrowing-down mutations that had already been characterized as high interest.

The most obvious mutation found in the screen was a SNP at the locus that encodes for LgtG, since it was expressed in every *in vivo* passaged isolate and not expressed in the parental or *in vitro* passaged isolates, and since its function had been previously assigned. The average coverage over the mutation was 24x, which is adequate for the identification of SNPs[82]. Further validation by Sanger sequencing was conducted to ensure the mutation present on the screen is valid in the bacterial population. LgtG is the enzyme responsible for the attachment of the β-chain onto the LOS structure of Gc[59]. Consequently, expression of the β-chain results in the presence of the 2C7 epitope[83], which is conserved in 95% of clinical samples[84]. Hence, Gulati and colleagues are currently investigating the role of immunization against 2C7 on Gc infection. Their findings suggest that immunizing with a peptide mimic of the 2C7 epitope enables mice to clear Gc infection more rapidly than do unvaccinated mice[46]. As such, these results further validate our infection model, selection process and screening method, suggesting that the other positively-

selected mutations are worth exploring further since they may also represent effective vaccine and drug targets.

4.5 – Methods

Genomic extraction

Parental, *in vivo* and *in vitro* passaged isolates were resuspended in 200µL of PBS++ and genomic DNA was extracted by a QIAGEN DNeasy Blood and Tissue Kit (QIAGEN, Hilden, Germany). Approximate quality and quantity was checked by a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, USA). Quant-iT^{TM} PicoGreen® dsDNA Assay Kit (Life Technologies, Oregon, USA) was used to accurately quantify amount of dsDNA from genomic extractions prior to library preparations.

Library preparation

Genomic DNA was prepared for whole genome sequencing by a Nextera XT library preparation kit (Illumina, San Diego, USA). Libraries were run on a 1% agarose gel to ensure libraries were of the correct length prior to pooling. Pooled library sample was submitted to the Donnelly Sequencing Centre for sequencing by Illumina MiSeq.

Quality check, assembly and alignment

Raw sequence reads were imported into FastQC to assess quality and potential problems with the dataset. All sequences were imported into Geneious 9.0.2, where sequence reads were trimmed and pair-ended. Sequences were assembled to the reference Gc MS11 genome published and annotated on the NCBI database using the Bowtie2 plugin with default settings. Multi-alignment of references, parental and passaged isolates were performed by Geneious Alignment, ClustalW and MUSCLE according to manufacturer settings. Sequence disagreements present in 3 or more passaged isolates but absent in parental were identified using the Geneious alignment file and mutations were verified by ClustalW and MUSCLE alignments.

Changes in amino acid sequence

Proteins encoded by NGFG_RS02510, NGFG_RS06385 and NGFG_RS09085 were imported into PredictProtein to predict potential binding site.

Chapter 5 – Summary and Future Directions

5.1 – Summary

Gc is a human-restricted pathogen that can cause devastating complications in women if left untreated. With the absence of an available vaccine and the rapid development of resistance to most relevant antibiotics, Gc has become a pressing global health concern. Consequently, there is a necessity to identify new vaccine and drug targets focused on preventing initial colonization and prolonged infection. The availability of a transgenic mouse line expressing human CEACAM3, -5 and -6 provides a good model for investigating the Gc lifestyle and host response during mucosal infection.

Intravaginal serial passaging of a lab-adapted Gc strain resulted in an increased proportion of mice colonized in subsequent infections. The *in vivo*-passaged isolates were phenotypically distinct from the original infecting strain in Opa and pilus expression, and in growth rate in nutrient-rich media. In this work, I have eliminated low pH, various tested concentrations of the AMPs LL-37, mCRAMP and mBD1, and naïve mouse serum as possible host immune factors driving the bacterial phenotype observed, at least when considered in isolation. Therefore, a genome-wide sequencing approach was performed to identify mutations that can be implicated in the increased colonization rate. A single mutation was present in 100% of *in vivo* passaged isolates but completely absent in all parental and *in vitro* passaged isolates. This mutation was associated with the *lipooligosaccharide glycosyl transferase G* gene, where mutations resulted in the complete translation of *lgtG* for *in vivo* passaged isolates. LgtG is responsible for the attachment of the β-chain on the LOS structure and is highly conserved in clinical isolates. As such, other groups have demonstrated that immunization against this epitope reduces time required for bacterial clearance in mice. When considered together, these findings imply that the LgtG modification is an important determinant of the infectivity of Gc.

Moreover, these findings support the validity of our infection model and screening approach, suggesting that other positively-selected factors may be equally important for Gc infection. This makes them biologically interesting and suggests that they might provide new avenues by which to prevent and/or treat gonococcal disease.

5.2 – Future Directions

5.2.1 – Validation of high priority hits

The genome-wide screen identified numerous mutations absent in the parental strain but present in most *in vivo* passaged isolates. While the most striking results was the complete switch in LgtG expression off (*in vitro*) to on (*in vivo*), other loci are also enticing. For example, certain mutations were consistently on *in vivo* but either on or off *in vitro*, suggesting that they may be useful to establish infection and less of a fitness drain during *in vitro* growth. Also, there were mutations that were on *in vitro* but off *in vivo*, suggesting that the phenotype might be helpful for unimpeded growth but imped initial colonization. In each case, it will be interesting to explore whether a targeted mutation in each single gene or a combination of genes could be attributed to the increased bacterial fitness observed. To accomplish this, site-directed mutagenesis should introduce point mutations in the original parental strain so that the specific effect of these mutations on the bacterial phenotype can be tested. These mutants should be used to challenge female CEABAC mice to test ability to colonize.

5.2.2 – Elucidation of mechanism driving bacterial selection

Although various possible innate immune factors were eliminated as driving selection in Chapter 3, many other possible mechanisms remain to be investigated. One aspect not examined in this work is the contributions of neutrophils on infection. As previously mentioned, neutrophils recognize invading pathogens, phagocytose and destroy most microbes through their characteristic oxidative burst and degranulation of bactericidal enzymes and peptides. Mice used in this study express CEACAM3 on neutrophils, which allows them to more effectively engulf Gc and triggers a potent bactericidal response. Isolation of bone marrow-derived neutrophils from CEABAC mice can allow for the assessment of differences in opsonin-independent phagocytosis and killing amongst isolates which would reflect their relative susceptibility to neutrophils encountered during infection. Testing bacterial survival of Gc isolates after incubation with neutrophils would provide insight into bactericidal effects possibly faced during infection. If neutrophils are driving bacterial selection, the parental isolate should have diminished bacterial numbers compared to *in vivo* passaged isolates.

Another avenue of the host immune response that could impact Gc phenotypic selection is variations in the levels of pro-inflammatory cytokines and chemokines recruited. To explore this, serum collected from mice originally infected with parental and passaged isolates should be

analyzed for differences in cytokine and chemokine profiles by enzyme linked immunosorbent assay. Mice infected with the parental isolate should express higher levels of pro-inflammatory cytokines and chemokines compared to *in vivo* passaged isolates if this mechanism is suspected of increasing bacterial infectivity.

As described in Chapter 2, I observed increased bacterial aggregation for *in vivo* passaged isolates compared to the parental. This suggests possible changes in surface exposed structures that promote intercellular interactions. Consequently, adherence of parental and passaged isolates to epithelial cells in the lower genital tract of CEABAC mice should be examined. To accomplish this, CEABAC mice should be infected with the parental or *in vivo* isolates, euthanized at various time points and immunofluorescence imaging performed to identify localization of bacteria.

5.2.3 – Comparison of transcriptomic and genomic hits of interest

Whole-genome sequencing is a great resource for determining where variability occurs, which allows one to narrow focus on what might contribute to changes in bacterial fitness. However, WGS does not provide insight into variations in the level of gene expression. Hence, RNASeq of isolates in combination with WGS results can aid in elucidating key factors responsible for the heightened infection phenotype.

5.2.4 – Identification of essential genes required for infection

The results of this study help to identify possible genomic mutations that are selected for during infection. This suggests that these mutations provide a survival advantage for bacteria, however it does not indicate that these mutations are absolutely necessary for establishing infection. To better elucidate which genes are essential for initial colonization, a transposon mutant library could be generated from one of the *in vivo* passaged isolates. Subsequent inoculation into the lower genital tract of CEABAC mice, bacterial recovery and sequencing could ascertain genes essential for infection.

5.2.5 – Identification of common adaptations and previously unidentified mutations contributing to infectivity

The overall goal of this study is to identify bacterial factors that are important for establishing infection. Therefore, comparisons between *in vivo* and *in vitro* passaged isolates allowed for the prioritization of mutations of interest. One caveat of this approach is the potential for *in vitro* passaged isolates to already have genomic mutations that are beneficial for infection but are not selected against *in vitro*. If these mutations do not provide a growth disadvantage *in vitro*, it can

be maintained within the genome, preventing its identification. Therefore, repeating this study using clinical isolates that are first passaged *in vitro* can reveal new factors not identified here. Additionally, the presence of common adaptations present in this study would further emphasize the importance these mutations have on bacterial fitness.

5.2.6 – Investigation of the role other human-specific innate immune factors have on infection

The presence of Gc-host specific factors has been emphasized throughout this thesis. In my study, I utilized a transgenic mouse line expressing human CEACAM3, -5 and -6, however, there are other human factors that can contribute to the maintenance of Gc infection. For example, Gc have been documented to bind inhibitors of the complement system such as fH and C4BP[19,20]. However, Gc is unable to recognize mouse fH and C4BP, therefore the availability of transgenic mouse lines expressing human fH or C4BP would contribute to our understanding of the role that the complement system plays on infection. Moreover, iron acquisition is a major factor limiting bacterial persistence. Hence, mice expressing human transferrin or lactoferrin, which the Gc can use as iron sources within human tissues[22], would allow us to compare the contribution iron to infection.

5.3 – Conclusion

This work utilizes serial passaging of Gc in a vaginal infection model in female mice expressing human CEACAM3, -5 and -6 to select for bacteria better capable of establishing initial colonization. Through the use of next-generation sequencing, I identify a list of possible genomic mutations, where one hit is currently being investigated as a promising vaccine target. As such, this approach represents a powerful approach to reveal and compare the adaptation of the genetically plastic Gc within their vaginal or other infectious niches.

References

1 WHO. Global action plan to control the spread and impact of antimicrobial resistance in *Neisseria gonorrhoeae*. (2012).

2 Edwards, J. L. & Apicella, M. A. The molecular mechanisms used by *Neisseria gonorrhoeae* to initiate infection differ between men and women. *Clin Microbiol Rev* **17**, 965-981, table of contents, doi:10.1128/CMR.17.4.965-981.2004 (2004).

3 Harrison, W. O. *et al.* A trial of minocycline given after exposure to prevent gonorrhea. *N Engl J Med* **300**, 1074-1078, doi:10.1056/NEJM197905103001903 (1979).

4 Marrazzo, J. M., Handsfield, H. H. & Sparling, P. F. in *Principles and practice of Infectious Diseases* (eds G.L. Mandell, J.E. Bennett, & R Dolin) 2753-2770 (Churchill Livingstone, 2010).

5 Fussenegger, M., Rudel, T., Barten, R., Ryll, R. & Meyer, T. F. Transformation competence and type-4 pilus biogenesis in *Neisseria gonorrhoeae*--a review. *Gene* **192**, 125-134 (1997).

6 Treangen, T. J., Ambur, O. H., Tonjum, T. & Rocha, E. P. The impact of the neisserial DNA uptake sequences on genome evolution and stability. *Genome Biol* **9**, R60, doi:10.1186/gb-2008-9-3-r60 (2008).

7 Unemo, M. & Shafer, W. M. Antibiotic resistance in *Neisseria gonorrhoeae*: origin, evolution, and lessons learned for the future. *Ann N Y Acad Sci* **1230**, E19-28, doi:10.1111/j.1749-6632.2011.06215.x (2011).

8 Ohnishi, M. *et al.* Ceftriaxone-resistant *Neisseria gonorrhoeae*, Japan. *Emerg Infect Dis* **17**, 148-149, doi:10.3201/eid1701.100397 (2011).

9 Unemo, M. *et al.* High-level cefixime- and ceftriaxone-resistant *Neisseria gonorrhoeae* in France: novel penA mosaic allele in a successful international clone causes treatment failure. *Antimicrob Agents Chemother* **56**, 1273-1280, doi:10.1128/AAC.05760-11 (2012).

10 Golparian, D., Hellmark, B., Fredlund, H. & Unemo, M. Emergence, spread and characteristics of *Neisseria gonorrhoeae* isolates with in vitro decreased susceptibility and resistance to extended-spectrum cephalosporins in Sweden. *Sex Transm Infect* **86**, 454-460, doi:10.1136/sti.2010.045377 (2010).

11 Prevention, C. f. D. C. a. (ed Department of Health and Human Services) (Atlanta: U.S, 2014).

12 Wang, J., Gray-Owen, S. D., Knorre, A., Meyer, T. F. & Dehio, C. Opa binding to cellular CD66 receptors mediates the transcellular traversal of *Neisseria gonorrhoeae* across polarized T84 epithelial cell monolayers. *Mol Microbiol* **30**, 657-671 (1998).

13 McGee, Z. A., Stephens, D. S., Hoffman, L. H., Schlech, W. F., 3rd & Horn, R. G. Mechanisms of mucosal invasion by pathogenic *Neisseria*. *Rev Infect Dis* **5 Suppl 4**, S708-714 (1983).

14 Jerse, A. E. *et al.* Multiple gonococcal opacity proteins are expressed during experimental urethral infection in the male. *J Exp Med* **179**, 911-920 (1994).

15 James, J. F. & Swanson, J. Studies on gonococcus infection. XIII. Occurrence of color/opacity colonial variants in clinical cultures. *Infect Immun* **19**, 332-340 (1978).

16 Seifert, H. S., Wright, C. J., Jerse, A. E., Cohen, M. S. & Cannon, J. G. Multiple gonococcal pilin antigenic variants are produced during experimental human infections. *J Clin Invest* **93**, 2744-2749, doi:10.1172/JCI117290 (1994).

17 Cannon, J. G. *et al.* in *10th International Pathogenic Neisseria Conference.*

18 Zipfel, P. F. & Skerka, C. Complement regulators and inhibitory proteins. *Nat Rev Immunol* **9**, 729-740, doi:10.1038/nri2620 (2009).

19 Ram, S. *et al.* Binding of C4b-binding protein to porin: a molecular mechanism of serum resistance of *Neisseria gonorrhoeae*. *J Exp Med* **193**, 281-295 (2001).

20 Ram, S. *et al.* Binding of complement factor H to loop 5 of porin protein 1A: a molecular mechanism of serum resistance of nonsialylated *Neisseria gonorrhoeae*. *J Exp Med* **188**, 671-680 (1998).
21 Apicella, M. A. *et al.* Modification by sialic acid of *Neisseria gonorrhoeae* lipooligosaccharide epitope expression in human urethral exudates: an immunoelectron microscopic analysis. *J Infect Dis* **162**, 506-512 (1990).
22 Schryvers, A. B. & Stojiljkovic, I. Iron acquisition systems in the pathogenic *Neisseria*. *Mol Microbiol* **32**, 1117-1123 (1999).
23 Gray-Owen, S. D. & Blumberg, R. S. CEACAM1: contact-dependent control of immunity. *Nat Rev Immunol* **6**, 433-446, doi:10.1038/nri1864 (2006).
24 Hamrick, T. S., Dempsey, J. A., Cohen, M. S. & Cannon, J. G. Antigenic variation of gonococcal pilin expression *in vivo*: analysis of the strain FA1090 pilin repertoire and identification of the *pilS* gene copies recombining with *pilE* during experimental human infection. *Microbiology* **147**, 839-849, doi:10.1099/00221287-147-4-839 (2001).
25 Hagblom, P., Segal, E., Billyard, E. & So, M. Intragenic recombination leads to pilus antigenic variation in *Neisseria gonorrhoeae*. *Nature* **315**, 156-158 (1985).
26 Murphy, G. L., Connell, T. D., Barritt, D. S., Koomey, M. & Cannon, J. G. Phase variation of gonococcal protein II: regulation of gene expression by slipped-strand mispairing of a repetitive DNA sequence. *Cell* **56**, 539-547 (1989).
27 Belland, R. J., Morrison, S. G., van der Ley, P. & Swanson, J. Expression and phase variation of gonococcal P.II genes in *Escherichia coli* involves ribosomal frameshifting and slipped-strand mispairing. *Mol Microbiol* **3**, 777-786 (1989).
28 Stern, A., Brown, M., Nickel, P. & Meyer, T. F. Opacity genes in *Neisseria gonorrhoeae*: control of phase and antigenic variation. *Cell* **47**, 61-71 (1986).
29 Connell, T. D., Shaffer, D. & Cannon, J. G. Characterization of the repertoire of hypervariable regions in the Protein II (*opa*) gene family of *Neisseria gonorrhoeae*. *Mol Microbiol* **4**, 439-449 (1990).
30 Yang, Q. L. & Gotschlich, E. C. Variation of gonococcal lipooligosaccharide structure is due to alterations in poly-G tracts in *lgt* genes encoding glycosyl transferases. *J Exp Med* **183**, 323-327 (1996).
31 Schneider, H. *et al.* Experimental human gonococcal urethritis: 250 *Neisseria gonorrhoeae* MS11mkC are infective. *J Infect Dis* **172**, 180-185 (1995).
32 John, C. M., Schneider, H. & Griffiss, J. M. *Neisseria gonorrhoeae* that infect men have lipooligosaccharides with terminal N-acetyllactosamine repeats. *J Biol Chem* **274**, 1017-1025 (1999).
33 Snyder, L. A., Butcher, S. A. & Saunders, N. J. Comparative whole-genome analyses reveal over 100 putative phase-variable genes in the pathogenic *Neisseria spp*. *Microbiology* **147**, 2321-2332, doi:10.1099/00221287-147-8-2321 (2001).
34 Jerse, A. E. Experimental gonococcal genital tract infection and opacity protein expression in estradiol-treated mice. *Infect Immun* **67**, 5699-5708 (1999).
35 Song, W. *et al.* Local and humoral immune responses against primary and repeat *Neisseria gonorrhoeae* genital tract infections of 17beta-estradiol-treated mice. *Vaccine* **26**, 5741-5751, doi:10.1016/j.vaccine.2008.08.020 (2008).
36 Packiam, M., Veit, S. J., Anderson, D. J., Ingalls, R. R. & Jerse, A. E. Mouse strain-dependent differences in susceptibility to *Neisseria gonorrhoeae* infection and induction of innate immune responses. *Infect Immun* **78**, 433-440, doi:10.1128/IAI.00711-09 (2010).
37 Muench, D. F. *et al.* Hydrogen peroxide-producing lactobacilli inhibit gonococci *in vitro* but not during experimental genital tract infection. *J Infect Dis* **199**, 1369-1378, doi:10.1086/597390 (2009).

38 Singer, A. The uterine cervix from adolescence to the menopause. *Br J Obstet Gynaecol* **82**, 81-99 (1975).
39 Islam, E. A. & Gray-Owen, S. D. in *Pathogenic Neisseria Genomics, Molecular Biology and Disease intervention* (eds J. K. Davies & C. M. Kahler) Ch. 12, 239 (Caister Academic Press, 2014).
40 Virji, M., Makepeace, K., Ferguson, D. J. & Watt, S. M. Carcinoembryonic antigens (CD66) on epithelial cells and neutrophils are receptors for Opa proteins of pathogenic *neisseriae*. *Mol Microbiol* **22**, 941-950 (1996).
41 Chen, T., Grunert, F., Medina-Marino, A. & Gotschlich, E. C. Several carcinoembryonic antigens (CD66) serve as receptors for gonococcal opacity proteins. *J Exp Med* **185**, 1557-1564 (1997).
42 Gray-Owen, S. D., Lorenzen, D. R., Haude, A., Meyer, T. F. & Dehio, C. Differential Opa specificities for CD66 receptors influence tissue interactions and cellular response to *Neisseria gonorrhoeae*. *Mol Microbiol* **26**, 971-980 (1997).
43 Boulton, I. C. & Gray-Owen, S. D. Neisserial binding to CEACAM1 arrests the activation and proliferation of CD4+ T lymphocytes. *Nat Immunol* **3**, 229-236, doi:10.1038/ni769 (2002).
44 Slevogt, H. *et al.* CEACAM1 inhibits Toll-like receptor 2-triggered antibacterial responses of human pulmonary epithelial cells. *Nat Immunol* **9**, 1270-1278, doi:10.1038/ni.1661 (2008).
45 Sintsova, A. *et al.* Selection for a CEACAM receptor-specific binding phenotype during *Neisseria gonorrhoeae* infection of the human genital tract. *Infect Immun* **83**, 1372-1383, doi:10.1128/IAI.03123-14 (2015).
46 Gulati, S. *et al.* Immunization against a saccharide epitope accelerates clearance of experimental gonococcal infection. *PLoS Pathog* **9**, e1003559, doi:10.1371/journal.ppat.1003559 (2013).
47 Nelson, J. F., Felicio, L. S., Randall, P. K., Sims, C. & Finch, C. E. A longitudinal study of estrous cyclicity in aging C57BL/6J mice: I. Cycle frequency, length and vaginal cytology. *Biol Reprod* **27**, 327-339 (1982).
48 Byers, S. L., Wiles, M. V., Dunn, S. L. & Taft, R. A. Mouse estrous cycle identification tool and images. *PLoS One* **7**, e35538, doi:10.1371/journal.pone.0035538 (2012).
49 Jerse, A. E. *et al.* Estradiol-Treated Female Mice as Surrogate Hosts for *Neisseria gonorrhoeae* Genital Tract Infections. *Front Microbiol* **2**, 107, doi:10.3389/fmicb.2011.00107 (2011).
50 Alexander, C. & Rietschel, E. T. Bacterial lipopolysaccharides and innate immunity. *J Endotoxin Res* **7**, 167-202 (2001).
51 Droge, W., Lehmann, V., Luderitz, O. & Westphal, O. Structural investigations on the 2-keto-3-deoxyoctonate region of lipopolysaccharides. *Eur J Biochem* **14**, 175-184 (1970).
52 Alakomi, H. L. *et al.* Lactic acid permeabilizes Gram-negative bacteria by disrupting the outer membrane. *Appl Environ Microbiol* **66**, 2001-2005 (2000).
53 Ganz, T. Defensins: antimicrobial peptides of innate immunity. *Nat Rev Immunol* **3**, 710-720, doi:10.1038/nri1180 (2003).
54 Kosciuczuk, E. M. *et al.* Cathelicidins: family of antimicrobial peptides. A review. *Mol Biol Rep* **39**, 10957-10970, doi:10.1007/s11033-012-1997-x (2012).
55 Packiam, M. *et al.* Phosphoethanolamine decoration of *Neisseria gonorrhoeae* lipid A plays a dual immunostimulatory and protective role during experimental genital tract infection. *Infect Immun* **82**, 2170-2179, doi:10.1128/IAI.01504-14 (2014).
56 Bals, R., Goldman, M. J. & Wilson, J. M. Mouse beta-defensin 1 is a salt-sensitive antimicrobial peptide present in epithelia of the lung and urogenital tract. *Infect Immun* **66**, 1225-1232 (1998).
57 Ricklin, D., Hajishengallis, G., Yang, K. & Lambris, J. D. Complement: a key system for immune surveillance and homeostasis. *Nat Immunol* **11**, 785-797, doi:10.1038/ni.1923 (2010).
58 Pridmore, A. C. *et al.* A lipopolysaccharide-deficient mutant of *Neisseria meningitidis* elicits attenuated cytokine release by human macrophages and signals via toll-like receptor (TLR) 2 but not via TLR4/MD2. *J Infect Dis* **183**, 89-96, doi:10.1086/317647 (2001).

59 Banerjee, A. *et al.* Identification of the gene (*lgtG*) encoding the lipooligosaccharide beta chain synthesizing glucosyl transferase from *Neisseria gonorrhoeae*. *Proc Natl Acad Sci U S A* **95**, 10872-10877 (1998).
60 Mandrell, R. E. *et al. In vitro* and *in vivo* modification of *Neisseria gonorrhoeae* lipooligosaccharide epitope structure by sialylation. *J Exp Med* **171**, 1649-1664 (1990).
61 Nairn, C. A. *et al.* Cytidine 5'-monophospho-N-acetylneuraminic acid or a related compound is the low Mr factor from human red blood cells which induces gonococcal resistance to killing by human serum. *J Gen Microbiol* **134**, 3295-3306, doi:10.1099/00221287-134-12-3295 (1988).
62 Parsons, N. J. *et al.* Cytidine 5'-monophospho-N-acetyl neuraminic acid and a low molecular weight factor from human blood cells induce lipopolysaccharide alteration in gonococci when conferring resistance to killing by human serum. *Microb Pathog* **5**, 303-309 (1988).
63 Steeghs, L. *et al.* Meningitis bacterium is viable without endotoxin. *Nature* **392**, 449-450, doi:10.1038/33046 (1998).
64 Harder, J., Bartels, J., Christophers, E. & Schroder, J. M. A peptide antibiotic from human skin. *Nature* **387**, 861, doi:10.1038/43088 (1997).
65 Faurschou, M. & Borregaard, N. Neutrophil granules and secretory vesicles in inflammation. *Microbes Infect* **5**, 1317-1327 (2003).
66 Eisenhauer, P. B. & Lehrer, R. I. Mouse neutrophils lack defensins. *Infect Immun* **60**, 3446-3447 (1992).
67 Gallo, R. L. *et al.* Identification of CRAMP, a cathelin-related antimicrobial peptide expressed in the embryonic and adult mouse. *J Biol Chem* **272**, 13088-13093 (1997).
68 Ganz, T. Extracellular release of antimicrobial defensins by human polymorphonuclear leukocytes. *Infect Immun* **55**, 568-571 (1987).
69 Sellers, R. S., Clifford, C. B. & Treuting, P. M. B., C. Immunological variation between inbred laboratory mouse strains: points to consider in phenotyping genetically immunomodified mice *Veterinary Pathology* **49**, 32-43 (2012).
70 Fleischmann, R. D. *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496-512 (1995).
71 Metzker, M. L. Sequencing technologies - the next generation. *Nat Rev Genet* **11**, 31-46, doi:10.1038/nrg2626 (2010).
72 Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304-1351, doi:10.1126/science.1058040 (2001).
73 Schatz, M. C., Delcher, A. L. & Salzberg, S. L. Assembly of large genomes using second-generation sequencing. (2010).
74 Land, M. *et al.* Insights from 20 years of bacterial genome sequencing. *Funct Integr Genomics* **15**, 141-161, doi:10.1007/s10142-015-0433-4 (2015).
75 Schadt, E. E., Turner, S. & Kasarskis, A. A window into third-generation sequencing. *Hum Mol Genet* **19**, R227-240, doi:10.1093/hmg/ddq416 (2010).
76 IMG. (2016).
77 Andrews, S. *FastQC: a quality control tool for high throughput sequence data*, <Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (2010).
78 Adey, A. *et al.* Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density *in vitro* transposition. *Genome Biol* **11**, R119, doi:10.1186/gb-2010-11-12-r119 (2010).
79 Goryshin, I. Y., Miller, J. A., Kil, Y. V., Lanzov, V. A. & Reznikoff, W. S. Tn5/IS50 target recognition. *Proc Natl Acad Sci U S A* **95**, 10716-10721 (1998).
80 Lamble, S. *et al.* Improved workflows for high throughput library preparation using the transposome-based Nextera system. *BMC Biotechnol* **13**, 104, doi:10.1186/1472-6750-13-104 (2013).
81 Kelley, L. A. & Sternberg, M. J. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**, 363-371, doi:10.1038/nprot.2009.2 (2009).

82 Illumina, I. Calling Sequencing SNPs. (Solexa, 2009).
83 Erwin, A. L., Haynes, P. A., Rice, P. A. & Gotschlich, E. C. Conservation of the lipooligosaccharide synthesis locus *lgt* among strains of *Neisseria gonorrhoeae*: requirement for *lgtE* in synthesis of the 2C7 epitope and of the beta chain of strain 15253. *J Exp Med* **184**, 1233-1241 (1996).
84 Gulati, S., McQuillen, D. P., Mandrell, R. E., Jani, D. B. & Rice, P. A. Immunogenicity of *Neisseria gonorrhoeae* lipooligosaccharide epitope 2C7, widely expressed in vivo with no immunochemical similarity to human glycosphingolipids. *J Infect Dis* **174**, 1223-1237 (1996).

Appendix – Permissions

Nature Publishing Group

Figure 1 & Figure 4
Images used from Gray-Owen & Blumberg (2006)
License Number: 4025500951203
License Date: January 10, 2017

Caister Academic Press

Figure 3
Image used from Islam & Gray-Owen (2014)
(http://www.caister.com/help/copyright.html)