GENETIC FLUX IN A GLOBAL COLLECTION OF INVASIVE *STREPTOCOCCUS PNEUMONIAE* GENOMES

TERESA MWIKALI MUTUA

I56/88033/2016

Project thesis presented to the University of Nairobi, Center for Biotechnology and Bioinformatics for the award of Master of Science in Bioinformatics

JUNE 2021

DECLARATION

This project is my own work and has not been submitted elsewhere for degree.

TERESA MWIKALI MUTUA

STUDENT REG NO: I56/88033/2016

STUDENT SIGNATURE: Teresa

DATE: 20th June 2021

This project has been submitted with my authorization as the University supervisor:

Dr. BENARD W. KULOHOMA

Center for Biotechnology and Bioinformatics (CEBIB), University of Nairobi.

SUPERVISOR SIGNATURE: Benard Kulohoma.

DATE: 28th June 2021

DEDICATION

I dedicate my work to my son Sebastian for being a great inspiration throughout my project work. He gave me a reason to work extra hard to create, a path which he can emulate. I am happy that one day he will understand that it is the will and effort to make a step that gives the energy to break a glass ceiling. I also dedicate this work to my hubby Mr. Chrispine Oguku for his support and understanding throughout the study process.

I dedicate this work and give special thanks to my parents, the Late Sebastian Mutua and Mrs. Georgina Mutua. My father believed in me and always encouraged me to pursue my dreams and interests. My greatest joy is to make him proud as the angels guard him. My mother set a path for me to follow, through her dedication to her own work and her strength encouraged me every day of my study.

ACKNOWLEDGEMENT

I thank God for the granting me this opportunity to pursue my dreams and His abundant grace and blessings. I thank my family for their support. My gratitude goes to the entire CEBIB fraternity for the support and guidance during my study period. A special gratitude to my supervisor Dr. Benard W. Kulohoma for his mentoring, guidance and assistance during my project work. I thank Mr. Kelvin Muteru for his advice and assistance during my study. I thank my fellow students for their advice and assistance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

CSF:	Cerebrospinal Fluid
NCBI:	National Centre for Biotechnology Information
MCL Algorithm:	Markov Cluster Algorithm
CEBIB:	Centre for Biotechnology and Bioinformatics
IPD:	Invasive Pneumococcal Disease (s)
PCV7:	7-valent pneumococcal conjugate vaccine
PCV13:	13-valent pneumococcal conjugate vaccine
PPSV23:	23-valent pneumococcal polysaccharide vaccine
HIV:	Human Immunodeficiency Virus
CDC:	Centre for Disease Control and Prevention
AT:	Anti-Toxins
EMBL:	European Molecular Biotechnology Laboratory
BLAST:	Basic Local Alignment Search Tool
WGS:	Whole Genome Shotgun
GLOOME:	Gain Loss Mapping Engine
BMX:	Bacterial Makeup eXplorer
FDA:	Food and Drug Administration
PCR:	Polymerase Chain Reaction
Ply:	Pneumolysin
AIDS:	Acquired Immunodeficiency Syndrome
CSP:	Competence Stimulating Peptide
HSP:	Heat Shock Protein

ABSTRACT

Streptococcus pneumoniae, also known as the pneumococcus, is a major cause of life-threatening bloodstream infections, and may cross the blood-brain barrier and cause meningitis. Invasive pneumococcal disease (IPD) affects all age groups, but the populations highest at risk of infection are children, the elderly, and individuals with compromised immunity. Despite the implementation of childhood immunization programs and effective antimicrobial agents, child mortality from pneumococcal meningitis still imposes a huge disease burden, even in developed countries. This study aimed to understand the differences in patterns of *Streptococcus pneumoniae* genome evolution through gene loss and gain events, and their effect on the propensity to cause meningitis compared to bacteremia. *Streptococcus pneumoniae* isolate genomes of strains retrieved from human cerebrospinal fluid (CSF) were compared to those retrieved from human peripheral blood. The two datasets were first each analyzed separately, followed by comparisons across the two subsets. Briefly, the sequences in each data subset were first broadly compared using an All vs All BLAST comparison. The BLAST results were then more accurately clustered into orthologous groups using a hidden Markov chain model algorithm called OrthoMCL. The resultant orthologous map generated was then annotated and processed using Bacterial Makeup eXplorer (BMX) to generate annotated phyletic patterns highlighting gene presence and absence. The phyletic patterns were further analyzed using the Gene Loss Mapping Engine (GLOOME) to determine the probability of genes acquisition or loss along the length of each genome in the dataset under study. The results were then analyzed to make inference of the general direction of evolution, which is gene gain or loss events, which are associated with propensity to cause meningitis or not; when comparing the meningitis and bacteremia associated data subsets. Among the known virulence proteins, putative bacteriocin transporter C39 protease domain BlpA2 and pneumococcal histidine triad protein D (bvh-11-2) showed more gene loss events in the meningitis set. The immunity protein PncB, pncF, immunity protein PncK and bacteriocin BlpO displayed more gene loss events in the bacteremia set. More gene loss events were observed in both bacteremia and meningitis sets for putative immunity protein PncM and putative membrane protein BlpL. Also, more gene gain in both sets was observed for putative uncharacterized protein PncC. There was more gene gain in bacteremia set for cell surface choline binding protein PcpA. The overall findings suggests that meningitis genomes were more conserved compared to those generated from bacteremia isolates. They highlight mechanisms that determine differences in invasive ability during infection since gene loss and acquisition primarily contribute to how bacteria genetically adapt to novel environments and diverge to form separate, evolutionarily distinct species and strains. Genetic flux can radically and rapidly increase fitness or alter some aspects of lifestyle, such as multidrug non-susceptibility.

CHAPTER 1

INTRODUCTION

1.1 Pneumococcal infections

Pneumococcal invasive disease (IPD) results from *Streptococcus pneumoniae* invasion of a host normally sterile sites, which include lungs, blood, heart, inner ear, and brain (Li et al., 2019). The population most at risk of the IPD are the children under the age of 5 years, the elderly and immunocompromised individuals (Brooks & Mias, 2018). The immunity of children is not well developed, exposing them to the risk of getting infected. Young children also have a high frequency of pneumococcal colonization due to their nature of interaction at various enclosed institutions like schools and children daycare, hence they are considered as the most likely vectors for pneumococcal strains horizontal dissemination in the community (Xu, Almudervar, Casey, & Pichichero, 2013). Immunity of the elderly (>65 years) is waning over time, hence predisposing them to the infection (Berical et al., 2016). Immunocompromised individuals include people living with HIV/AIDS, functional or anatomical asplenia, genetic immune deficiencies and people with cancer among other chronic conditions. In the year 2015 the global prevalence and case fatality rate for IPD was 36.4/100,000 and 0.68/100,000 in children under 5 years, and 107.5/100000 and 19.89 in adults aged 50 years and above, respectively (Brooks & Mias, 2018) (Table 1).

Table 1: Occurrence of pneumococcal diseases from 1997 to 2015 as reported by the Center for Disease Control and Prevention. This table shows the morbidity and mortality rates from pneumococcal diseases among different age cohorts. Table reproduced from (Brooks & Mias, 2018).

Year	1997		2007		2012		2014		2015	
Age	Case rate	Death rate	Case rate	Death rate	Case rate	Death rate	Case rate	Death rate	Case rate	Death rate
<1	142.9	4.02	40.51	0.9	15.7	0.24	15.9	0.48	18.4	0.24
1	178.7	0.9	32.39	0.23	13.6	0.24	10.3	0	12.9	0.24
2-4	31	0.15	13.03	0.08	5.9	0	6.3	0.08	5.1	0.16
5-17	4.8	0.14	2.19	0.14	1.9	0.14	1.4	0.05	1.3	0
18-347	9.3	0.52	4.19	0.22	2.8	0.1	2.7	0.18	2.5	0.08
35-49	18.9	1.65	11.89	0.98	7.5	0.6	6.6	0.7	6.7	0.5
50-64	23.5	2.72	20.59	2.33	15.9	1.53	15.1	1.64	15	1.53
65-74	61.7	11.02	39.26	6.37	29.6	4.24	19.1	2.41	18.2	2.3
75-84							28.2	3.46	29	4.5
≥85							42.6	8.01	45.3	11.56

Invasive pneumococcal diseases can be prevented through hygiene, healthy diet and vaccination against the pneumococcus. Hygiene involves regular hand washing, body cleanliness, and proper food and drinks handling as defined by the Food and Drug Administration (FDA) (*The FDA Food Safety Modernization Act*, 2011). Healthy diet is achieved through proper preparation and storage of food to ensure maximum preservation of the nutrients and prevention of food contaminants to help in boosting immunity (Magni et al., 2017). Antibiotics can be used to prevent pneumococcal infections through reduction of bacterial load by inhibiting growth of, or killing the bacteria (Bistrović et al., 2018). Penicillin was initially the preferred antibiotic against *Streptococcus pneumoniae,* discovered by Alexander Fleming in 1928, and paving way for development of other antibiotics (Berical et al., 2016). Pneumococci develop resistance to antibiotics with continued use and transmit the resistant genes with their progeny, creating a need for the development of novel antibiotics (van der Poll & Opal, 2009). Vaccination can also be used for prevention of pneumococcal diseases. There is a regimen for pneumococcal vaccines given to all children under 5 years incorporated in childhood immunization schedules. The seven-valent pneumococcal conjugate vaccine (PCV7) is among the vaccines developed and approved for children in the year 2000, with a marked improvement to disease depletion, showing 64% decrease in < 2 years old children and 54% decrease in > 65 years old adults by 2005 (Berical et al., 2016). The thirteen-valent pneumococcal conjugate vaccine (PCV13) was also approved in the year 2010 for use in children as young as 6 weeks. The continued vaccine evolution and development, led to an increase in occurrence of IPD by nonvaccine serotypes, which prompted for more readjustments and specifications in the childhood administration strategies of the vaccines (Gerdes, 2013; Plumptre et al., 2013).

1.2 Bacteriology of *Streptococcus pneumoniae*

1.2.1 Classification of the pneumococcus

Streptococcus pneumoniae is a gram-positive, facultative anaerobic bacteria and it's common for its highly invasiveness nature (Tomos, n.d.). This bacterium is an alpha-hemolytic (α-hemolytic) pathogen when under aerobic conditions and beta-hemolytic (β-hemolytic) when under anaerobic conditions (Hajaj et al., 2012).

1.2.2 Invasive pneumococcal diseases diagnosis

In addition to clinical signs and symptoms, a Gram-positive stain and laboratory culture of a sample collected from the either peripheral blood, CSF, nasopharynx or middle ear is used to diagnose invasive pneumococcal diseases. Polymerase chain reaction (PCR) is among the most significant rapid methods to perform a molecular-based detection and differentiation of *Streptococcus pneumoniae*. However, PCR may be susceptible to contamination and inhibitors, which could lead to misdiagnosis (Yamamoto, 2002).

1.3 Epidemiology of the invasive pneumococcal diseases

1.3.1 Geographic distribution of the invasive pneumococcal diseases

IPD cover a wide geographical region affecting both developed and developing countries with Africa having the highest incidences (O'Brien et al., 2009). The incidence of IPD in Asia and Latin America is reported to be higher compared to North America and Europe. Children are the main carriers of *Streptococcus pneumonia* bacteria, especially in developing countries and among some indigenous societies of the developed countries (World Health Organization, 2019). The incidence of IPD and age distribution of cases among children may vary in different countries depending on the socio-economic status. More specifically, the incidence of meningitis correlate with child mortality rate and varies geographically (O'Brien et al., 2009). Geographical region is among the factors associated with the prevalence of the known > 90 *Streptococcus pneumonia* serotypes, with less serotypes being associated with IPD morbidity over time, due to effective vaccines intervention (van der Poll & Opal, 2009; World Health Organization, 2019).

1.3.2 The pneumococcal disease burden

Globally, the prevalence of pneumococcal diseases is approximately 14 million cases annually (Benard Kulohoma, 2012; O'Brien et al., 2009) and approximately 300, 000 deaths each year in 0-59 months old children(Wahl et al., 2018). Despite the development of better interventions, that includes antibiotics and vaccines to prevent and manage IPD, there is still continuous need for novel IPD mitigating strategies. This is because of the antibiotic efficacy being challenged by increased antibiotics resistance (Brooks & Mias, 2018), and the bacteria are also able to escape the vaccine (Brueggemann et al., 2013).

1.4 Aim and Objectives

1.4.1 Aim

This study aimed to highlight differences in patterns of *Streptococcus pneumoniae* genome evolution associated with gene loss and gene gain that leads to the propensity to cause meningitis compared to bacteremia.

1.4.2 Hypothesis

Pneumococci with a propensity to cause meningitis display a different pattern of genetic flux compared to those that cause bacteremia.

1.4.3 Objectives

1. To establish the cumulative number of gene gain and loss events along the genomes of pneumococci associated with bacteremia compared to those associated with meningitis
2. To establish the probability of gene gain and loss for all genes, core genes and accessory genes in the genomes of pneumococci associated with bacteremia compared to those associated with meningitis

1.5 Justification

Streptococcus pneumoniae are commensal bacteria found on human nasopharynx of healthy individuals (Benard Kulohoma, 2012). However, this bacteria can invade the host's normally sterile compartments (blood, middle ear, lungs and CSF) and lead to invasive pneumococcal diseases that comprise; bacteremia, pneumonia, otitis media and meningitis (Feldman & Anderson, 2014; Ogunniyi et al., 2012; Orihuela et al., 2004). *Streptococcus pneumoniae* display resistance to multiple antibiotics due its rapid genetic mutation hence prompting continuous research (Marks et al., 2012). Pneumococcal multidrug resistance threatens to reverse gains made in disease management and continuous analyses are required to understand mechanisms involved in development multiple lineages capable of circumventing current interventions (Pan et al., 2018). Although evolutionary biology and the population genetics of the pneumococcus is well understood, it remains unclear how some pneumococci are able to breach the blood-brain-barrier and cause meningitis, while others are not. Meningitis, a severe form of IPD, is associated with high mortality and permanent neurological impairment (Meichanetzidis et al., 2018). There is limited knowledge on the genetics of *Streptococcus pneumoniae* associated with the propensity to cause meningitis (Li et al., 2019). This study improves knowledge on *Streptococcus pneumoniae* genetic makeup as well as gene gain and gene loss events in bacteremia and meningitis causing strains, and their ability to facilitate breaching the blood-brain barrier. This study characterizes the gene gain and gene loss in a global collection of invasive *Streptococcus pneumoniae* isolate genomes. The findings highlight different patterns associated with mechanisms that determine differences in invasiveness during infection since gene loss and acquisition primarily contribute to how bacteria genetically adapt to novel environments and diverge to form separate, evolutionarily distinct species and strains.

CHAPTER 2

LITERATURE REVIEW

2.1 Epidemiology of the pneumococcal diseases

Streptococcus pneumoniae causes diseases globally mostly affecting children and elderly; with an estimate of almost a million deaths in children of <5 years in 2013 (Feldman & Anderson, 2016). In children, *Streptococcus pneumoniae* asymptomatically colonizes the upper respiratory system preceding the IPD (Adegbola et al., 2014).The presence of the bacteria in nasopharynx without causing any symptoms is called carriage, and it is high in children compared to adults (Berical et al., 2016). The case of morbidity and mortality is even more amongst immunocompromised children, for instance, children with HIV/AIDS have 40-fold increase of the risk of the infections compared to children without HIV/AIDS (Kulohoma et al., 2017).

There is a significant improvement in the reduction of the number of deaths with time, although drug resistance and poor host immunity pose a new challenge (Kim et al., 2016). Worldwide immunization programs have been implemented to prevent deaths in children under 5 years (Cecchini et al., 2018). However, these vaccination programs have led to selection of strains with serotypes absent in vaccines (Berical et al., 2016; Masomian et al., 2020).

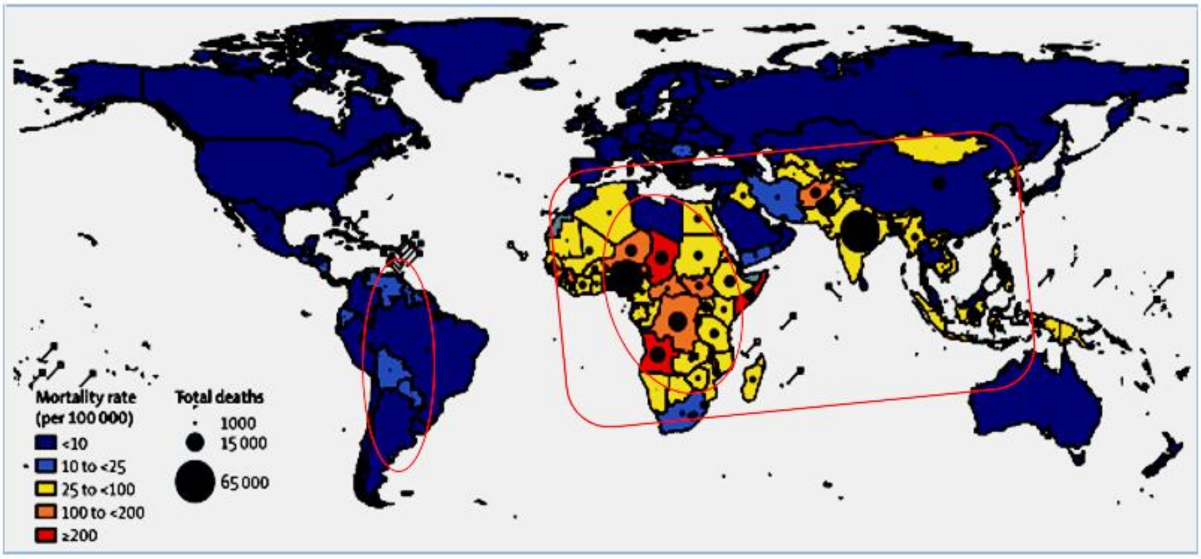


Figure 1: Worldwide distribution of children < 5 years old pneumococcal deaths per 100,000 children population. This figure shows the global prevalence of child mortality, among HIV negative young children as a result of invasive pneumococcal diseases in the year 2015. There is a disproportionate disease burden in Africa, Asia, and the Middle East. Reproduced from (Wahl et al., 2018).

The morbidity and mortality due to drug resistance results from bacterial genetic transformation and recombination (Henriques-Normark & Tuomanen, 2013) . There is a marked increase in β -lactam resistance, a result of susceptibility tests conducted in different laboratories (Cecchini et al., 2018). This continued antibiotics non-susceptibility and vaccine escape in the recent pneumococcal eradication measures, has led to research for other IPD management strategies. For example, the utilization of pneumococcal toxin-antitoxin (TA) genes as drug development candidates (Chan & Espinosa, 2016). Similar to bacteriocins, TA systems consist of closely related genes, which convert a toxin to an antitoxin, to give stability to bacteria by making them immune to the toxin. Bacteriocins on the other hand, are encoded on plasmids and include a bacteriocin gene and an immunity gene, which convert the toxin to a non-toxin (Bayramoglu et al., 2017; Pei & Grishin, 2001).

2.2 Pathogenicity of the pneumococcus

Streptococcus pneumoniae is a commensal colonizing the respiratory mucosa and the colonization is enabled by an extracellular pneumococcal capsule, acting as a protective layer to the bacteria (Shenoy & Orihuela, 2016). The polysaccharide capsule found over the outer surface of *Streptococcus pneumoniae* is very heterogeneous with nearly a hundred different capsular serotypes identified (Tennant et al., 2016). As the most relevant virulence factor of pneumococci, the polysaccharide capsule protects it from phagocytosis (Bogaert et al., 2004). The bacteria are transmitted into the host system through contact with an exposed surface or through coinfection, for example with Influenza virus, which compromises the host immune system (Mücke et al., 2020).

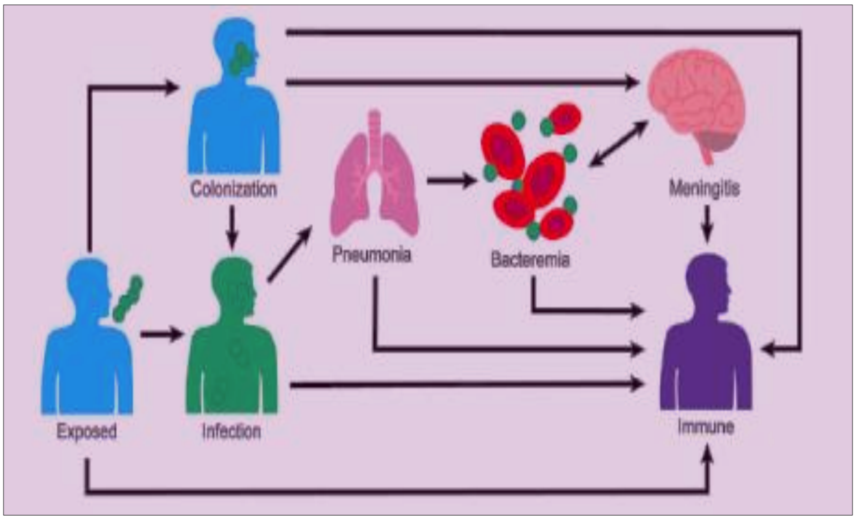


Figure 2: *Streptococcus pneumoniae* colonization and infection routes. The common pathways to IPD in the human body after exposure.

Non-invasive bacteria colonize the nasopharynx and the lower respiratory system after exposure to the bacteria. Bacteria invade the host system infecting the blood, ears, sinuses, joints and peritoneum. Further, the bacteria cross the blood-brain-barrier causing meningitis. Individuals with strong immunity can easily resist the infection. Reproduced from (Zivich et al., 2018).

Pneumococci produce harmful antigens, which facilitate its pathogenesis (Brooks & Mias, 2018). These antigens include: autolysin and pneumolysin (Ply) among others (Popowicz et al., 2017). Pneumolysin is not associated with colonization of the pneumococcus. Autolysin activates the release of pneumolysin while the pneumolysin leads to cell lysis of the host cells as well as induction of complement pathway activation (Popowicz et al., 2017). In the bloodstream, the bacteria are confronted by the host complement system as a way of host-pathogen clearance (Andre et al., 2017). *Streptococcus pneumoniae* has preferential adherence to different sites of the brain as well as an immediate activation of the brain local immunity upon bloodstream infection. Pneumococcal pathophysiology leads to blood-brain barrier dysfunction, hence central nervous system invasion through the choroid plexus epithelium causing meningitis (Iovino et al., 2013; Prager et al., 2017).

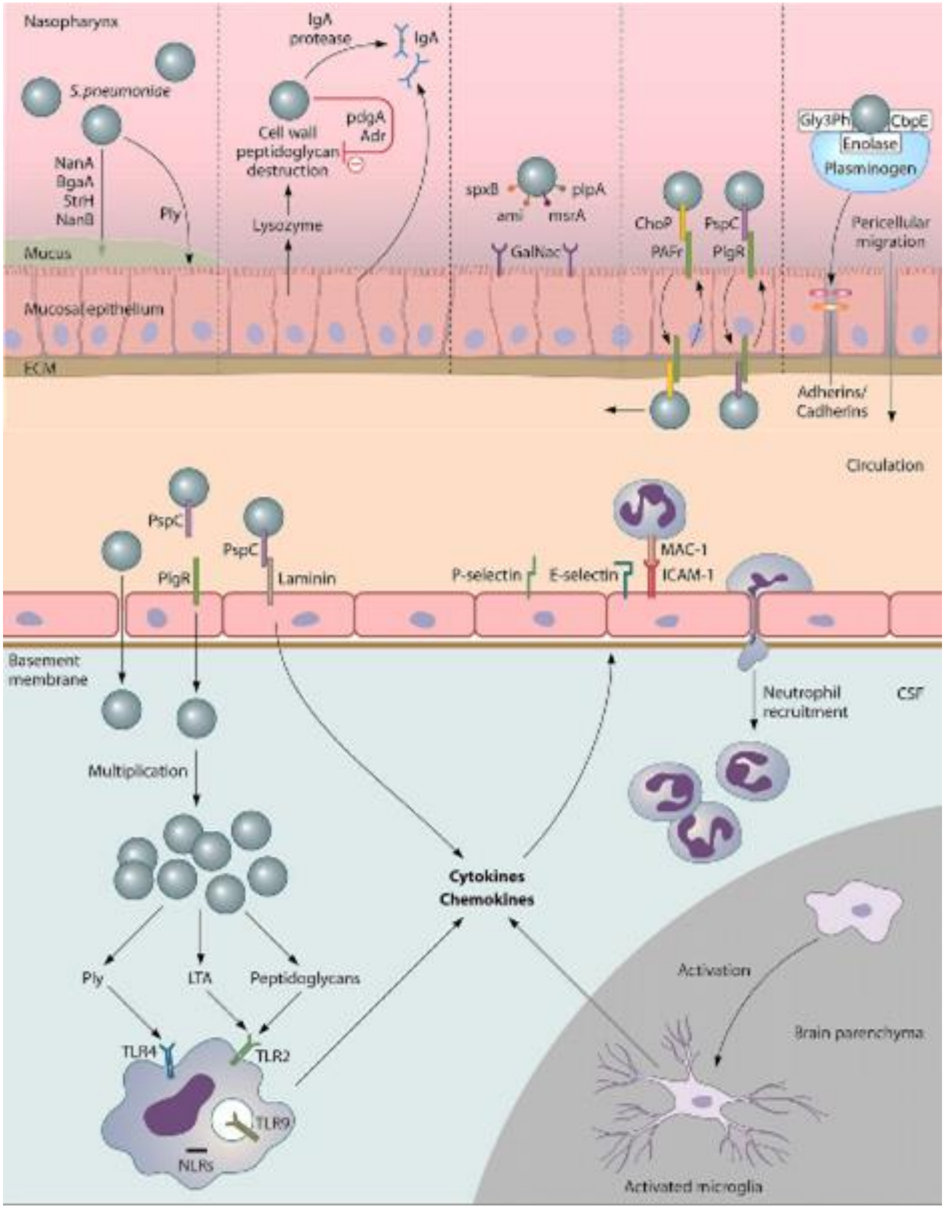


Figure 3: The pneumococcus transmission from nasopharynx through blood to brain, by crossing the Blood Brain Barrier (BBB). This figure shows the process of the *Streptococcus pneumoniae* colonizing the human respiratory system and invading the human blood and finally crossing the blood-brain-barrier into the cerebrospinal fluid. Reproduced from (Iovino et al., 2016)

2.3 The pneumococcal disease management

2.3.1 The use of antibiotics against *Streptococcus pneumoniae*

IPD has successfully been managed using antibiotics (El Khoury et al., 2017), although rapid development of antibiotic resistance is challenging disease management (Zivich et al., 2018). There is also the occurrence of multidrug resistance where strains of *Streptococcus pneumoniae* are resistant to two or more antimicrobials (Pan et al., 2018).

In 1917, acquired resistance to optochin was first reported in humans, and 22 years later, cases of human pneumococcal meningitis showed development of treatment acquired resistance to sulfonamides (Kim et al., 2016). In 1967, there was marked resistance to penicillin reported from clinical pneumococci isolates (Muley et al., 2016), and this, together with trimethoprim-sulfamethoxazole and erythromycin resistant pneumococci, spread fast worldwide in 1970s and 1980s (Muley et al., 2016). There was also reported cases of chloramphenicol and tetracycline resistance (Raddaoui et al., 2015). There was relatively lower resistance to fluoroquinolones compare to the other antibiotics (Kim et al., 2016). Multidrug resistant strains were documented in the 21st century among isolates of specific serotypes (Cornick et al., 2014; Ip et al., 2001; Kim et al., 2016). The genetic mutations enabling for drug resistance are shown in Table 2.

Table 2: Molecular mechanisms responsible for most observed cases of pneumococcal antibiotic resistance. Antibiotics used over time to manage pneumococcal diseases and the respective mutations leading to their resistance. Table reproduced from (Kim et al., 2016)

Antibiotic	**Mechanism (s)**
β-lactams (penicillin and cephalosporins)	Mutations in penicillin-binding (transpeptidase) domains of *pbp* genes (primarily *pbp2x, pbp2b,* and *pbp1a*); mutations in aminoacyl-tRNA ligase gene (*murM*); mutations in other genes, including *pdgA, ciaH-ciaR,* and *stkP*
Macrolides	*Erm (*235 rRNA methyltransferases*) (ermB* and rarely *ermTR), mef*-mediated efflux [*mef (*A*)* or *mef €,* mutations in 23SrRNA genes or L4 or L22 ribosomal protein genes (*rplD* and *rplV,* respectively)
Fluoroquinolones	Mutations in DNA gyrase (primarily *gyrA*) and/or topoisomerase IV genes (primarily *parC*), pmrA-mediated efflux
Tetracycline	Ribosomal protection proteins, primarily Tet (M) and rarelyTet (O)
Rifampin	Mutations in *rpoB* encoding the β-subunit of RNA polymerase

Antibiotic	**Mechanism (s)**
Chloramphenicol	Inactivation of chloramphenicol by *cat*-encoded chloramphenicol acetyltransferase
Trimethoprim-sulfamethoxazole	Mutations in the dihydrofolate reductase gene (*folA*) and dihydropterolate synthetase gene (*folP*)
Ketolides	Mutations in 23S rRNA or L4 or L22 ribosomal protein genes (*rplD* and *rplV*), *ermB* with deletion or mutation in leader sequence
Oxazolidinones	Mutations in 23S rRNA genes, deletions in L4 ribosomal protein gene *rplD*

2.3.2 Vaccines used against *Streptococcus pneumoniae*

A vaccine is a biological concoction containing live or dead microorganism, an agent of a microbe, or an extract of the microbe, aimed at boosting immunity to a specific infection, by stimulating the immune system to identify the agent as alien, destroy and remember it in case of recurrence (Burton, 2017). This acts as a preventive measure of reducing disease occurrence and severity (Alderson, 2016).

Currently, the US Food and Drug Administration (FDA) approved pneumococcal vaccines include PCV13 and 23-valent pneumococcal polysaccharide vaccine (PPSV23) although PCV7 and PCV10 are in use in other countries. The PPSV23 is recommended after conjugated vaccines immunization series for at-risk children, to provide protection against the pneumococcal capsular serotypes causing disease (Zivich et al., 2018). The PCV is recommended to healthy children under the age of 2 years.

Table 3: Pneumococcal vaccine approval dates, serotypes, and general effect on pneumococcal disease. The symbols * and † indicate serotypes found uniquely in the PPSV23 and PCV13 vaccines respectively. Reproduced from (Daniels et al., 2016).

Vaccine	**FDA Approval**	**Serotypes Contained in Vaccine*†**	**Pneumococcal Disease Effect from Vaccine Serotypes**
PPSV23	June 1983	1, 2*, 3, 4, 5, 6B, 7F, 8*, 9N*, 9V, 10A*, 12F*, 14, 15*, 17F*, 18C, 19A, 19F, 20*, 22F*, 23F, and 33F*	• Reduced invasive diseases • No effect on carriage
PCV7	February 2000	4, 6B, 9V, 14, 18C, 19F, and 23F	• Reduced invasive diseases • Reduced carriage • Protective herd effective • Increase in 19A infections
PCV13	February 2010	1, 3, 4, 5, 6A†, 6B, 7F, 9V, 14, 18C, 19A, and 23F	• Reduced invasive diseases • Reduced carriage • Increase in 35B infections

Table 4: PPSV23 and PCV13 vaccination recommendations for children and adults aged between 5 to 64 years of age with medical conditions. CSF, cerebrospinal fluid; HIV, human immunodeficiency virus; PPSV23, 23-valent pneumococcal polysaccharide vaccine; PCV13, 13-valent pneumococcal conjugate vaccine. Reproduced from (Daniels et al., 2016).

Condition	**Number of Doses**	
	PPSV23	**PCV13**
Chronic heart disease, chronic lung disease, diabetes mellitus	1	0
Sickle cell disease, functional or anatomical asplenia	2	1 administered before 23-valent
CSF leakage, cochlear implant, congenital immunodeficiency, HIV infection, chronic renal failure, cancer, transplant recipient	1	1 administered before 23-valent

Table 5: PCV13 vaccination catch-up dose recommendations for healthy children under 5 years of age. PCV7, 7-valent pneumococcal conjugate vaccine; PCV13, 13-valent pneumococcal conjugate vaccine. *Minimum interval between doses depends on age at which the first dose was administered (CDC, 2019; World Health Organization, 2019).

Previous PCV7 Dose	**Age (months)**	**Number of Doses to Complete Primary Series**	**Booster Dose**
None	<6	3	Yes
None	7-11	2	Yes
None	12-23	2	No
-	24-59	-	Yes
1 or 2	7-11	1	Yes
None or 1 given <12 months	12-23	2	No
1 given >12 months	12-23	1	No
2-3 doses given <12 months	12-23	1	No

2.4 Pneumococcal virulence factors and vaccine candidates

The pneumococcal virulence factors include pneumolysin, autolysin, neuraminidases, enolase, pneumococcal surface proteins A and C, choline-binding proteins, hyaluronate lyase, pneumococcal adhesion and virulence A, the metal-ion-binding proteins PsaA, PiaA and PiuA, the capsule and the cellwall (Brooks & Mias, 2018; Feldman & Anderson, 2016; M. J. Jedrzejas, 2003; Mitchell & Mitchell, 2010). Each of these factors have specific roles during colonization and infection(Bryant et al., 2016; Kadioglu et al., 2008). There is a different distribution of invasive *Streptococcus pneumoniae* genotypes in every region globally hence different genotypes of the pneumococcal conjugate vaccines are introduced in each region independently after a thorough understanding of the existing genetic structure and the trend of prevalence of the genotypes (Brueggemann et al., 2013). There are several protein antigens, including pneumococcal surface protein A, choline-binding protein A and pneumolysin, which can cause an immune response for protection, hence qualify as vaccine candidates (Abry et al., 2015; Cecchini et al., 2018). The ideal vaccine candidates should be capable of eliciting a T-cell dependent immune response; and should be conserved throughout all genotypes to ensure broad coverage in the population (Abry et al., 2015).

2.5 Pneumococcal genomics and reverse vaccinology

The genetic composition of bacteria varies according to the lifestyle they are adapted to, based on the niche they colonize. The pneumococcus is able to colonize a variety of host ecological environments, and this is displayed by its ability to colonize the nasopharynx as a commensal and to invade different-normally sterile sites, for example the middle-ear, blood, lungs and CSF, causing invasive disease (Obolski et al., 2019; Orihuela et al., 2004).

The genome of *Streptococcus pneumoniae* rapidly evolves to adapt to the host environments within a single infection, referred as selective adaptation (Nelson et al., 2007). *Streptococcus pneumoniae* mutations occur generally leading to different genetic structures, which dictate the course of infection of each specific strain, hence different strains invade different host body compartments (Brooks & Mias, 2018).

Streptococcus pneumoniae rapidly evolve multidrug resistance and have virulence antigens that can be transferred among strains through horizontal gene transfer (Andam & Hanage, 2015a). For example, bacteriocin encoded plasmids are horizontally transferred through physical contact using some mobile genetic material (Andam & Hanage, 2015b). The frequent horizontal transfer of genes leading to high resistance of antibiotics, makes it difficult to generate vaccines and drugs unless these interventions target virulence factors present across all genotypes (Henriques-Normark & Normark, 2014). Genomic information allows vaccine development without using the pathogen but rather an *in silico* approaches (Burton, 2017). This process eases the procedure of vaccine candidate selection and design (Delany et al., 2013).

2.6 Protein antigens with promising medical intervention ability

Bacteriocins proteins have posed a challenge to drug development, due to their pathogenicity potential and ability to enable persistent pathogen survival (Koedel et al., 2002). These bacteria produced proteinaceous toxins survive by killing their co-existent strains as well as their closely related peptides and use their DNA to strengthen their adaptability, hence prolonging pathogen survival (Weiser et al., 2018).

2.6.1 ABC transporters

ABC transporters are a class of trans-membrane exporters and importers of various substrates, consisting of four proteins, two ATPases (ATP-binding proteins) and two permeases (membrane-spanning proteins). Bacterial pathogen virulence is a consequence of the ABC transporters influence to cellular processes which among them is antimicrobial resistance

(Basavanna et al., 2009). Some *Streptococcus pneumoniae* ABC transporters are relevant considering they could cause impact on the pathogenesis of pneumococcal infections, given that they aid in nutrition, growth and virulence of the bacteria (Basavanna et al., 2009). ABC transporters substrate-binding proteins are attached to the outer surface of the pneumococcal membrane, hence are exposed potential vaccine target (Garmory & Titball, 2004).

2.6.2 Autolysin

The autolysins (ALs) responsible for pneumococcal virulence include LytA, LytB and LytC. The major pneumococcal autolysin is an enzyme encoded by the *lytA* gene involved in cell-wall destruction. This enzyme has a transposable arrangement with a C choline binding terminal domain and N terminal N-acetylmuramoyl L-ala- nine amidase domain (Mellroth et al., 2012; Whatmore et al., 1999) The pathogenesis of autolysin can cause virulence directly or indirectly. Directly by damaging the cell-wall and releasing its products, which can be highly inflammatory, promoting pathogenesis (Berry et al., 1989). Indirectly by obtruding cell lysis and the successive release of virulence factors like the pneumolysin, which is transported in an inactive form from the cell (Mark J Jedrzejas, 2001). The activity of LytA partly contribute to the resistance of some antibiotics like penicillin and vancomycin (Mellroth et al., 2012). Both LytA and LytC are involved in the release of the cell wall degradation products and pneumolysin among other cytoplasmic components (Herta et al., 2018; Martner et al., 2008).

2.6.3 Bacteriocin

Bacteriocins are lantibiotics, heat-stable peptides or heat-labile proteins (Gram-positive bacteria bacteriocin groups) produced by one bacteria strain and acts against the closely related strains, mostly found in gram-positive bacteria. The lantibiotics group contain bacteriocins, that are modified after translation. The heat-stable peptides group contain small unmodified bacteriocins (Kadioglu et al., 2008). The third group of proteins produced by the Gram positive bacteria is the heat-labile proteins, which are composed of large bacteriocins (Giersing et al., 2016). The *Streptococcus pneumoniae* bacteriocins producing genes are induced by the competence stimulating peptide (CSP) and regulated by the *comDE* genetic locus (Iannelli et al., 2005).

Bacteriocins can be beneficial during colonization of *Streptococcus pneumoniae* by inhibiting the colonization of other competing and closely related bacteria to the same environment. Due to this ability, bacteriocin could be among the contributing factors of *Streptococcus pneumoniae* pathogenesis. Several genomes of the pneumococcus have been availed making it easy to analyze the genetic functions and regulation of the bacteriocins producing genes. Different bacteriocin clusters produce different amounts of toxins hence more analysis of the different strains producing bacteriocins is needed to better understand the pathogenicity of these proteins (Lux et al., 2007). The advantages of bacteriocin production to the pneumococcus is not only for bactericidal purposes but also for introduction of genetic material to the host environment hence promoting horizontal gene transfer. This enables recombination events enhancing the pathogenicity of the bacteria (Steinmoen et al., 2003).

2.6.4 CAAX protease

CAAX protease is an enzyme that is used for protein prenylation in the cell to enhance cell membrane attachment. The CAAX prenyl protease belong to a family of putative intra-membrane metalloproteases, together with their homologous prokaryotic bacteriocin-processing enzymes (Pei & Grishin, 2001). In *Streptococcus pneumoniae,* CAAX proteases are involved in bacteriocin processing (Lux et al., 2007). The activation and processing of bacteriocins and immunity proteins is also implicated by CAAX protease homologs, the *pncP and pncO* encoding proteins (Pei & Grishin, 2001).

2.6.5 Transcriptional regulator

Transcription regulation is the ability to control gene transcription, which is the conversion of DNA to RNA, hence regulating gene expression. The orchestration of gene expression is achieved through the multiplication or depletion of the production of various gene products. Transcription factors are proteins that contain DNA-binding domain (DBD), produced from transcription factor genes to perform the function of orchestrating specific gene activities (Brivanlou & Darnell, 2002). There are various transcriptional regulators identified necessary for *Streptococcus pneumoniae* virulence. There are some putative transcriptional regulators that aid in the survival of the bacteria in different host environments by regulating tissue specific virulence factors (Hava et al., 2003). The site of transcription initiation is the potential site of transcriptional regulator activity, where DNA promoter elements bind RNA polymerase and other proteins to activate transcription. Transcriptional regulators also regulate their own transcription (Hava et al., 2003).

2.6.6 Competence stimulating peptide

The competence of a bacteria is dependent on its ability to take up DNA from its neighboring environment. Competence stimulating peptide (CSP) is a heptadecapeptide molecule used for signaling and is produced and increased as the quorum sensing bacterial cells density increase, hence it is required for *Streptococcus pneumoniae* competency (Steinmoen et al., 2003). CSP is produced by propeptide, ComC (Competence stimulating peptide C), where ComC is the precursor form of the CSP and it is encoded by ComC gene, after which the mature CSP is transported by the ABC transporter (ComAB) to the outside of the cell (Y. Yang & Tal-Gan, 2019). Once the mature CSP attains its threshold, it binds and activates a membrane-localized histidine kinase receptor, ComD. In response, through phosphorylation, ComD activates its response regulator, ComE, which activates transcription of different genes by binding to their promoter regions. These genes include ComX, which encipher a sigma factor that leads to initiation of competence, and Com A B and Com C D E lead to the control of pneumococcal quorum sensing components hence integration of gene expression (Y. Yang & Tal-Gan, 2019). ComA is the membrane-associated peptide permease and ComE is the cognate transcriptional regulator (Merrifield et al., 2016).

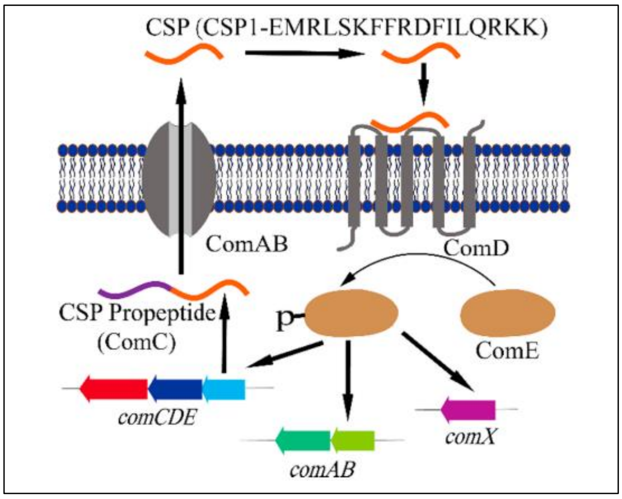


Figure 4: Competence stimulating peptide (CSP).

ComC produces CSP which is then transported to the outside of the cell by comAB. Mature CSP activates comD which then activates comE leading to transcription of comX, comAB and comCDE (Y. Yang & Tal-Gan, 2019).

2.6.7 Regulatory proteins

Regulatory proteins are enciphered by a capsular polysaccharide (CPS) gene cluster during the synthesis of the capsule, which is a major *Streptococcus pneumoniae* virulence factor. The biogenesis of the CPS relies on capsular regulatory proteins, CpsB, CpsC and CpsD. CpsB is the major regulatory protein since it's needed to dephosphorylate CpsD. In case where there is absence of CpsB, there occurs a rapid accumulation of phosphorylated CpsD which lead to marked decrease in the production of CPS. During drug development targeting the bacteria, CpsB is a relevant drug target due to its role in the biogenesis of CPS.

Regulatory proteins are the major proteins that are highly associated with *Streptococcus pneumoniae*, which enable adaption and resistance of the bacteria (Mücke et al., 2020). These systems are responsible for the transport of sugars and amino acids, and enable pneumococci to better adapt to their environment. The major function of this regulatory proteins is to detect a better environment that can support pneumococci (Gómez-mejia et al., 2018). Thus, understanding regulatory frameworks will help identify the major characteristics of antibiotics that will help in killing the bacteria during drug development.

2.6.8 Response regulator

In *Streptococcus pneumoniae*, different response regulators (RRs) have different roles in various strains. The cytosolic DNA-binding RR, which functions as a transcriptional regulator, together with histidine kinase, make the two-component signal transduction system that is vital in the survival of *Streptococcus pneumoniae* through the regulation of different cellular processes (Hendriksen et al., 2007). An autophosphorylating histidine kinase transfers the phosphor group to the response regulator, which leads to a change in the response regulator protein hence the initiation of the regulatory function (Hendriksen et al., 2007).

2.6.9 Foldase protein PrsA precursor

Foldases are a type of molecular chaperones that aid in correct, non-covalent, ATP-dependent folding of proteins during peptide structure formation (Mücke et al., 2020). This lead to proper activity of secreted proteins required in bacterial replication in host cells and spread to adjacent cells (Cahoon & Freitag, 2015).

2.6.10 Heat shock protein

Antibiotic resistance and tolerance are the global leaders in the increasing cases of antibiotic-resistance bacteria. Despite that many genes have been discovered to cause drug resistance and tolerance; heat shock protein has also contributed significantly to the low penicillin susceptibility (Hien et al., 2011). Heat shock protein contributes to this activity by modulating the cell wall and biosynthetic enzymes in the body. More notably, the action of heat shock protein enhances the activity of the bacteria in the body as it highly contributes to a higher level of resistance and severity of streptococcus pneumonia. A significant number of stresses more especially from antibiotics and DNA damage mainly causes the induction of heat shock protein. Pneumococcal heat shock proteins may be exposed to different level of fates as a result of the levels of stresses from the antibiotics (Hien et al., 2011).

2.6.11 Histidine kinase

Based on their domain organization, histidine kinases are divided into two groups, where in one, the N-terminal transmembrane region represents the sensor domain and the C-terminal transmitter domain contains the conserved histidine residue (Iannelli et al., 2005).

Competence transfer of genes across *S. pneumoniae* genome in the presence of CSPs involves the transmembrane histidine kinase comD, which activates the response regulator come (Iannelli et al., 2005). CSP binds onto the transmembrane histidine kinase comD receptor, which prompts autophosphorylation. The process therefore transfers the phosphoryl group into the response regulator comE. ComE, a DNA-binding protein is a part of the com CDE operon. The comE regulator protein interacts with components in a location near to that of the promoter of the com CDE operon, and therefore regulates its expression. Com CDE operon is involved in the transfer and transformation of genes across the *S. pneumoniae* genome.

The ComD receptor possess similar organizational structure to that of histidine kinases with the N-terminal representing the sensor domain. ComD receptor is responsible for the specificity of the competence phenotype in *S. pneumoniae* (Iannelli et al., 2005). The introduction of synthetic CSP1 and CSP2 to induce competence revealed a pattern of specificity that led to the conclusion that the histidine kinase comD receptor is responsible for specificity of the competence pherotype of *S. pneumoniae* (Iannelli et al., 2005). The histidine kinase comD receptor characteristics and functionality could be used to control the specificity of competence transfer of genes across *Streptococci* by abrogating the activation of the comE regulator (G. Yang et al., 2015). This technique that prevents transformation of genes in the genome could therefore be used in the formulation of pneumococcal vaccines for the prevention of pneumococcal disease (Zhu & Lau, 2011a).

2.6.12 NADH oxidase

Streptococcus pneumoniae attacks and resides in tissues with reduced oxygen levels and in some cases in areas with complete absence of oxygen. To survive in the respiratory organs, the pathogen adapts in ways similar to other anaerobic pathogens, by possessing enzymes that catalyze the oxidation of O2. NADH oxidase is an enzyme purified as a soluble flavoprotein, that detoxifies *S. pneumoniae* environment by catalyzing the reduction of molecular oxygen into water (Auzat et al., 1999). The presence of NADH oxidase in *S. pneumoniae* is essential for the regulation of competence for genetic exchange through metabolisms. The presence of NADH oxidase allows *S. pneumoniae* to re-oxidize part of the glycolytic NADH oxidase with oxygen rather than with pyruvate. This further improves the effectiveness of glucose catabolism. The ability of the enzyme to catalyze its own NADH gives it a metabolic advantage. This process utilizes oxygen as a substrate. Therefore, NADH oxidase enzyme senses the presence of oxygen and transduces the signal into metabolic changes that alter the physiology of the bacteria (Yu et al., 2001).

NADH oxidase in *S. pneumoniae* functions as an adhesin, aiding the adhesion process of the bacteria to the targeted mammalian cells (Muchnik et al., 2013). The bacteria encapsulate during adhesion to the targeted cells. Recombinant NOX (rNOX) interferes with the encapsulating bacteria and prevents its attack on the target cells. Concentrating the target cells with rNOX is a likely possibility of preventing the adhesion of the bacteria to A549 cells. This could also be achieved by neutralizing NOX enzyme residing in the cell-walls with an anti-rNOX anti-serum. Using this knowledge, scientists could derive vaccines and medicine that reduce the habitation of the *S. pneumoniae* in respiratory organs.

2.6.13 Enolase

The enolase surface protein is classified among the anchorless proteins of *S. pneumoniae*. The protein has also been identified as a plasminogen-binding protein, sufficiently enabling pneumococcal colonization of the mucosal cells (Bergmann et al., 2001; Kolberg et al., 2006). Plasminogen binding degrades the extracellular matrix proteins and further enables the migration of bacteria across the human extracellular matrix. This protein gains the advantage of withstanding the counter activity of polyclonal rabbit anti-enolase antibodies. The procedure revealed that a low amount enolase on the surface of vitro-grown pneumococci.

Under scientific experiment, a low amount of surface-exposed protein of enolase on vitro-grown pneumococci suggested that enolase could not be efficient enough for the development of vaccines against *S. pneumoniae*. However, the process of pneumococcal colonization being multifactorial, might affect the behaviour of enolase proteins as well as other proteins (Kolberg et al., 2006). The small amount existing on the vitro-grown pneumococci is enough to generate pneumococcal pathogens since the protein's efficiency of plasminogen binding is very high (Kolberg et al., 2006).

2.6.14 Sialidase A

Sialidases are one of the virulence factors that enable faster growth. Sialidase is divided into three major types; NanA, NanB and NanC. Pneumococcal sialidases play a major role in colonization by bacteria especially in crossing the blood-brain-barrier to infect the meninges (Xu et al., 2011). Sialidase A has played an essential role in the determination of the best drugs that can be used to eradicate the colonization action of the bacteria by inhibiting the removal of sialic acid (Xiao et al., 2019).

2.6.15 Pneumolysin

Pneumolysin (Ply) is a pore forming toxin known as a major virulent factor across all serotypes of the *S. pneumoniae,* and it facilitates crossing of the blood-brain-barrier (Hirst et al., 2004). Ply lacks the N-terminal which is a signal peptide for export hence its release is depended on autolysis of the Pneumococci which happens mostly during the stationary growth phase of the bacteria (Martner et al., 2008; Price & Camilli, 2009). Autolysins (ALs) aid in the lysis of the cell wall for the release of Ply. The major AL is the N- acetyl-muramoyl-1-alanine amidase (LytA) and the others include LytB and LytC (Martner et al., 2008). During the early stages of infection, when Ply is in low concentrations, it causes some effects on the host cell, including cell apoptosis, activation of the complementary system and modulation of a proinflammatory state to the immune system. In the late stages of infection, when Ply is in high concentrations, it causes lethal effects to host cells which include direct tissue damage (Hirst et al., 2004). Due to its effects, Ply leads to increased bacteria invasion to the brain tissue since the microglia are already actively involved in neuroinflammation as well as their main job which is protection of the central nervous system (CNS) (Hupp et al., 2019). There is already several research projects developing Ply targeting vaccination strategies and antibacterial chemotherapy (Cockeran et al., 2005). The pneumococcal toxin can also be applied in preventive strategies.

2.6.16 Serine protease

Serine protease orthologs are the major factors that contribute to virulence of the pneumococcus. The major serine protease that is involved in *Streptococcus pneumonia* virulence is the high-temperature requirement A (HtrA), which play a role in competence by helping in the survival of host environmental stress. Other serine proteases include the cell-wall associated serine protease A (PrtA) which is involved after intraperitoneal sepsis and subtilase family protein (SFP) (Stoppelaar et al., 2013).

2.6.17 Pyruvate oxidase

Pyruvate oxidase is a protein encoded by the SpxB gene and acts to decarboxylate pyruvate to form acetyl phosphate, hydrogen peroxide (H_2O_2) and carbon dioxide (CO_2) (Spellerberg et al., 1996). Although the H_2O_2 obtained from pyruvate oxidase is harmful to *Streptococcus pneumoniae* survival when in high concentrations, it also helps the pneumococcus to compete for the host environment against other inhabitants (Bryant et al., 2016). Pneumococcal H_2O_2 is also toxic to eukaryotic host cells and during the stationary growth phase of the pneumococcus, it leads to apoptosis-like cell death (Bryant et al., 2016).

2.6.18 Sortase

Sortase A (srtA) is an enzyme found on the gram-positive bacteria cell surface, which plays a roles in the virulence of *Streptococcus pneumoniae* by anchoring specific protein to the cell wall envelop (Kumari et al., 2020; Paterson & Mitchell, 2006). The attachment of these proteins is done through the processing of sorting signals at the recognition (LPXTG) motif (Gianfaldoni et al., 2009). SrtA is also involved in the colonization and the pathogenesis of the pneumococcus through adhesion to host cells (Gianfaldoni et al., 2009).

2.6.19 Serine/ Threonine kinase protein

Serine/ Threonine kinase is a protein encoded by the StkP gene (Echenique et al., 2004), located on the pneumococcus cell membrane with an N terminal and C terminal. The N terminal kinase domain extends towards the cytoplasm, the C terminal extends towards the extracellular region with four Penicillin-binding protein and Ser/Thr protein kinase Associated (PASTA) domains, and a short transmembrane region exists between the two terminals (Pi et al., 2018). The pneumococcus StkP play an essential role in virulence among other cellular processes. It also aids in the survival of the bacteria as well as its resistance to different stress conditions (Saskova et al., 2007).

2.6.20 Iron-compound ABC transporter

Pathogenic bacteria depend on the acquisition of iron for survival. S. *pneumoniae* possess two genetic loci, pneumococcal iron uptake (piu) and pneumococcal iron acquisition (pia) that encode for homologues of ABC transporters that are necessary for iron uptake for the bacteria (Brown et al., 2002). The lipoprotein components piuA and piaA encoded by the iron ABC transporters piu and pia are required for *Streptococcus pneumoniae* virulence (Jomaa et al., 2005). The iron ABC transporters function differently with pia, the dominant one, being responsible for hydroxamate siderophores transport and piu being responsible for heme transport from hemoglobin (Cheng et al., 2013; Whalan et al., 2005). Presence of single mutations in the piu or pia loci result to the genome's reduced ability to source iron molecules from hemoglobin (Brown et al., 2002). In vitro, the effect of the double mutation is attenuation of virulence (Brown et al., 2002).

2.6.21 Peptide pheromone

Pheromones are quorum sensing compounds that regulate the competence for genetic transformation in bacteria (Manuscript, 2014; Morrison, 1996; Piccoli et al., 1996). *Streptococcus pneumonia* genome has acquired characteristics that make it antibiotic resistance, a major concern in the medical research field. The attainment of the antibiotic resistance characteristics is regulated by a peptide pheromone (BlpC), competence-stimulating peptide (CSP) (Zhu & Lau, 2011b). The peptide binds to a ComD receptor, which then initiates its cognate transcriptor factor ComE to prompt DNA uptake into the S. pneumonia genome. The receptor also controls the genome's virulence factors expressed during infection. The peptide pheromone plays an important role in protecting the pneumococci as well as in propelling the genome through adherence with epithelial cells and colonization of the pharynx (Zhu & Lau, 2011b). DNA transformation into the genome, regulated by the peptide pheromone is necessary to gain the genome virulence and antibiotic resistance collectively obtained from other species. Competitive inhibitors of CSPs efficiently inhibits horizontal gene transfer from other species as well as attenuates virulence in the genome. The use of peptide analogues prevents the transfer of genes from other species of the genome. Gene transfer across different species of the genome enables the acquisition of antibiotic resistance characteristics (Piccoli et al., 1996; Pinchas & Lacross, 2015; Zhu & Lau, 2011b).

2.6.22 Pneumococcal histidine triad (D and E)

Pneumococcal histidine Triad protein D and E, (PhtD and PhtE) are among Pht proteins necessary for the adherence of pneumococcal strains to the pharyngeal and mucosal surface (Melin et al., 2010). PhtD is highly conserved in *Streptococcus pneumoniae* strains as well as being a virulent factor of the bacteria (Ochs et al., 2016). Both the PhtD and PhtE have been considered over time as vaccine candidates since they produce a protective immunity against sepsis, although PhtE is less effective (Plumptre et al., 2013).

The proteins also promote the colonization of the pharynx region by the pneumococcal strains. The conserved surface proteins are associated with the following functions; defending epithelial cells against complement deposition, consuming zinc ions as well as helping in the adherence of pneumococcal bacteria to epithelial cells in the nasopharynx (Ogunniyi et al., n.d.). The protein PhtE among other polyhistidine proteins produce immunogenic impact when delivered in vaccine formulations by reducing the adhesion of bacteria to the epithelial cells (Ogunniyi et al., n.d.).

2.6.23 Pneumococcal surface protein A

Pneumococcal surface protein A (PspA) is a protein found on the surface of the pneumococcus cell wall and acts as virulence factor for the bacteria (Hollingshead et al., 2006; Leonor et al., 2003; Tu et al., 1999). PspA has three subsets, family 1, family 2, and family 3 according to the DNA and protein sequence variability. Each family subset has one or more clades with each clade representing a PspA antigen (Baril et al., 2006). PspA attaches itself to phosphorylcholine residues on the host cell wall as the N-terminus is exposed to the surface of the bacteria (Pujanauski et al., 2020). Lack of PspA in the pneumococci reduced the bacteria's virulence, since the bacteria became more susceptible to deactivation by complement component C3 deposition (Tu et al., 1999).

Recombination events in pneumococcal strains enable them to escape the host immune response by swapping antigens that elicit an immune response, with those that do not. These properties also enable the acquisition of drug resistance and survival in different environments of the host body compartments. There is still limited understanding on the genetic differences between meningitis and bacteremia associated strains. This study examined the difference in genetic patterns of gene acquisition and loss that contribute to the propensity to cause meningitis compared to bacteremia.

CHAPTER 3

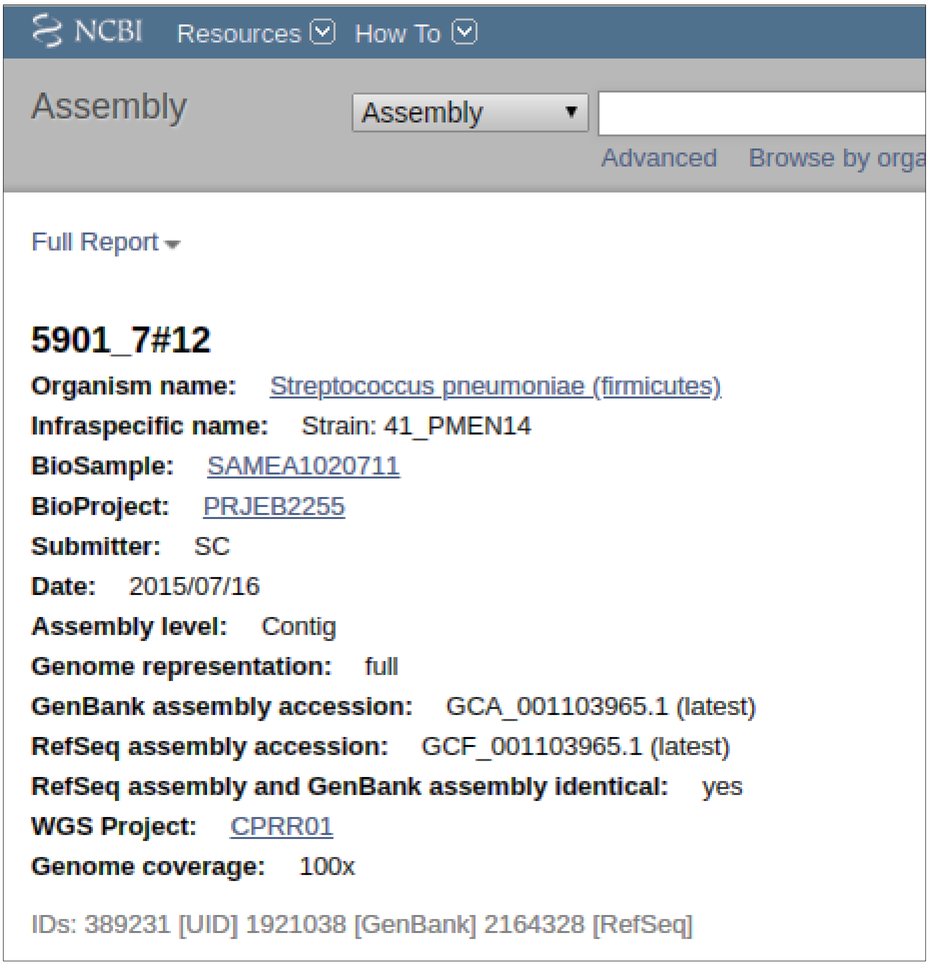
MATERIALS AND METHODS

3.1 Data retrieval and preprocessing

The data was obtained from the National Center for Biotechnology Information (NCBI) GenBank database (https://www.ncbi.nlm.nih.gov/) for each strain and cleaned. All missing variables were identified and marked with "N/A". The records contained the metadata for 209 strains of *Streptococcus pneumoniae* among which 147 strains were from human blood and 62 strains were from human cerebrospinal fluid. The data retrieved intentionally included data from all available geographical locations which gave a global overview of the study. The metadata included assembly number, whole genome shotgun (WGS) sequencing accession number, biosample number, stain taxonomy number and serotype.

The specification of geographical location allowed the inclusion of a global set of pneumococci, since different countries and continents have differing prevalence in serotypes and sequence types associated with invasive pneumococcal disease.

GenBank files were retrieved from the NCBI site in the GenBank file format (.gbff). The GenBank files contained the whole genome sequence for each strain. The strain taxonomy number was included in the metadata. The first step of pneumococcal genome data retrieval used the strain taxonomy number to search from all datasets, then narrowed the search down to assembly database which gave a full assembly report. The full assembly report included the GenBank assembly accession number, which was recorded alongside the strain taxonomy number to link the GenBank files data with their respective strains metadata. The full assembly report also included the WGS project link which led to the nucleotide database, with the whole genome shotgun sequencing project of the strain. The record of the whole sequencing project had a WGS sequencing link that led to the GenBank file download link which had the complete genome sequence with a '.gbff.gz' extension. All the files were downloaded for all strains with a complete genome.

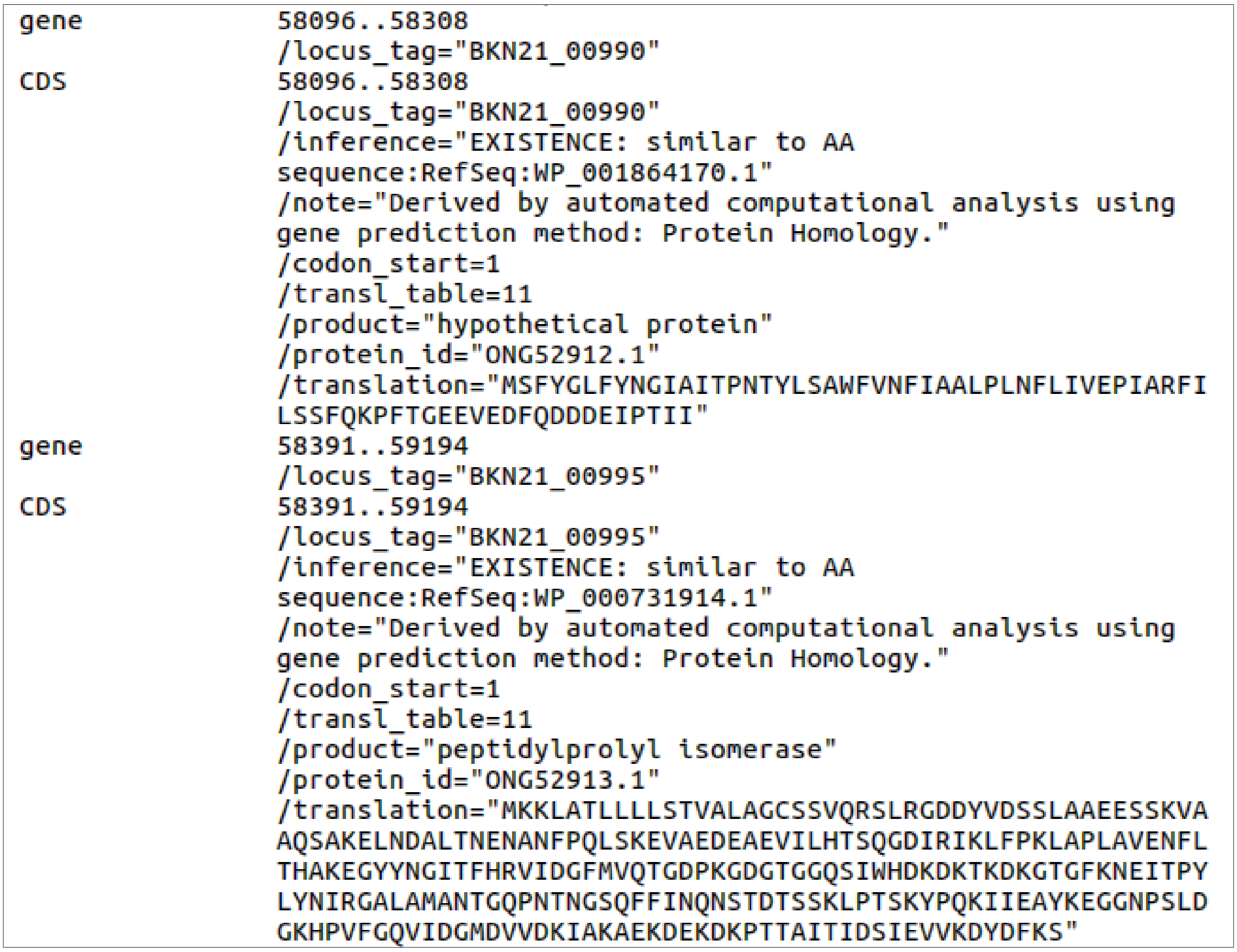


5901_7#12

Organism name: Streptococcus pneumoniae (firmicutes)
Infraspecific name: Strain: 41_PMEN14
BioSample: SAMEA1020711
BioProject: PRJEB2255
Submitter: SC
Date: 2015/07/16
Assembly level: Contig
Genome representation: full
GenBank assembly accession: GCA_001103965.1 (latest)
RefSeq assembly accession: GCF_001103965.1 (latest)
RefSeq assembly and GenBank assembly identical: yes
WGS Project: CPRR01
Genome coverage: 100x

IDs: 389231 [UID] 1921038 [GenBank] 2164328 [RefSeq]

Figure 5: National Center for Biotechnology Information (NCBI) assembly database page. The assembly page contains a biosample link, the latest GenBank assembly accession number and the WGS project link. Through the biosample link, the extra metadata about the strain can be accessed, including sample collection date, geographical location, host age and host disease.



```
gene                58096..58308
                    /locus_tag="BKN21_00990"
CDS                 58096..58308
                    /locus_tag="BKN21_00990"
                    /inference="EXISTENCE: similar to AA
                    sequence:RefSeq:WP_001864170.1"
                    /note="Derived by automated computational analysis using
                    gene prediction method: Protein Homology."
                    /codon_start=1
                    /transl_table=11
                    /product="hypothetical protein"
                    /protein_id="ONG52912.1"
                    /translation="MSFYGLFYNGIAITPNTYLSAWFVNFIAALPLNFLIVEPIARFI
                    LSSFQKPFTGEEVEDFQDDDEIPTII"
gene                58391..59194
                    /locus_tag="BKN21_00995"
CDS                 58391..59194
                    /locus_tag="BKN21_00995"
                    /inference="EXISTENCE: similar to AA
                    sequence:RefSeq:WP_000731914.1"
                    /note="Derived by automated computational analysis using
                    gene prediction method: Protein Homology."
                    /codon_start=1
                    /transl_table=11
                    /product="peptidylprolyl isomerase"
                    /protein_id="ONG52913.1"
                    /translation="MKKLATLLLLSTVALAGCSSVQRSLRGDDYVDSSLAAEESSKVA
                    AQSAKELNDALTNENANFPQLSKEVAEDEAEVILHTSQGDIRIKLFPKLAPLAVENFL
                    THAKEGYYNGITFHRVIDGFMVQTGDPKGDGTGGQSIWHDKDKTKDKGTGFKNEITPY
                    LYNIRGALAMANTGQPNTNGSQFFINQNSTDTSSKLPTSKYPQKIIEAYKEGGNPSLD
                    GKHPVFGQVIDGMDVVDKIAKAEKDEKDKPTTAITIDSIEVVKDYDFKS"
```

Figure 6: Gene ID, product names and their respective protein sequences. These detailed datasets contain information of each strain of the *Streptococcus pneumoniae*.

The GenBank files were separated into two subsets: meningitis-associated strain files and bacteremia-associated strain files placed in different folders for easier computing. These data subsets were used for comparisons between human cerebrospinal fluid isolates and human blood isolates to determine the specific genetic features in each set.

3.2 Data analysis

A customized Perl script (Appendix 2) was first used to covert GenBank file format files to European Molecular Biology Laboratory (EMBL) file format and another script (Appendix 2) were used to covert EMBL files to FASTA file format. The files were concatenated prior to performing an all-vs-all BLAST.

An All-versus-All BLAST, is a BLAST sequence search where each sequence was compared to all other sequences in the data sub-set as illustrated in Figure 7.

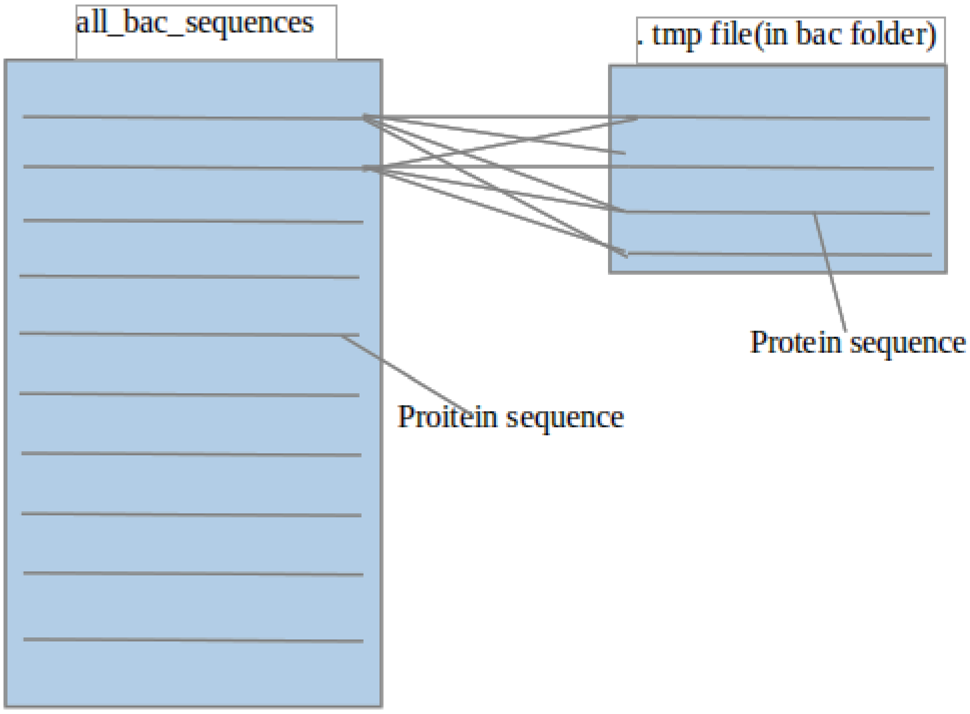


Figure 7: All vs All BLAST. Each bacteremia file contained several protein sequences which were concatenated to create a list of all bacteremia sequences. Each sequence in the list was compared with all the sequences in each bacteremia file to form an All-vs-All BLAST database.

All-vs-All BLAST database was generated which was run in OrthoMCL with default settings to generate an initial orthologous map. The BLAST results were further analyzed using

OrthoMCL using default settings (BLASTP E-value cut-off 1e-5 and inflation index 2.5) to assign sequences into clusters based on homology (orthologs and paralogs) (Li, Stoeckert, & Roos, 2003). The OrthoMCL results were converted to comma-separated values (CSV) format, displaying clearly the genes and their clusters. An example is shown in Table 6. The OrthoMCl files for meningitis and bacteremia can be found in Appendices 6 and 7 respectively.

Table 6: OrthoMCL clusters demonstration and their respective genes.

In each cluster, different genes have same taxa hence more gene and fewer taxa. Each raw represent a list of genes from different genomes in one cluster.

OrthoMCL Cluster	No. of Genes	No. of Taxa	Gene ID (Genome)	Gene ID (Genome)	Gene ID (Genome)
Cluster0	187 genes	62 taxa	gene1 (genome1)	gene8 (genome1)	gene15 (genome2)
Cluster1	181 genes	62 taxa	gene2 (genome4)	gene9 (genome3)	gene16 (genome1)
Cluster2	136 genes	61 taxa	gene3 (genome2)	gene10 (genome4)	gene17 (genome2)
Cluster3	130 genes	54 taxa	gene4 (genome5)	gene11 (genome4)	gene18 (genome3)
Cluster4	130 genes	49 taxa	gene5 (genome3)	gene12 (genome5)	gene19 (genome3)
Cluster5	124 genes	62 taxa	gene6 (genome5)	gene13 (genome2)	gene20 (genome4)
Cluster6	124 genes	62 taxa	gene7 (genome4)	gene14 (genome1)	gene21 (genome5)

The *Streptococcus pneumoniae* ATCC 700669 genome was used as reference for annotation.

3.2.1 Annotation using a reference genome

An annotation list was first generated from the *Streptococcus pneumoniae* ATCC 700669 complete genome, for each gene. Annotations were then transferred to the ortholog clusters from the *Streptococcus pneumoniae* ATCC 700669 genome.

The annotated datasets were further processed using BMX to organize them into files with phyletic patters 1's and 0's representing presence and absence of genes respectively Table 7 (Kulohoma, 2015). The analysis with BMX first allowed the definition of the core genes, which consists of genes shared by both the meningitis and bacteremia set (n=1937 genes); and a set of genes that were unique to either the meningitis (n=351 genes), and bacteremia (n=620 genes) data subsets. This enabled the identification of orthologs and paralogs and helped to distinguish the core and accessory genomes.

Table 7: OrthoMCL clusters and the respective binary representation of absence or presence of genes. 1t represents presence of genes in different strains genomes' and 0s represent absence of genes in different strains genomes.

Cluster	Genome1	Genome2	Genome3	Genome4	Genome5
0	1	1	0	0	0
0	1	0	0	0	0
1	1	0	1	1	0
2	0	1	0	1	0
2	0	1	0	0	0
3	0	0	1	1	1
4	0	0	1	0	1
4	0	0	1	0	0
5	0	1	0	1	1
6	1	0	0	1	1

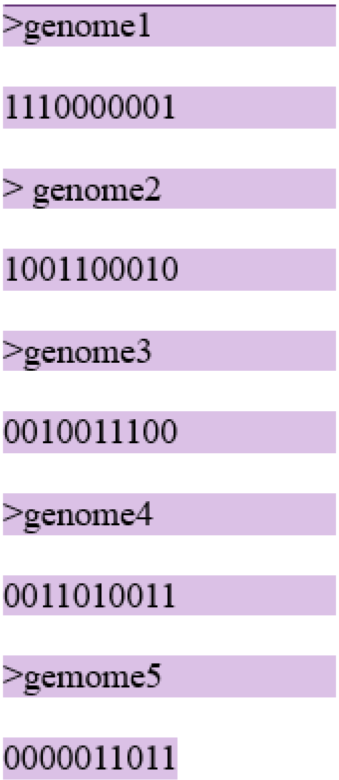
These datasets were transposed (Table 8) and then converted into FASTA format phyletic pattern files (Figure 8), which is the input format for GLOOME analysis.

A cluster may have two sets of values in cases where the BMX program distinguishes orthologs from paralogs, hence if they are present, the cluster line is repeated**.**

Table 8: Transposed data with orthoMCL clusters and binary form of gene presence or absence. 1t represents presence of genes in different strains genomes' and 0s represent absence of genes in different strains genomes.

Cluster	0	0	1	2	2	3	4	4	5	6
Genome1	1	1	1	0	0	0	0	0	0	1
Genome2	1	0	0	1	1	0	0	0	1	0
Genome3	0	0	1	0	0	1	1	1	0	0
Genome4	0	0	1	1	0	1	0	0	1	1
Genome5	0	0	0	0	0	1	1	0	1	1

The data was reorganized into binary FASTA files as shown in Figure 8 below:



```
>genome1
1110000001
> genome2
1001100010
>genome3
0010011100
>genome4
0011010011
>gemome5
0000011011
```

Figure 8: Binary FASTA arrangement demonstration. The 1s represent gene presence and 0s represent gene absence. Genome 1 to genome 5 represent all the genomes that were studied.

3.2.2 Gain Loss Mapping Engine (GLOOME) analysis

In every genome, from the first position to the last, aligns data showing the genes present as 1's, and genes absent as 0's. This gives us a matrix of absence and presence of genes across the entire genome of every single strain studied. The probability of gain and loss is required to determine the genetic flux, from all the strains. Genetic flux explains the probability of gaining or losing a gene along the genome, and provides an indication of the adaptability to survive in each niche, either blood or CSF. GLOOME produces results in the form of an easy to visualize matrix, with probabilities showing whether there is likely to be gene gain or loss at a particular locus (Benard & Kulohoma, 2012). The statistical representation of the GLOOME data showed the probability of gain and loss adjoining the position along the genomes for both the meningitis data set and the bacteremia data set.

Each of the meningitis and bacteremia data sets produced from GLOOME were annotated to produce comma separated version files containing cluster number, Annotation (gene ID), product (Gene name), position, expectation of gain, expectation of loss, gain and loss. During annotation, the shared meningitis and bacteremia file was split into two files, one as the shared meningitis data file and the second as the shared bacteremia data file, for easier and independent analysis. The cumulative expectation of gain and loss events was computed to aid in construction of visualization graphs.

3.3 Visualization of the GLOOME results

Huge chunks of data can be troublesome when interpreting, hence a summary of the data should be made, to make it more understandable. The results obtained in this study are no exception, a reason for summarization and image representation as a way of making the results easier for understanding. Images make the entire information, however large, to be viewed in one glance and as well easily understood. To make more sense of the results obtained in this study, the data was represented in line graphs and box plots.

CHAPTER 4

RESULTS

4.1 Gene gain and loss events along *Streptococcus pneumoniae* genomes

Gene gain and loss events were quantified with color codes as shown in Figure 9 and Figure 10 below.

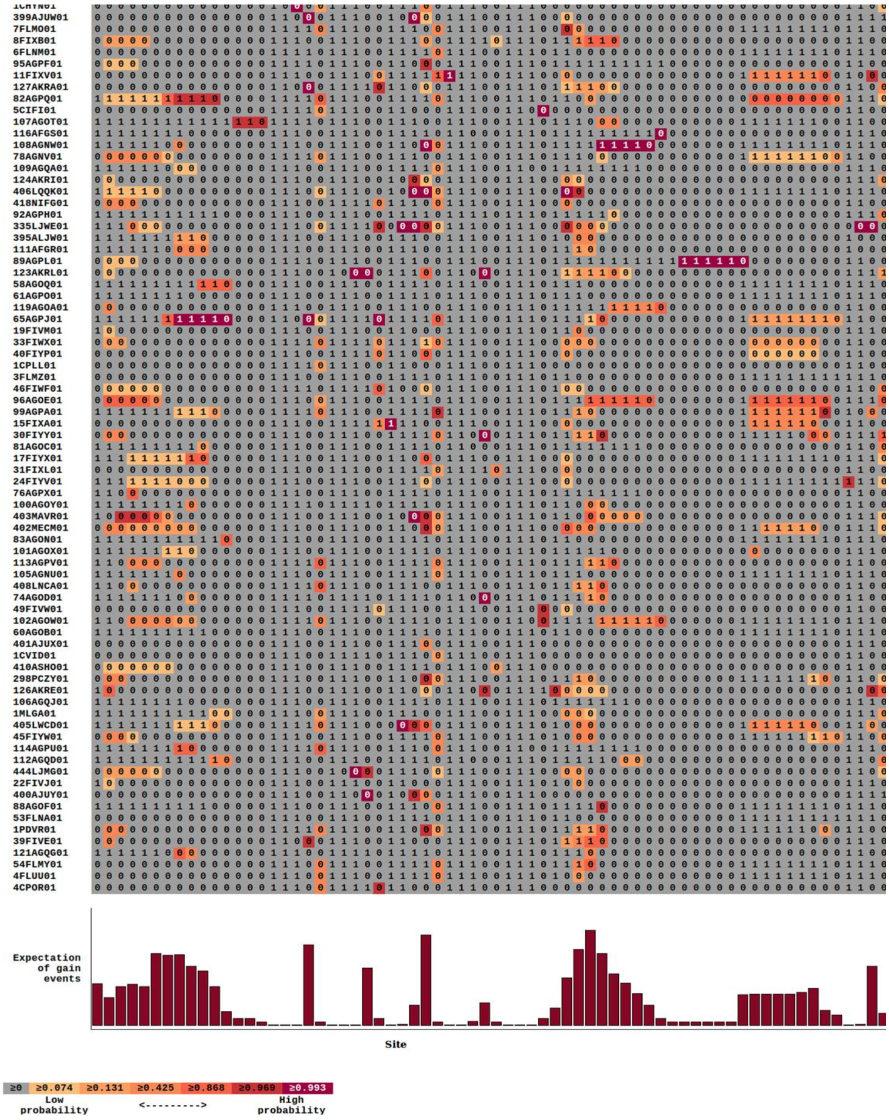
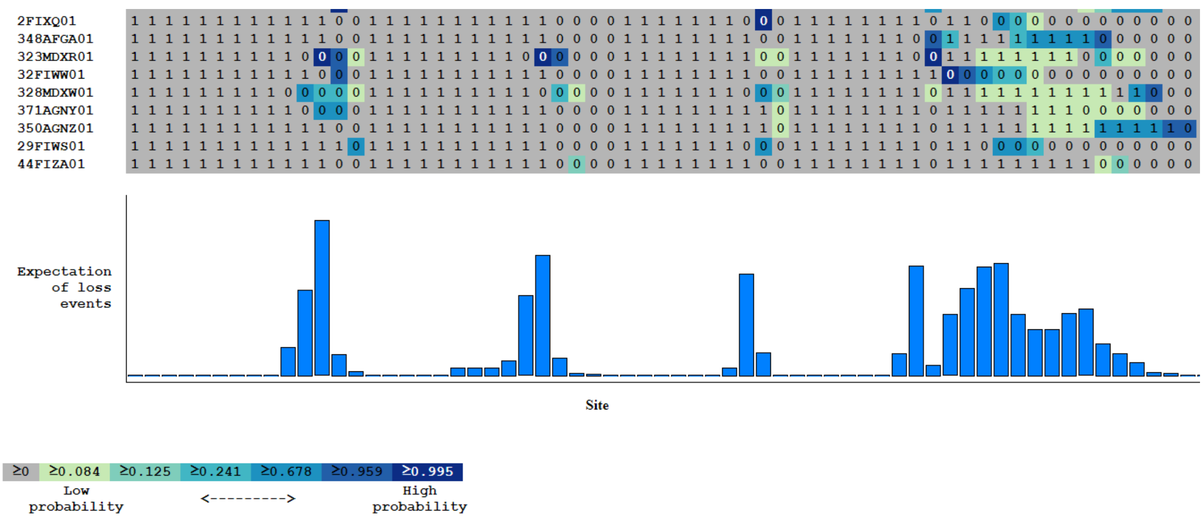


Figure 9: Color-coded gene gain demonstration data produced from GLOOME. The first column shows all the genomes, displayed using the file names in which each strain was

contained. The columns on the color-coded binary matrix displays the position along each genome and the rows represent the genes gain or loss on each genome for all the clusters. The expectation of gene gain events was shown on different loci, coded from a dark red color to show high probability and yellow color to show low probability. The grey color showed 0 (zero) probability of gene gain.



```
2FIXQ01    1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0
348AFGA01  1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0
323MDXR01  1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0
32FIWW01   1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
328MDXW01  1 1 1 1 1 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0
371AGNY01  1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0
350AGNZ01  1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0
29FIWS01   1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0
44FIZA01   1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0
```

Figure 10: Color-coded gene loss demonstration data produced from GLOOME. The gene loss events were coded from high probability with dark blue color to low probability with green color. The grey color showed 0 (zero) probability of gene loss.

4.2 Probability of *Streptococcus pneumoniae* gene gain and loss

Bacteremia and meningitis strain genomes were analyzed separately, to identify differences in the pattern of gene gain and loss. These differences highlight mechanisms that allow certain strains to invade the CSF (Benard Kulohoma, 2012). The GLOOME output was the probability of gene gain and loss for each locus represented in the phyletic map. The average of the expectations of gain or loss for each of the loci was established to determine the average probability of either gene gain or loss per site (Table 9). These average values helped to determine the magnitude of gene gain or loss across the entire genome for the strains in the subset under consideration. Comparisons of the average expectation of gene gain and loss per site enabled the evaluation of the direction of evolution for the dataset under consideration. It was observed that there were more gene loss than gene gains events, with the exception of the unique set of accessory genes in the bacteremia sub-set.

Table 9: Average expectation of gene gain or loss events per site in unique and shared genes in bacteremia and meningitis strains. Gene gain and loss data was extracted from the expectation of gain and loss events data, by converting all values below 0.95 to 0s and above 0.95 to 1s.

	Average expectation of Gene Gain per site	Average expectation of Gene Loss per site	Gene gain over Gene loss
Unique Meningitis	3.9761123077	4.0265632051	0.9874704817
Unique Bacteremia	3.4637956157	2.5514061667	1.3576025883
Shared Meningitis	2.4366164663	2.9449318159	0.8273931686
Shared Bacteremia	3.0187111763	3.3940442662	0.8894142031

Confounding in the bacteremia set was reduced by matching its size to that of the meningitis set. Two subsets (n=62 genomes) of randomly selected genome strains were analyzed and compared to the meningitis set. (Table 10). In general, the results were consistent with those from the previous analysis. Gene loss events were greater than gene gain events in the meningitis subset, the opposite was true for the bacteremia subset.

Table 10: Average expectation of gene gain or loss events per site in 62 strains sets of meningitis and bacteremia genomes. The first meningitis and bacteremia rows represent the originally analyzed shared meningitis and shared bacteremia data. The second meningitis data was analyzed together with the randomly selected two bacteremia sets.

	Average expectation of Gene Gain per site	Average expectation of Gene Loss per site
1 Meningitis	2.44	2.94
1 Bacteremia	3.02	3.39
2 Meningitis	2.40	2.78
2Bacteremia II	1.57	1.43

4.2.1 *Streptococcus pneumoniae* virulence factors

The Figure 11 below highlights some of the important virulent factors and their expectations of gene gain or loss in the meningitis and bacteremia subsets. This provides a good indication on whether they should be prioritized for incorporation into vaccines and as potential antibiotic target sites.

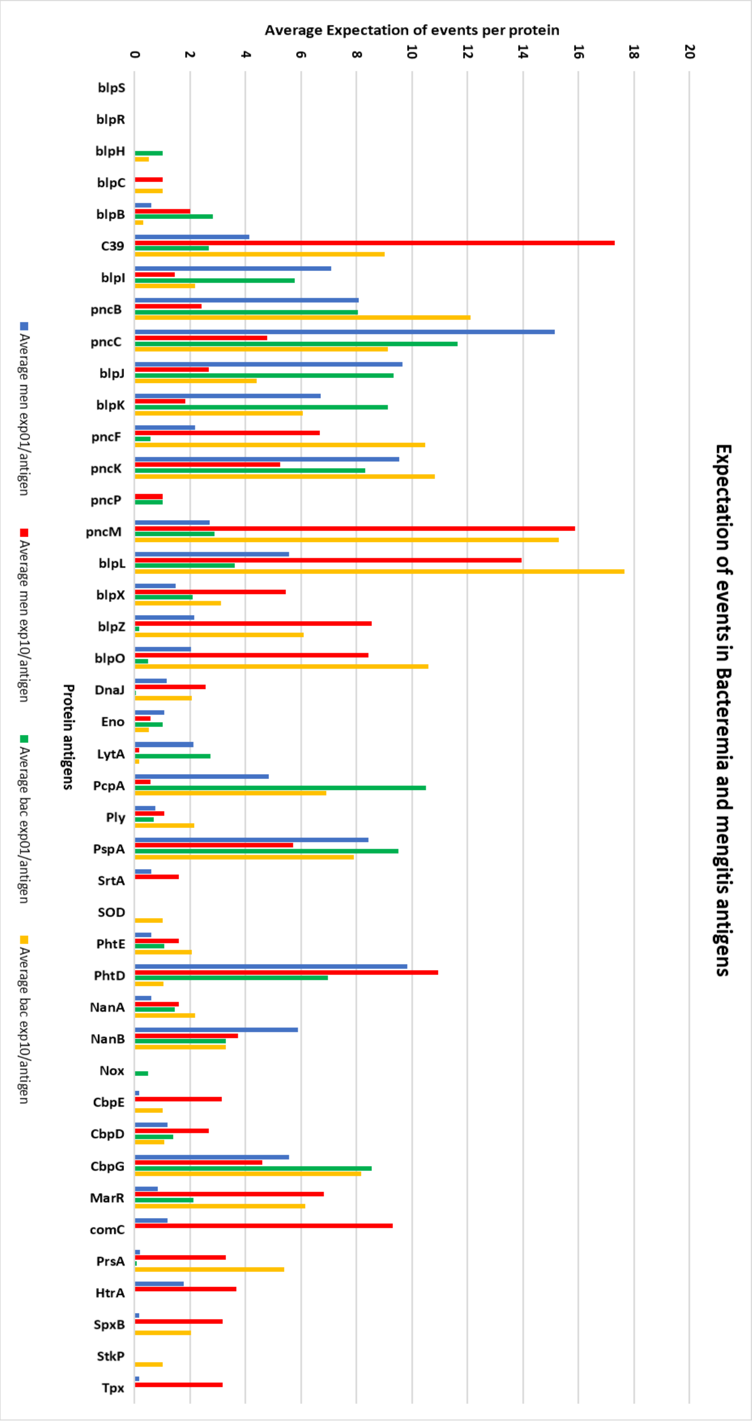


Figure 11: Important virulence factors and their expectations of gene gain or loss in the meningitis and bacteremia subsets. The ‘exp01 Men’ and ‘exp10 Men’ indicates the

expectation of gain and expectation of loss in meningitis dataset respectively. Same applies to the ‘exp01 Bac’ and ‘exp10 Bac’ in bacteremia.

The putative bacteriocin transporter C39 protease domain BlpA2 and pneumococcal histidine triad protein D (bvh-11-2) showed more gene loss events in the meningitis set. The immunity protein PncB, pncF, immunity protein PncK and bacteriocin BlpO displayed more gene loss events in the bacteremia set. More gene loss events were observed in both bacteremia and meningitis sets for putative immunity protein PncM and putative membrane protein BlpL. Also, more gene gain in both sets was observed for putative uncharacterized protein PncC. There was more gene gain in bacteremia set for cell surface choline binding protein PcpA.

4.2.2 Comparison of genetic flux within data subsets

There was a total of 62 strains associated with meningitis that were analyzed. Unique meningitis associated genes were 351, while the shared meningitis associated genes were 1937. The total meningitis genes were 2288.The set was analyzed twice through GLOOME, giving the expectation of gain and expectation of loss of gene in every position along the genome. The total cumulative expectation of gene gain was 7702.14465, total cumulative expectation of gene loss was 9308.92947, average expectation of gene gain was 2.44, and average expectation of gene loss was 2.94 with a ratio of 0.83 for shared meningitis strains. The total cumulative expectation of gene gain was 310.13676, total cumulative expectation of gene loss was 314.07193, average expectation of gene gain was 3.98, average expectation of gene loss was 4.03 with a ratio of 0.99 for unique meningitis strains.

This information highlights that there are more gene loss compared to gene gain events as shown in Figure 12.

A total of 147 strains associated with the propensity to cause bacteremia were analyzed. Unique bacteremia associated genes were 620, while the shared bacteremia associated genes were 1937. The total bacteremia genes were 2557. Despite that the cumulative data of the unique bacteremia sets show otherwise, the overall gene loss is high with low probability of gene gain. The unique bacteremia set has more gene gain and low gene loss probability. The total cumulative expectation of gene gain was 748.179853, total cumulative expectation of gene loss was 551.103732, average expectation of gene gain was 3.46, and average expectation of gene loss was 2.55 with a ratio of 1.36 for unique meningitis strains.

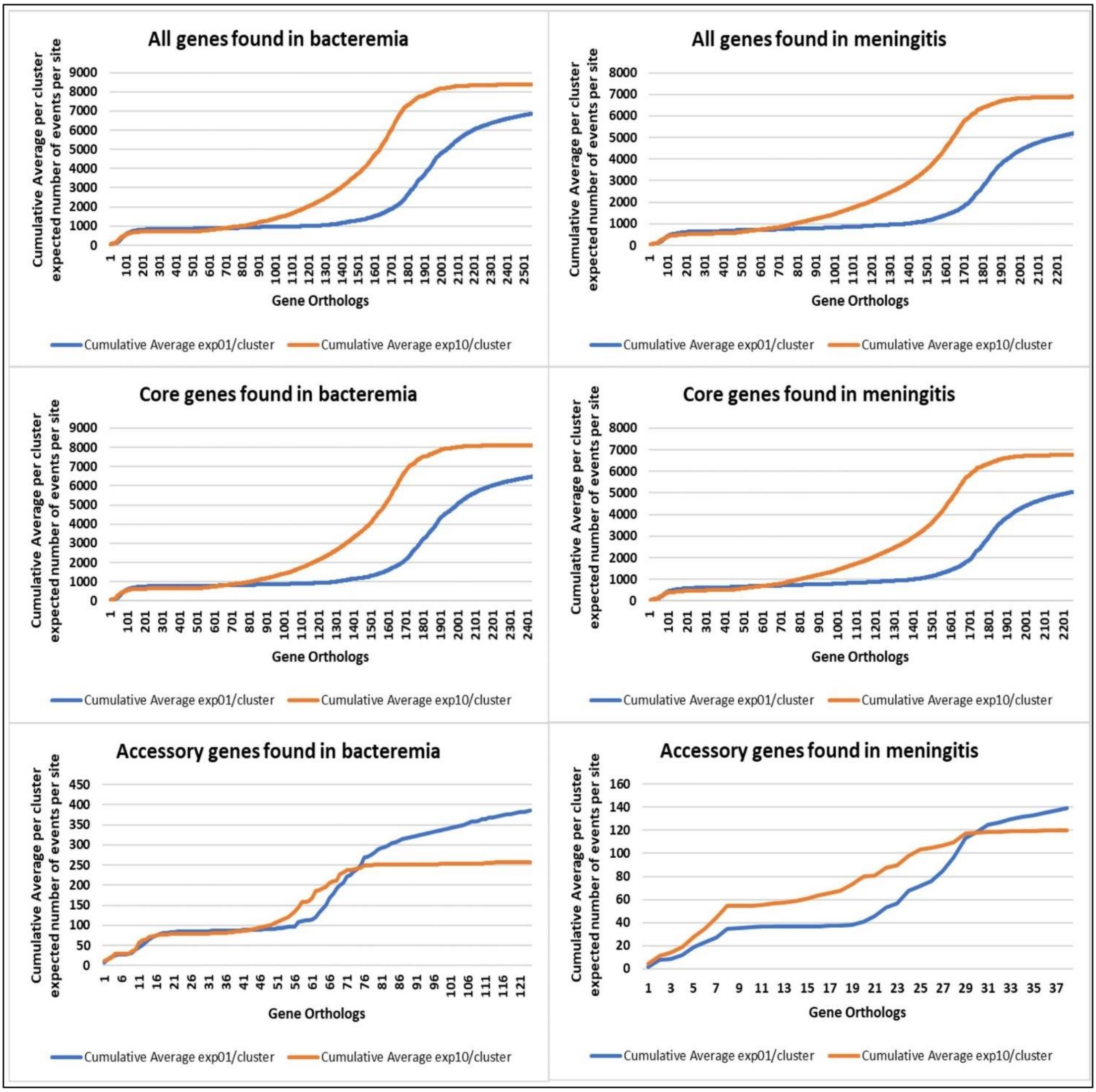


Figure 12: Cumulative expected number of events per site.

Generally, there was more gene loss compared to gene gain events. However, in bacteremia dataset, the accessory genes had more gene gain compared to gene loss events.

Apart from genes uniquely identified in the meningitis set, there was a significant different in genetic flux (Figure 13).

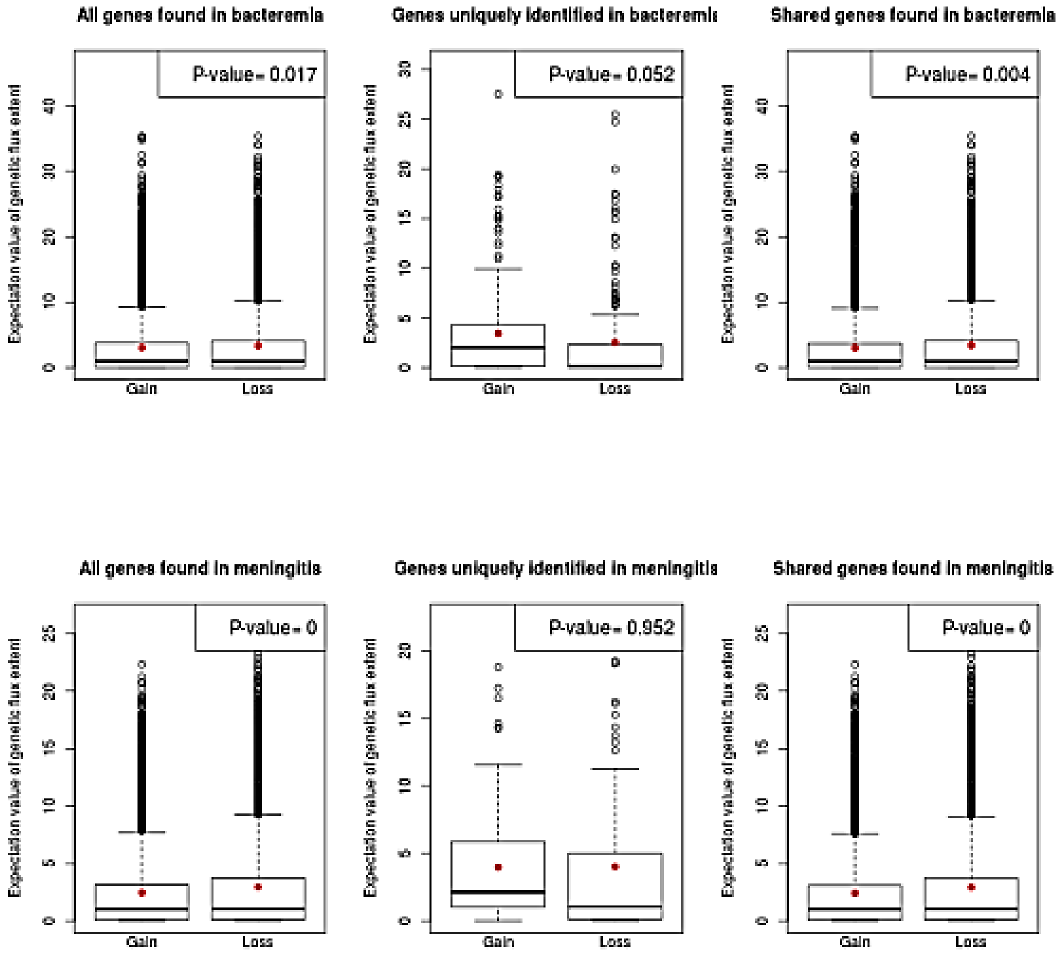


Figure 13: Expected value of genetic flux events box plot.

The p-value, which tests for null hypothesis value, uses a standard cut-off of 5% (α=0.05) below which the null hypothesis is rejected and vice versa (Chavalarias et al., 2016).

There was genetic flux among both of the strain sets showing evidence of genetic influence to their invasiveness to different body compartments. It shows that all of these genomes have certain genes, which should be conserved for the ability to infect the brain, and chance for the invasion might lower if a strain loses the conserved genes. Perhaps the ability to be able to invade the CSF in the meningitis subset is because they have a more conserved set of genes, compared to bacteremia dataset (Benard & Kulohoma, 2012).

CHAPTER 5

DISCUSSION

Genetic flux is the probability to gain or lose genes throughout the phylogenetic evolution (Benard Kulohoma, 2012). It explains the process of obtaining new genetic components which among them can include the propensity to cause, or not to cause meningitis. Over the period of development of a strain, there occurs events which lead to access of new genes, through modification of the old genes or acquisition from the neighboring environment. This process changes the genetic makeup of the strain as well as its traits, expressed by the new genes. During carriage the genetic composition of an organism is altered through recombination events, which determines the niche of adaptation. For the pneumococcus to cross the blood-brain-barrier and cause meningitis, it has to possess certain genes that lack in strains that remain in the blood causing bacteremia. The understanding of the overall direction of pathogen evolution paves way for specific gene target in designing diagnostics, vaccines and drugs.

To determine the probability of gene gain and gene loss, a GLOOME analysis was done. The GLOOME results gave a clear visual representation of the level of gene gain and gene loss probability with a bar graph at the bottom of both meningitis and bacteremia data sets showing the expectation of gain or loss along the genomes. This implies that the probability of gene gain and loss varied at different points along the genomes. The cumulative expectation of gain and loss events was computed to quantify the genetic change over the phylogenetic evolution period. Generally, there was more gene loss events as compared to gene gain events. However, in the bacteremia sub-set, the accessory genes displayed more gene gain events than the gene loss events. This suggests that over time *Streptococcus pneumoniae* strains gain or lose genetic material to adapt to new environments or to develop resistance against drugs. This implies that perhaps the strains that were able to cause meningitis had a different genetic makeup from those that caused bacteremia. It also suggests that to cross the blood-brain-barrier, the strains needed specific traits from the respective genes.

The phyletic pattern of some of the challenging *Streptococcus pneumoniae* virulent factors was analyzed to show the specific proteins and their gene gain and loss events. The transport accessory protein showed more gene loss in both meningitis and bacteremia sets while the competence stimulating peptide showed more gene loss in the meningitis sub-set only. Also, the pneumococcal surface protein A displayed more gene gain in both meningitis and bacteremia sets. This implies that these specific proteins could be relevant during the development of diagnostics, drugs and vaccines against the *Streptococcus pneumoniae* bacteria.

During evolution, different strains of the same species loss and gain genes, leading to conservation or loss of traits from the parent genome, as well as acquisition of new traits. When genes are preserved from the parent genome, they are referred as core genes and they are shared by all the strains that has a common decent. The genes which are obtained from recombination over the evolutionary period tend to be unique to certain strains, and are called the accessory genes. In this study, the probability of gene gain and gene loss was established for all genes, core genes and accessory genes of each meningitis and bacteremia sets. It was observed that meningitis causing strains were more conserved compared to the bacteremia causing strains. This implies that the meningitis causing strains were enabled to cross the blood-brain-barrier to infect the brain by their genetic structure. It hence validates the argument that the genetic flux pattern of the *Streptococcus pneumonia* strains that had the propensity to cause meningitis is different from those that cause bacteremia.

CHAPTER 6

CONCLUSION AND RECOMMENDATIONS

This study suggests that accessory genes are essential for invasion of different body compartments. The probability of gene loss in all genes and core genes in both meningitis and bacteremia sets is more compared to gene gain. However, in the accessory genes for both meningitis and bacteremia there is more gene gain with more genetic flux in the bacteremia set. The observation from this study confirms that genetic flux affects the propensity to cause meningitis as compared to the propensity to cause bacteremia. This study highlights some of the known virulent factors affected by genetic flux. Future studies should evaluate the benefit of combining antigens associated with specific disease outcomes to improve diagnosis, increase coverage of vaccines and provide specific antibiotic targets.

REFERENCES

Abry, M. F., Kimenyi, K. M., Osowo, F. O., Odhiambo, W. O., Sewe, S. O., & Kulohoma, B. W. (2015). Genetic diversity of the Pneumococcal CbpA: Implications for next-generation vaccine development. *Human Vaccines and Immunotherapeutics*. https://doi.org/10.1080/21645515.2015.1021521

Adegbola, R. A., DeAntonio, R., Hill, P. C., Roca, A., Usuf, E., Hoet, B., & Greenwood, B. M. (2014). Carriage of Streptococcus pneumoniae and other respiratory bacterial pathogens in low and lower-middle income countries: A systematic review and meta-analysis. *PLoS ONE*, *9*(8). https://doi.org/10.1371/journal.pone.0103293

Alderson, M. R. (2016). Status of research and development of pediatric vaccines for Streptococcus pneumoniae. *Vaccine*, *34*(26), 2959–2961. https://doi.org/10.1016/j.vaccine.2016.03.107

Andam, C. P., & Hanage, W. P. (2015a). Mechanisms of genome evolution of Streptococcus. *Infection, Genetics and Evolution*, *33*, 334–342. https://doi.org/10.1016/j.meegid.2014.11.007

Andam, C. P., & Hanage, W. P. (2015b). Mechanisms of genome evolution of Streptococcus. *Infection, Genetics and Evolution*, *33*, 334–342. https://doi.org/10.1016/j.meegid.2014.11.007

Andre, G. O., Converso, T. R., Politano, W. R., Ferraz, L. F. C., Ribeiro, M. L., Leite, L. C. C., & Darrieux, M. (2017). Role of Streptococcus Pneumoniae proteins in evasion of complement-mediated immunity. *Frontiers in Microbiology*, *8*(FEB), 1–20. https://doi.org/10.3389/fmicb.2017.00224

Auzat, I., Chapuy-regaud, S., Santos, D. Dos, Ogunniyi, A. D., Thomas, I. Le, Garel, J., James, C., Trombe, M., & Yvette, G. (1999). *The NADH oxidase of Streptococcus pneumoniae : its involvement in competence and virulence*. *34*, 1018–1028.

Baril, L., Dietemann, J., & Béniguel, L. (2006). *Pneumococcal surface protein A (PspA) is effective at eliciting T cell- mediated responses during invasive pneumococcal disease in adults*. 277–286. https://doi.org/10.1111/j.1365-2249.2006.03148.x

Basavanna, S., Khandavilli, S., Yuste, J., Cohen, J. M., Hosie, A. H. F., Webb, A. J., Thomas, G. H., & Brown, J. S. (2009). *Screening of Streptococcus pneumoniae ABC Transporter Mutants Demonstrates that LivJHMGF , a Branched-Chain Amino Acid ABC Transporter , Is Necessary for Disease Pathogenesis* □. *77*(8), 3412–3423. https://doi.org/10.1128/IAI.01543-08

Bayramoglu, B., Toubiana, D., Vliet, S. Van, Inglis, R. F., & Shnerb, N. (2017). Bet-hedging in bacteriocin producing Escherichia coli populations : the single cell perspective. *Nature Publishing Group*, *January*, 1–10. https://doi.org/10.1038/srep42068

Benard, W., & Kulohoma, B. (2012). *Genetic antigen diversity and gene flux among meningitic and bacteraemia-associated pneumococci from Malawi*.

Bergmann, S., Rohde, M., Gursharan, S., & Hammerschmidt, S. (2001). *a -Enolase of Streptococcus pneumoniae is a plasmin (ogen) -binding protein displayed on the bacterial cell surface*. *40*.

Berical, A. C., Harris, D., Dela Cruz, C. S., & Possick, J. D. (2016). Pneumococcal vaccination strategies: An update and perspective. *Annals of the American Thoracic Society*, *13*(6), 933–944. https://doi.org/10.1513/AnnalsATS.201511-778FR

Berry, A. M., Lock, R. A., Hansman, D., & Paton, J. C. (1989). *Contribution of Autolysin to Virulence of Streptococcus pneumoniae*. *57*(8), 2324–2330.

Bistrović, A., Krstulović, L., Stolić, I., Drenjančević, D., Talapko, J., Taylor, M. C., Kelly, J. M., Bajić, M., & Raić-Malić, S. (2018). Synthesis, anti-bacterial and anti-protozoal activities of amidinobenzimidazole derivatives and their interactions with DNA and RNA. *Journal of Enzyme Inhibition and Medicinal Chemistry*, *33*(1), 1323–1334.

https://doi.org/10.1080/14756366.2018.1484733

Bogaert, D., De Groot, R., & Hermans, P. W. M. (2004). Streptococcus pneumoniae colonisation: The key to pneumococcal disease. In *Lancet Infectious Diseases* (Vol. 4, Issue 3, pp. 144–154). https://doi.org/10.1016/S1473-3099(04)00938-7

Brivanlou, A. H., & Darnell, J. E. (2002). Transcription: Signal transduction and the control of gene expression. *Science*, *295*(5556), 813–818. https://doi.org/10.1126/science.1066355

Brooks, L. R. K., & Mias, G. I. (2018). Streptococcus pneumoniae's virulence and host immunity: Aging, diagnostics, and prevention. In *Frontiers in Immunology*. https://doi.org/10.3389/fimmu.2018.01366

Brown, J. S., Gilliland, S. M., Ruiz-albert, J., & Holden, D. W. (2002). *Characterization of Pit , a Streptococcus pneumoniae Iron Uptake ABC Transporter*. *70*(8), 4389–4398. https://doi.org/10.1128/IAI.70.8.4389

Brueggemann, A. B., Muroki, B. M., Kulohoma, B. W., Karani, A., Wanjiru, E., Morpeth, S., Kamau, T., Sharif, S., & Scott, J. A. G. (2013). Population genetic structure of Streptococcus pneumoniae in Kilifi, Kenya, prior to the introduction of pneumococcal conjugate vaccine. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0081539

Bryant, J. C., Dabbs, R. C., Oswalt, K. L., Brown, L. R., Rosch, J. W., Seo, K. S., Donaldson, J. R., McDaniel, L. S., & Thornton, J. A. (2016). Pyruvate oxidase of Streptococcus pneumoniae contributes to pneumolysin release. *BMC Microbiology*, *16*(1), 1–12. https://doi.org/10.1186/s12866-016-0881-6

Burton, D. R. (2017). What are the most powerful immunogen design vaccine strategies?: Reverse vaccinology 2.0 shows great promise. *Cold Spring Harbor Perspectives in Biology*, *9*(11). https://doi.org/10.1101/cshperspect.a030262

Cahoon, L. A., & Freitag, N. E. (2015). Identification of conserved and species-specific functions of the Listeria monocytogenes PrsA2 secretion chaperone. *Infection and Immunity*, *83*(10), 4028–4041. https://doi.org/10.1128/IAI.00504-15

CDC. (2019). *Catch-Up Guidance for Healthy 1 Children 4 Months through 4 Years of Age Pneumococcal Conjugate Vaccine: PCV*. *February*, 1–3. www.cdc.gov/vaccines/schedules/downloads/child/0-18yrs-child-combined-schedule.pdf.

Cecchini, P., Goldblatt, D., Brown, J. S., Whiting, G., McIlgorm, A., Lam, O., Entwisle, C., Ercoli, G., Ramos-Sevillano, E., Chan, W.-Y., Wheeler, J. X., Bailey, C., & Green, N. (2018). A novel, multiple-antigen pneumococcal vaccine protects against lethal Streptococcus pneumoniae challenge . *Infection and Immunity*, *December*. https://doi.org/10.1128/iai.00846-18

Chan, W. T., & Espinosa, M. (2016). The Streptococcus pneumoniae pezAT Toxin–Antitoxin System Reduces β-Lactam Resistance and Genetic Competence. *Frontiers in Microbiology*, *7*(August), 1322. https://doi.org/10.3389/fmicb.2016.01322

Chavalarias, D., Wallach, J. D., Ho, A., Li, T., & Ioannidis, J. P. A. (2016). *Evolution of Reporting*. *94305*(11), 1141–1148. https://doi.org/10.1001/jama.2016.1952

Cheng, W., Li, Q., Jiang, Y., Zhou, C., & Chen, Y. (2013). *Structures of Streptococcus pneumoniae PiaA and Its Complex with Ferrichrome Reveal Insights into the Substrate Binding and Release of High Affinity Iron Transporters*. *8*(8). https://doi.org/10.1371/journal.pone.0071451

Cockeran, R., Anderson, R., & Feldman, C. (2005). *Pneumolysin as a vaccine and drug target in the prevention and treatment of invasive pneumococcal disease*. *June 2014*.

Cornick, J. E., Harris, S. R., Parry, C. M., Moore, M. J., Jassi, C., Kamng’ona, A., Kulohoma, B., Heyderman, R. S., Bentley, S. D., & Everett, D. B. (2014). Genomic identification of a novel co-trimoxazole resistance genotype and its prevalence amongst Streptococcus pneumoniae in Malawi. *Journal of Antimicrobial Chemotherapy*. https://doi.org/10.1093/jac/dkt384

Daniels, C. C., Rogers, P. D., & Shelton, C. M. (2016). A review of pneumococcal vaccines: Current polysaccharide vaccine recommendations and future protein antigens. *Journal of Pediatric Pharmacology and Therapeutics*, *21*(1), 27–35. https://doi.org/10.5863/1551-6776-21.1.27

Delany, I., Rappuoli, R., & Seib, K. L. (2013). Vaccines, reverse vaccinology, and bacterial pathogenesis. *Cold Spring Harbor Perspectives in Medicine*, *3*(5). https://doi.org/10.1101/cshperspect.a012476

Echenique, J., Kadioglu, A., Romao, S., & Andrew, P. W. (2004). *Protein Serine / Threonine Kinase StkP Positively Controls Virulence and Competence in Streptococcus pneumoniae*. *72*(4), 2434–2437. https://doi.org/10.1128/IAI.72.4.2434

El Khoury, J. Y., Boucher, N., Bergeron, M. G., Leprohon, P., & Ouellette, M. (2017). Penicillin induces alterations in glutamine metabolism in Streptococcus pneumoniae. *Scientific Reports*, *7*(1), 1–15. https://doi.org/10.1038/s41598-017-15035-y

Feldman, C., & Anderson, R. (2014). Recent advances in our understanding of Streptococcus pneumoniae infection. *F1000Prime Reports*, *6*(September). https://doi.org/10.12703/P6-82

Feldman, C., & Anderson, R. (2016). Epidemiology, virulence factors and management of the pneumococcus. *F1000Research*, *5*(0), 2320. https://doi.org/10.12688/f1000research.9283.1

Garmory, H. S., & Titball, R. W. (2004). *MINIREVIEW ATP-Binding Cassette Transporters Are Targets for the Development of Antibacterial Vaccines and Therapies*. *72*(12), 6757–6763. https://doi.org/10.1128/IAI.72.12.6757

Gerdes, K. (2013). Prokaryotic toxin-antitoxins. *Prokaryotic Toxin-Antitoxins*, *9783642332*, 1–365. https://doi.org/10.1007/978-3-642-33253-1

Gianfaldoni, C., Maccari, S., Pancotto, L., Rossi, G., Hilleringmann, M., Pansegrau, W., Sinisi, A., Moschioni, M., Masignani, V., Rappuoli, R., Giudice, G. Del, & Ruggiero, P. (2009). *Sortase A Confers Protection against Streptococcus pneumoniae in Mice* □. *77*(7), 2957–2961. https://doi.org/10.1128/IAI.01516-08

Giersing, B. K., Modjarrad, K., Kaslow, D. C., Moorthy, V. S., Bavdekar, A., Cichutek, K., Cravioto, A., Fritzell, B., Graham, B. S., Karron, R., Lanata, C. F., Powell, M., Shao, Y., & Smith, P. (2016). Report from the World Health Organization's Product Development for Vaccines Advisory Committee (PDVAC) meeting, Geneva, 7-9th Sep 2015. *Vaccine*, *34*(26), 2865–2869. https://doi.org/10.1016/j.vaccine.2016.02.078

Gómez-mejia, A., Gámez, G., & Hammerschmidt, S. (2018). International Journal of Medical Microbiology Streptococcus pneumoniae two-component regulatory systems : The interplay of the pneumococcus with its environment. *International Journal of Medical Microbiology*, *308*(6), 722–737. https://doi.org/10.1016/j.ijmm.2017.11.012

Hajaj, B., Yesilkaya, H., Benisty, R., David, M., Andrew, P. W., & Porat, N. (2012). Thiol peroxidase is an important component of streptococcus pneumoniae in oxygenated environments. *Infection and Immunity*, *80*(12), 4333–4343. https://doi.org/10.1128/IAI.00126-12

Hava, D. L., Hemsley, C. J., & Camilli, A. (2003). Transcriptional regulation in the Streptococcus pneumoniae rlrA pathogenicity islet by RlrA. *Journal of Bacteriology*, *185*(2), 413–421. https://doi.org/10.1128/JB.185.2.413-421.2003

Hendriksen, W. T., Silva, N., Bootsma, H. J., Blue, C. E., Paterson, G. K., Kerr, A. R., De Jong, A., Kuipers, O. P., Hermans, P. W. M., & Mitchell, T. J. (2007). Regulation of gene expression in Streptococcus pneumoniae by response regulator 09 is strain dependent. *Journal of Bacteriology*, *189*(4), 1382–1389. https://doi.org/10.1128/JB.01144-06

Henriques-Normark, B., & Normark, S. (2014). Bacterial vaccines and antibiotic resistance. *Upsala Journal of Medical Sciences*, *119*(2), 205–208. https://doi.org/10.3109/03009734.2014.903324

Henriques-Normark, B., & Tuomanen, E. I. (2013). The pneumococcus: Epidemiology, microbiology, and pathogenesis. *Cold Spring Harbor Perspectives in Medicine*, *3*(7), 1–15. https://doi.org/10.1101/cshperspect.a010215

Herta, T., Bhattacharyya, A., Bollensdorf, C., Kabus, C., García, P., Suttorp, N., Hippenstiel, S., & Zahlten, J. (2018). *DNA-release by Streptococcus pneumoniae autolysin LytA induced Krueppel-like factor 4 expression in macrophages*. *March*, 1–14. https://doi.org/10.1038/s41598-018-24152-1

Hien, T. D.-, Young, H.-, & Hye, E.-. (2011). *Heat- Shock Protein ClpL / HSP100 Increases Penicillin Tolerance in Streptococcus pneumoniae*. *72*, 126–128.

Hirst, R. A., Kadioglu, A., Callaghan, C. O., & Andrew, P. W. (2004). *The role of pneumolysin in pneumococcal pneumonia and meningitis*. https://doi.org/10.1111/j.1365-2249.2004.02611.x

Hollingshead, S. K., Baril, L., Ferro, S., King, J., Coan, P., Briles, D. E., Epi, P., Group, S., & Baril, L. (2006). *Pneumococcal surface protein A (PspA) family distribution among clinical isolates from adults over 50 years of age collected in seven countries Printed in Great Britain*. 215–221. https://doi.org/10.1099/jmm.0.46268-0

Hupp, S., Grandgirard, D., Mitchell, T. J., Leib, S. L., Hathaway, L. J., & Iliev, A. I. (2019). *Pneumolysin and the bacterial capsule of Streptococcus pneumoniae cooperatively inhibit taxis and motility of microglia*. *7*, 1–14.

Iannelli, F., Oggioni, M. R., & Pozzi, G. (2005). *Sensor domain of histidine kinase ComD confers competence pherotype specificity in Streptoccoccus pneumoniae*. *252*, 321–326. https://doi.org/10.1016/j.femsle.2005.09.008

Iovino, F., Orihuela, C. J., Moorlag, H. E., Molema, G., & Bijlsma, J. J. E. (2013). Interactions between Blood-Borne Streptococcus pneumoniae and the Blood-Brain Barrier Preceding Meningitis. *PLoS ONE*, *8*(7). https://doi.org/10.1371/journal.pone.0068408

Iovino, F., Seinen, J., Henriques-Normark, B., & van Dijl, J. M. (2016). How Does Streptococcus pneumoniae Invade the Brain? *Trends in Microbiology*, *24*(4), 307–315. https://doi.org/10.1016/j.tim.2015.12.012

Ip, M., Lyon, D. J., Yung, R. W. H., Chan, C., & Cheng, A. F. B. (2001). Macrolide resistance in Streptococcus pneumoniae in Hong Kong. *Antimicrobial Agents and Chemotherapy*, *45*(5), 1578–1580. https://doi.org/10.1128/AAC.45.5.1578-1580.2001

Jedrzejas, M. J. (2003). Pneumococcal Virulence Factors: Structure and Function. *Microbiology and Molecular Biology Reviews*, *65*(2), 187–207. https://doi.org/10.1128/mmbr.65.2.187-207.2001

Jedrzejas, Mark J. (2001). *Pneumococcal Virulence Factors : Structure and Function*. *65*(2), 187–207. https://doi.org/10.1128/MMBR.65.2.187

Jomaa, M., Yuste, J., Paton, J. C., Jones, C., Dougan, G., & Brown, J. S. (2005). *Antibodies to the Iron Uptake ABC Transporter Lipoproteins PiaA and PiuA Promote Opsonophagocytosis of Streptococcus pneumoniae*. *73*(10), 6852–6859. https://doi.org/10.1128/IAI.73.10.6852

Kadioglu, A., Weiser, J. N., Paton, J. C., & Andrew, P. W. (2008). The role of Streptococcus pneumoniae virulence factors in host respiratory colonization and disease. In *Nature Reviews Microbiology*. https://doi.org/10.1038/nrmicro1871

Kim, L., McGee, L., Tomczyk, S., & Beall, B. (2016). Biological and epidemiological features of antibiotic-resistant Streptococcus pneumoniae in pre- and post-conjugate vaccine eras: A United States perspective. In *Clinical Microbiology Reviews*.

https://doi.org/10.1128/CMR.00058-15

Koedel, U., Scheld, W. M., & Pfister, H. W. (2002). Pathogenesis and pathophysiology of pneumococcal meningitis. In *Lancet Infectious Diseases*. https://doi.org/10.1016/S1473-3099(02)00450-4

Kolberg, J., Aase, A., Bergmann, S., Herstad, T. K., Rødal, G., Frank, R., Rohde, M., & Hammerschmidt, S. (2006). *Streptococcus pneumoniae enolase is important for plasminogen binding despite low abundance of enolase protein on the bacterial cell surface Streptococcus pneumoniae enolase is important for plasminogen binding despite low abundance of enolase protein on . June*. https://doi.org/10.1099/mic.0.28747-0

Kulohoma, B. W. (2015). BMX: A tool for computing bacterial phyletic composition from orthologous maps. *BMC Research Notes*. https://doi.org/10.1186/s13104-015-1017-z

Kulohoma, B. W., Marriage, F., Vasieva, O., Mankhambo, L., Nguyen, K., Molyneux, M. E., Molyneux, E. M., Day, P. J. R., & Carrol, E. D. (2017). Peripheral blood RNA gene expression in children with pneumococcal meningitis: a prospective case–control study. *BMJ Paediatrics Open*. https://doi.org/10.1136/bmjpo-2017-000092

Kumari, P., Nath, Y., Murty, U. S., Bonaventura, G. Di, & Chifiriuc, M. C. (2020). *Sortase A Mediated Bioconjugation of Common Epitopes Decreases Biofilm Formation in Staphylococcus aureus*. *11*(July), 1–8. https://doi.org/10.3389/fmicb.2020.01702

Leonor, M., Oliveira, S., Monedero, V., Miyaji, E. N., Leite, L. C. C., Lee, P., & Pe, G. (2003). *Expression of Streptococcus pneumoniae antigens , PsaA (pneumococcal surface antigen A) and PspA (pneumococcal surface protein A) by Lactobacillus casei*. *227*, 25–31. https://doi.org/10.1016/S0378-1097(03)00645-1

Li, Y., Metcalf, B. J., Chochua, S., Li, Z., Walker, H., Tran, T., Hawkins, P. A., Gierke, R., Pilishvili, T., McGee, L., & Beall, B. W. (2019). Genome-wide association analyses of invasive pneumococcal isolates identify a missense bacterial mutation associated with meningitis. *Nature Communications*, *10*(1), 178. https://doi.org/10.1038/s41467-018-07997-y

Lux, T., Nuhn, M., Hakenbeck, R., & Reichmann, P. (2007). *Diversity of Bacteriocins and Activity Spectrum in Streptococcus pneumoniae* □. *189*(21), 7741–7751. https://doi.org/10.1128/JB.00474-07

Magni, P., Bier, D. M., Pecorelli, S., Agostoni, C., Astrup, A., Brighenti, F., Cook, R., Folco, E., Fontana, L., Gibson, R. A., Guerra, R., Guyatt, G. H., Ioannidis, J. P. A., Jackson, A. S., Klurfeld, D. M., Makrides, M., Mathioudakis, B., Monaco, A., Patel, C. J., … Peracino, A. (2017). Perspective: Improving nutritional guidelines for sustainable health policies: Current status and perspectives. *Advances in Nutrition*, *8*(4), 532–545. https://doi.org/10.3945/an.116.014738

Manuscript, A. (2014). *NIH Public Access*. *38*(3), 473–492. https://doi.org/10.1111/1574-6976.12046.Peptide

Marks, L. R., Reddinger, R. M., & Hakansson, A. P. (2012). High levels of genetic recombination during nasopharyngeal carriage and biofilm formation in Streptococcus pneumoniae. *MBio*, *3*(5), 1–13. https://doi.org/10.1128/mBio.00200-12

Martner, A., Dahlgren, C., Paton, J. C., & Wold, A. E. (2008). *Pneumolysin Released during Streptococcus pneumoniae Autolysis Is a Potent Activator of Intracellular Oxygen Radical Production in Neutrophils* □. *76*(9), 4079–4087. https://doi.org/10.1128/IAI.01747-07

Masomian, M., Ahmad, Z., & Gew, L. T. (2020). *Development of Next Generation Streptococcus pneumoniae Vaccines Conferring Broad Protection*. 1–23.

Meichanetzidis, K., Turner, C. J., Farjami, A., Papić, Z., & Pachos, J. K. (2018). Free-fermion descriptions of parafermion chains and string-net models. *Physical Review B*, *97*(12), 417–425. https://doi.org/10.1103/PhysRevB.97.125104

Melin, M., Paolo, E. Di, Tikkanen, L., Jarva, H., Neyt, C., Ka, H., Meri, S., Poolman, J., & Va, M. (2010). *Interaction of Pneumococcal Histidine Triad Proteins with Human Complement □*. *78*(5), 2089–2098. https://doi.org/10.1128/IAI.00811-09

Mellroth, P., Daniels, R., Eberhardt, A., Rönnlund, D., Blom, H., Widengren, J., Normark, S., & Henriques-normark, B. (2012). *LytA , Major Autolysin of Streptococcus pneumoniae , Requires Access to Nascent Peptidoglycan * □*. *287*(14), 11018–11029. https://doi.org/10.1074/jbc.M111.318584

Merrifield, M., Hotez, P. J., Beaumier, C. M., Gillespie, P., Strych, U., Hayward, T., & Bottazzi, M. E. (2016). Advancing a vaccine to prevent human schistosomiasis. *Vaccine*, *34*(26), 2988–2991. https://doi.org/10.1016/j.vaccine.2016.03.079

Mitchell, A. M., & Mitchell, T. J. (2010). Streptococcus pneumoniae: Virulence factors and variation. In *Clinical Microbiology and Infection*. https://doi.org/10.1111/j.1469-0691.2010.03183.x

Morrison, D. A. (1996). *Regulation of competence for genetic transformation in Streptococcus pneumoniae by an auto-induced peptide pheromone and a two-component regulatory system*. *21*, 853–862.

Muchnik, L., Adawi, A., Ohayon, A., Dotan, S., Malka, I., Azriel, S., Shagan, M., Portnoi, M., Kafka, D., Nahmani, H., Porgador, A., Gershoni, J. M., Morrison, D. A., Mitchell, A., Tal, M., Ellis, R., Dagan, R., & Nebenzahl, Y. M. (2013). *NADH Oxidase Functions as an Adhesin in Streptococcus pneumoniae and Elicits a Protective Immune Response in Mice*. *8*(4). https://doi.org/10.1371/journal.pone.0061128

Mücke, P. A., Maaß, S., Kohler, T. P., Hammerschmidt, S., & Becher, D. (2020). Proteomic adaptation of streptococcus pneumoniae to the human antimicrobial peptide LL-37. *Microorganisms*, *8*(3). https://doi.org/10.3390/microorganisms8030413

Muley, V., Ghadage, D., Yadav, G., & Bhore, A. (2016). Study of invasive pneumococcal infection in adults with reference to penicillin resistance. *Journal of Laboratory Physicians*, *9*(1), 31. https://doi.org/10.4103/0974-2727.187918

Nelson, A. L., Roche, A. M., Gould, J. M., Chim, K., Ratner, A. J., Weiser, J. N., Al, N. E. T., & Mmun, I. N. I. (2007). Capsule Enhances Pneumococcal Colonization by Limiting. *Infection and Immunity*, *75*(1), 83–90. https://doi.org/10.1128/IAI.01475-06

O'Brien, K. L., Wolfson, L. J., Watt, J. P., Henkle, E., Deloria-Knoll, M., McCall, N., Lee, E., Mulholland, K., Levine, O. S., & Cherian, T. (2009). Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: global estimates. *The Lancet*, *374*(9693), 893–902. https://doi.org/10.1016/S0140-6736(09)61204-6

Obolski, U., Gori, A., Lourenço, J., Thompson, C., Thompson, R., French, N., Heyderman, R. S., & Gupta, S. (2019). Identifying genes associated with invasive disease in S. pneumoniae by applying a machine learning approach to whole genome sequence typing data. *Scientific Reports*, *9*(1), 4049. https://doi.org/10.1038/s41598-019-40346-7

Ochs, M. M., Williams, K., Sheung, A., Lheritier, P., Visan, L., Rouleau, N., Proust, E., de Montfort, A., Tang, M., Mari, K., Hopfer, R., Gallichan, S., & Brookes, R. H. (2016). A bivalent pneumococcal histidine triad protein D-choline-binding protein A vaccine elicits functional antibodies that passively protect mice from Streptococcus pneumoniae challenge. *Human Vaccines and Immunotherapeutics*, *12*(11), 2946–2952. https://doi.org/10.1080/21645515.2016.1202389

Ogunniyi, A. D., Grabowicz, M., Mahdi, L. K., Cook, J., Gordon, D. L., Sadlon, T. A., & Paton, J. C. (n.d.). *Pneumococcal histidine triad proteins are regulated by the Zn 2 □ - dependent repressor AdcR and inhibit complement deposition through the recruitment of complement factor H*. 1–8. https://doi.org/10.1096/fj.08-119537

Ogunniyi, A. D., Mahdi, L. K., Trappetti, C., Verhoeven, N., Mermans, D., Van der Hoek, M. B., Plumptre, C. D., & Paton, J. C. (2012). Identification of Genes That Contribute to the Pathogenesis of Invasive Pneumococcal Disease by In Vivo Transcriptomic Analysis . *Infection and Immunity*, *80*(9), 3268–3278. https://doi.org/10.1128/iai.00295-12

Orihuela, C. J., Radin, J. N., Sublett, J. E., Gao, G., Kaushal, D., & Tuomanen, E. I. (2004). Microarray analysis of pneumococcal gene expression during invasive disease. *Infection and Immunity*, *72*(10), 5582–5596. https://doi.org/10.1128/IAI.72.10.5582-5596.2004

Pan, F., Zhang, H., Dong, X., Ye, W., He, P., Zhang, S., Zhu, J. X., & Zhong, N. (2018). Comparative genomic analysis of multidrug-resistant Streptococcus pneumoniae isolates. *Infection and Drug Resistance*, *11*, 659–670. https://doi.org/10.2147/IDR.S147858

Paterson, G. K., & Mitchell, T. J. (2006). *The role of Streptococcus pneumoniae sortase A in colonisation and pathogenesis*. *8*, 145–153. https://doi.org/10.1016/j.micinf.2005.06.009

Pei, J., & Grishin, N. V. (2001). Type II CAAX prenyl endopeptidases belong to a novel superfamily of putative membrane-bound metalloproteases. *Trends in Biochemical Sciences*, *26*(5), 275–277. https://doi.org/10.1016/s0968-0004(01)01813-8

Pi, E., Cian, M. B., Olivero, N. B., & Perez, D. R. (2018). *Crosstalk between the serine / threonine kinase StkP and the response regulator ComE controls the stress response and intracellular survival of Streptococcus pneumoniae*.

Piccoli, L., Microbiologia, S., & Chirurgiche, S. (1996). *Competence for Genetic Transformation in Encapsulated Strains of Streptococcus pneumoniae : Two Allelic Variants of the Peptide Pheromone*. *178*(20), 6087–6090.

Pinchas, M. D., & Lacross, N. C. (2015). *An Electrostatic Interaction between BlpC and BlpH Dictates Pheromone Specificity in the Control of Bacteriocin Production and Immunity in Streptococcus pneumoniae*. *197*(7), 1236–1248. https://doi.org/10.1128/JB.02432-14

Plumptre, C. D., Ogunniyi, A. D., & Paton, J. C. (2013). Vaccination against Streptococcus pneumoniae using truncated derivatives of polyhistidine triad protein D. *PLoS ONE*, *8*(10). https://doi.org/10.1371/journal.pone.0078916

Popowicz, N. D., Lansley, S. M., Cheah, H. M., Kay, I. D., Carson, C. F., Waterer, G. W., Paton, J. C., Brown, J. S., & Lee, Y. C. G. (2017). Human pleural fluid is a potent growth medium for Streptococcus pneumoniae. *PLoS ONE*, *12*(11), 1–14. https://doi.org/10.1371/journal.pone.0188833

Prager, O., Friedman, A., & Nebenzahl, Y. M. (2017). Role of neural barriers in the pathogenesis and outcome of Streptococcus pneumoniae meningitis (Review). In *Experimental and Therapeutic Medicine*. https://doi.org/10.3892/etm.2017.4082

Price, K. E., & Camilli, A. (2009). *Pneumolysin Localizes to the Cell Wall of Streptococcus pneumoniae □*. *191*(7), 2163–2168. https://doi.org/10.1128/JB.01489-08

Pujanauski, L., Colino, J., Flora, M., Torres, R. M., Tuomanen, E., & Snapper, C. M. (2020). *Pneumococcal Surface Protein A Plays a Major Role in Streptococcus pneumoniae − Induced Immunosuppression*. https://doi.org/10.4049/jimmunol.1502709

Raddaoui, A., Simoes, A. S., Baaboura, R., Felix, S., Achour, W., Othman, T. Ben, Bejaoui, M., Sa-Leao, R., & Hassen, A. Ben. (2015). Serotype distribution, antibiotic resistance and clonality of streptococcus pneumoniae isolated from immunocompromised patients in Tunisia. *PLoS ONE*, *10*(10), 1–10. https://doi.org/10.1371/journal.pone.0140390

Saskova, L., Nova, L., Basler, M., & Branny, P. (2007). *Eukaryotic-Type Serine / Threonine Protein Kinase StkP Is a Global Regulator of Gene Expression in Streptococcus pneumoniae □ †*. *189*(11), 4168–4179. https://doi.org/10.1128/JB.01616-06

Shenoy, A. T., & Orihuela, C. J. (2016). Anatomical site-specific contributions of pneumococcal virulence determinants. *Pneumonia*, *8*(1), 7. https://doi.org/10.1186/s41479-016-0007-9

Spellerberg, B., Cundell, D. R., Sandros, J., Pearce, B. J., Ida, I., & Masure, H. R. (1996). *Pyruvate oxidase , as a determinant of virulence in Streptococcus pneumoniae*. *19*, 803–813.

Steinmoen, H., Teigen, A., & Ha, L. S. (2003). *Competence-Induced Cells of Streptococcus pneumoniae Lyse Competence-Deficient Cells of the Same Strain during Cocultivation*. *185*(24), 7176–7183. https://doi.org/10.1128/JB.185.24.7176

Stoppelaar, S. F. De, Bootsma, H. J., Zomer, A., Roelofs, J. J. T. H., Hermans, P. W. M., Veer, C. Van, & Poll, T. Van Der. (2013). *Streptococcus pneumoniae Serine Protease HtrA , but Not SFP or PrtA , Is a Major Virulence Factor in Pneumonia*. *8*(11). https://doi.org/10.1371/journal.pone.0080062

Tennant, S. M., MacLennan, C. A., Simon, R., Martin, L. B., & Khan, M. I. (2016). Nontyphoidal salmonella disease: Current status of vaccine research and development. *Vaccine*, *34*(26), 2907–2910. https://doi.org/10.1016/j.vaccine.2016.03.072

The FDA Food Safety Modernization Act. (2011).

Tomos, I. (n.d.). *Prevention of Invasive Pneumococcal Disease (IPD)*.

Tu, A. T., Fulgham, R. L., Crory, M. A. M. C., Briles, D. E., & Szalai, A. J. (1999). *Pneumococcal Surface Protein A Inhibits Complement Activation by Streptococcus pneumoniae*. *67*(9), 4720–4724.

van der Poll, T., & Opal, S. M. (2009a). Pathogenesis, treatment, and prevention of pneumococcal pneumonia. *The Lancet*, *374*(9700), 1543–1556. https://doi.org/10.1016/S0140-6736(09)61114-4

van der Poll, T., & Opal, S. M. (2009b). Pathogenesis, treatment, and prevention of pneumococcal pneumonia. *Lancet (London, England)*, *374*(9700), 1543–1556. https://doi.org/10.1016/S0140-6736(09)61114-4

Wahl, B., O'Brien, K. L., Greenbaum, A., Majumder, A., Liu, L., Chu, Y., Lukšić, I., Nair, H., McAllister, D. A., Campbell, H., Rudan, I., Black, R., & Knoll, M. D. (2018). Burden of Streptococcus pneumoniae and Haemophilus influenzae type b disease in children in the era of conjugate vaccines: global, regional, and national estimates for 2000–15. *The Lancet Global Health*, *6*(7), e744–e757. https://doi.org/10.1016/S2214-109X(18)30247-X

Weiser, J. N., Ferreira, D. M., & Paton, J. C. (2018). Streptococcus pneumoniae: transmission, colonization and invasion. In *Nature reviews. Microbiology* (Vol. 16, Issue 6). https://doi.org/10.1038/s41579-018-0001-8

Whalan, R. H., Funnell, S. G. P., Bowler, L. D., Hudson, M. J., Robinson, A., & Dowson, C. G. (2005). *PiuA and PiaA , iron uptake lipoproteins of Streptococcus pneumoniae , elicit serotype independent antibody responses following human pneumococcal septicaemia*. *43*, 73–80. https://doi.org/10.1016/j.femsim.2004.07.010

Whatmore, A. M., Barcus, V. A., & Dowson, C. G. (1999). *Genetic Diversity of the Streptococcal Competence (com) Gene Locus*. *181*(10), 3144–3154.

World Health Organization. (2019). Pneumococcal conjugate vaccines in infants and children under 5 years of age: WHO position paper – February 2019. *Weekly Epidemiological Record*, *94*(8), 85–104. http://www.who.int/immu-%0Ahttps://www.who.int/immunization/policy/position_papers/who_pp_pcv_2019_summary.pdf?ua=1

Xiao, K., Wang, X., & Yu, H. (2019). *Comparative studies of catalytic pathways for Streptococcus pneumoniae sialidases NanA , NanB and NanC*. *December 2018*, 1–13. https://doi.org/10.1038/s41598-018-38131-z

Xu, G., Kiefel, M. J., Wilson, J. C., Andrew, P. W., Oggioni, M. R., & Taylor, G. L. (2011). *Three Streptococcus pneumoniae Sialidases : Three Different Products*. 1718–1721.

Yamamoto, Y. (2002). PCR in diagnosis of infection: Detection of bacteria in cerebrospinal fluids. *Clinical and Diagnostic Laboratory Immunology*, *9*(3), 508–514. https://doi.org/10.1128/CDLI.9.3.508-514.2002

Yang, G., Sau, C., Lai, W., Cichon, J., & Li, W. (2015). *蚊子网状进化HHS Public Access*. *344*(6188), 1173–1178. https://doi.org/10.1126/science.1249098.Sleep

Yang, Y., & Tal-Gan, Y. (2019). Exploring the competence stimulating peptide (CSP) N-terminal requirements for effective ComD receptor activation in group1 Streptococcus pneumoniae. *Bioorganic Chemistry*, *89*(March), 102987. https://doi.org/10.1016/j.bioorg.2019.102987

Yu, J., Bryant, A. P., Marra, A., Lonetto, M. A., Ingraham, K. A., Chalker, A. F., Holmes, D. J., Holden, D., Rosenberg, M., & Mcdevitt, D. (2001). *Characterization of the Streptococcus pneumoniae NADH oxidase that is required for infection*. 431–438.

Zhu, L., & Lau, G. W. (2011a). Inhibition of competence development, horizontal gene transfer and virulence in streptococcus pneumoniae by a modified competence stimulating peptide. *PLoS Pathogens*, *7*(9). https://doi.org/10.1371/journal.ppat.1002241

Zhu, L., & Lau, G. W. (2011b). *Inhibition of Competence Development , Horizontal Gene Transfer and Virulence in Streptococcus pneumoniae by a Modified Competence Stimulating Peptide*. *7*(9). https://doi.org/10.1371/journal.ppat.1002241

Zivich, P. N., Grabenstein, J. D., Becker-Dreps, S. I., & Weber, D. J. (2018). Streptococcus pneumoniae outbreaks and implications for transmission and control: a systematic review. *Pneumonia*, *10*(1). https://doi.org/10.1186/s41479-018-0055-4

APPENDICES

Appendix 1: Prioritized antigens during *Streptococcus pneumoniae* research

Gene	Synonym	Putative Function
spiR1	blpS	Regulatory protein
spiR2	blpR	Response regulator
spiH	blpH	Histidine kinase
spiP	blpC	Peptide pheromone
spiD	blpB	Transport accessory protein
spiA	blpA	ABC transporter, ATPase
spiB	blpA	ABC transporter, transmembrane domain
spiC	blpA	ABC transporter, C39 protease domain
pncA	blpI	Bacteriocin
pncB	NA	Immunity protein
pncC	NA	Hypothetical protein
pncD	blpJ	Bacteriocin
pncE	blpU/thmA	Bacteriocin
pncE2	blpK	Bacteriocin
pncF	SP0534	Hypothetical protein
pncG	SP0535	Immunity protein
pncH	blpL	Hypothetical protein
pncI	blpM	Bacteriocin
pncJ	blpN	Bacteriocin
pncK	NA	Immunity protein
pncL	SP0542	Hypothetical protein

pncM	NA	Immunity protein
pncN	blpX	Immunity protein
pncO	blpY	CAAX protease
pncP	SP0547	CAAX protease
pncQ	blpZ	Immunity protein
pncR		Bacteriocin
pncS		Hypothetical protein
pncT		Bacteriocin
pncU		Bacteriocin
pncV	blpO	Bacteriocin
pncW	spr0470	Hypothetical protein, fusion
Blp		bacteriocin-like peptide
BlpC		Peptide pheromone
PspC	CbpA/spsA	Surface Protein C,Choline-binding protein A
CbpD		Choline-binding protein D
CbpG		Putative serine protease, Choline-binding protein G
CbpH		Choline-binding protein H
CbpI		Choline-binding protein I
comA		membrane-associated peptide permease
comC		Competence stimulating peptide (CSP)
comD		membrane-localized histidine kinase receptor
comE		cognate transcriptional regulator
comX		Competence stimulating peptide (CSP), alternative sigma factor
DnaJ	hsp40	heat shock protein (hsp)
Eno		Phosphopyruvate hydratase (Enolase)

Hyl	Hyaluronate lyase
LytA	Autolysin (N-Acetyl-Muramoyl-L-Alanine Amidase)
LytB	Peptidoglycan hydrolase, Endo-β-N-Acetylglucosamidase
LytC	Peptidoglycan hydrolase, Lysozyme (1,4-β-N-Acetylmuramidase)
BgaA	surface-associated exoglycosidase,β-galactosidase
HtrA	heat shock-induced serine protease
iga1	immunoglobulin A 1 protease precursor
Nox	NADH oxidase
NanA	Sialidase A
NanB	Neuraminidase B
NanC	Neuraminidase C
PcpA	Choline-Binding Protein A
PcsB	putative murein hydrolase, protein required for cell separation B
CbpE	choline-binding protein
RrgA	Pilus-1 Tip Protein (Adhesin)
RrgB	Pilus-1 Backbone Protein
RrgC	Pilus-1 Anchore Protein
PAVA	Adherence and Virulence protein A
PhtA	Pneumococcal Histidine Triad A
PhtB	Pneumococcal Histidine Triad B
PhtD	Pneumococcal Histidine Triad D
PhtE	Pneumococcal Histidine Triad E
PiaA	Iron-Compound ABC Transporter
PiuA	Iron-Compound ABC Transporter

PppA	Pneumococcal Protective Protein A, Non-Heme Iron-Containing Ferritine
Ply	Pneumolysin
PotD	Polyamine transport protein D
PpmA	Foldase Protein PrsA, Proteinase Maturation A
PsaA	Pneumococcal Surface Adhehesin A
PspA	Pneumococcal Surface protein A
SpxB	Pyruvate oxidase
SOD	Superoxide dismutase
SP0189	Hypothetical protein
SP0376	DNA-binding response regulator
mntE	manganese efflux pump
SP1633	Response regulator
SP1651	Thiol peroxidase
StkP	Serine/Threonine Protein Kinase
SrtA	Sortase enzyme A
SrtH	Sortase enzyme H
Usp45	PcsB, Secreted 45-KDa Protein

Appendix 2: Scripts used

To convert genbank file format files to embl file format

- This can be done one by one as follows:

```
perl perlfile1.pl <NAME of GENOME file eg sequence.gb>  <the output.name for the results eg sequence.embl>
```

- or All 209 genomes as follows:

```
for f in *.gbff *.gb ; do perl perlfile1.pl $f ${f%}.embl ; done
```

Genbank to EMBL (perlfile1.pl)

```
#!/usr/local/bin/perl -w
use strict;
use Bio::SeqIO;

if (@ARGV != 2) {  die "USAGE: gb2embl.pl <genbank file> <output embl file> \n"; }

my $seqio = Bio::SeqIO->new ('-format' => 'genbank', '-file' => "$ARGV[0]");
my $seqout = new Bio::SeqIO ('-format' => 'embl', '-file' => ">$ARGV[1]");
while ( my $seq = $seqio->next_seq) {
 $seqout->write_seq ($seq)
}
```

To retrieve multi FASTA files from each of the embl files generated.

- To run the analysis for 1 file after the other:

```
perl perlfile2.pl <EMBL file name eg using the result from the example above sequence.embl >
```

- To run the analysis for all genomes in one as follows:

```
for f in *.embl ; do perl perlfile2.pl $f ; done
```

EMBL to multi FASTA (perlfile2.pl)

```
#!/usr/local/bin/perl -w

use strict;

use Bio::SeqIO;

use Getopt::Std;

use File::Basename;

use Data::Dumper;


#Author: np1@sanger.ac.uk


sub usage {


die <<EOF;
Usage:
$0: [-p <pvalue cutoff>] [-i <percentage ID cutoff>] [-m <mcl inflation>] [-s produce sybil format output] <embl files>

 Defaults: -p  1e-5
       -i  off
       -m  1.5
                -s  off

EOF

}

my %options;
getopts ( 'si:p:m:o:', \%options );

my $pvalue = "";
my $id = "";
my $inflation = "";
#store defaults for output
my $p = 0;
my $i = 0;
my $m = 1.5;
my $sybil = 0;
my $mcl;
my $pid;
my %count_by_organism;
if (defined $options{p}){
  $pvalue = "--pv_cutoff=$options{p}";
  $p = $options{p};
}
if (defined $options{i}){
  $id = "--pi_cutoff=$options{i}";
  $i = $options{i};
}
if (defined $options{m}){
  $inflation = "--inflation=$options{m}";
  $m = $options{m};
}

defined $options{o} and $mcl = $options{o};

exists $options{s} and $sybil = 1;

unless (scalar (@ARGV) > 1){
  warn "Need two or more files for comparison";
  usage;
}

foreach (@ARGV){
  unless (-e $_) {
          warn "You must specify the embl file locations: $_ does not exist\n";
          usage;
  }
}

my @tmp_files;
my %products;


print STDERR "Producing protein files...\n";

#write files to protein fasta
          foreach my $file (@ARGV){

           open (FASTATMP, ">$file.tmp") or die "Could not write to $file.tmp: $!";
           push (@tmp_files, "$file.tmp");
           my $stream = Bio::SeqIO -> new ( -file => $file, -format => 'embl');

           while (my $seq = $stream->next_seq ()){

                    #include psudo genes but add pseudo to name
                    foreach my $feature ($seq-> get_SeqFeatures ()){
```

```
            #print STDERR $feature->location->to_FTstring . "\n";
            if ( $feature -> primary_tag () eq 'CDS'){

                my $id;

                if ($feature ->has_tag ('systematic_id')){

                  ($id) = $feature -> get_tag_values ('systematic_id');
                }
                elsif ($feature ->has_tag ('temporary_systematic_id')){

                  ($id)        =        $feature        ->        get_tag_values
('temporary_systematic_id');
                }
                elsif ($feature ->has_tag ('locus_tag')){

                  ($id) = $feature -> get_tag_values ('locus_tag');
                }
                else {
                  warn"Couldn't find an id for the CDS\n";
                  next;
                }

                if ($feature ->has_tag ('pseudo')){
```

```
  $id = 'pseudo'. $id;
}


if (length ($feature-> spliced_seq -> translate -> seq) > 10){
          if ($feature -> has_tag ('product')){
            my @products = $feature->get_tag_values ('product');
            my $products = join (',', @products);
            $products{$id} = $products;
          }

          my $seq = $feature -> spliced_seq -> translate -> seq;

          $seq =~ s/\*//g;

          print FASTATMP ">$id\n$seq\n";
          $count_by_organism{$file}++;
}
else {

          print STDERR "Excluding $id.... too short\n";
}
```

```
                              }


                      }

              }


   close FASTATMP;


}
```

To join the bac and men files separately.

Making a folder with only bac *.tmp files and men *.tmp files.

Creating a joined file in respective folders, eg

```
cat *.tmp > 1.all_bac_sequences


cat *.tmp > 2.all_men_sequences
```

BLAST

1. Installing blast

2. Runing blast as follows:

```
bash run_blastp.sh
```

Run_blastp.sh

```
#!/bin/env bash

#SBATCH -p batch

#SBATCH -J map.step.1
```

```
#SBATCH -n 4
module load blast/2.7.1+


    makeblastdb -in 1.all_bac_sequences -input_type fasta -dbtype prot
    makeblastdb -in 2.all_men_sequences -input_type fasta -dbtype prot
    for f in 1.all_bac_sequences 2.all_men_sequences ; do blastp -query $f -db $f -outfmt 6 -evalue 1e-10 -num_threads 4 > ${f%}_blastp_result.txt ; done
```

Make all.gg_file for all the *.tmp files and then merge them

This is a file with just the fasta header line minus the ‘>’ symbol.

The command for each file is

```
cat <file name> | grep ‘>’ | sed ’s/>//g’ > <output file name>
```

Annotation

Annotate.sh file

```
cat _bac_CP001993.embl | grep -1 "/locus_tag=" | grep -v "gene" | grep -v "CDS" | grep -v "\-\-" | grep -v "\/transl" | grep -v "tRNA" | grep -v "rRNA" | grep -v "ncRNA" | grep -v "tmRNA" | uniq | sed 's/FT          \///g' | tr '\n' ' ' | sed 's/note=//g' > 1.bac_CP001993.embl.annotations.txt

cat _bac_tigr4_AE005672.embl | grep -1 "/locus_tag=" | grep -v "gene" | grep -v "CDS" | grep -v "\-\-" | grep -v "\/transl" | grep -v "tRNA" | grep -v "rRNA" | grep -v "ncRNA" | grep -v
```

```
"tmRNA" | uniq | sed 's/FT                    \//g' | tr '\n' ' ' | sed 's/note=//g' > 1.bac_tigr4_AE005672.embl.annotations.txt
```

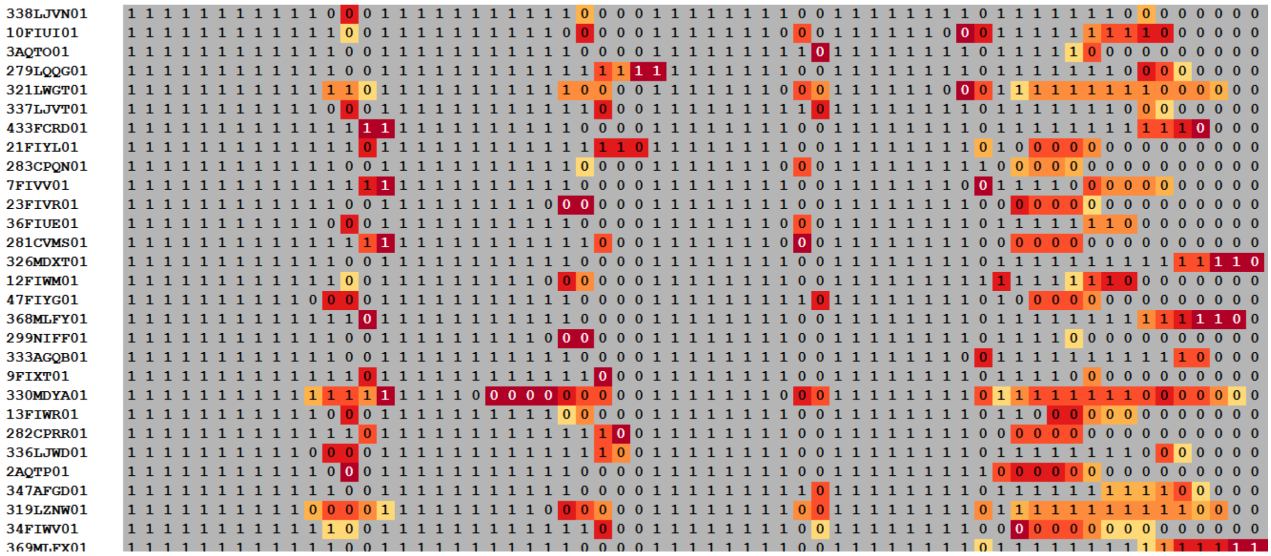
```
cat _men_331.MDYB01.1.embl | grep 'locus_tag\|/product=' | uniq | sed 's/FT \/product\=//g' | tr '\n' ' ' > 1.men_331.MDYB01.1.embl.annotations.txt
```

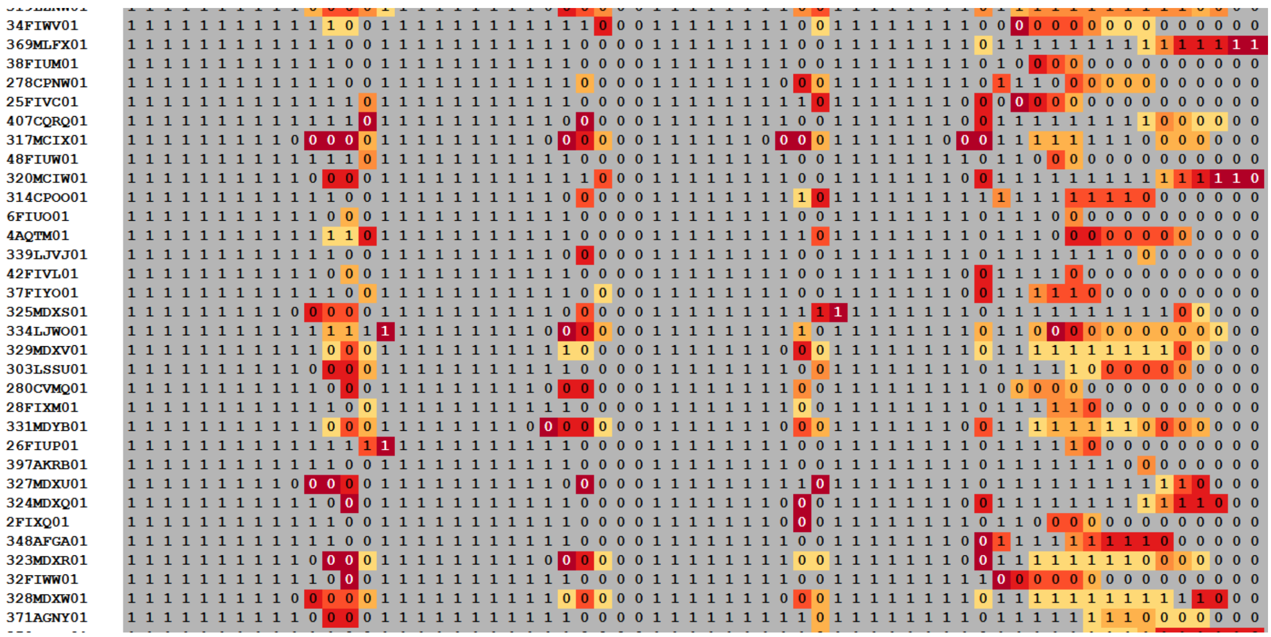
```
cat _men_327.MDXU01.1.embl | grep 'locus_tag\|/product=' | uniq | sed 's/FT \/product\=//g' | tr '\n' ' ' > 1.men_327.MDXU01.1.embl.annotations.txt
```

Appendix 3: Gain Loss Mapping Engine (GLOOME)

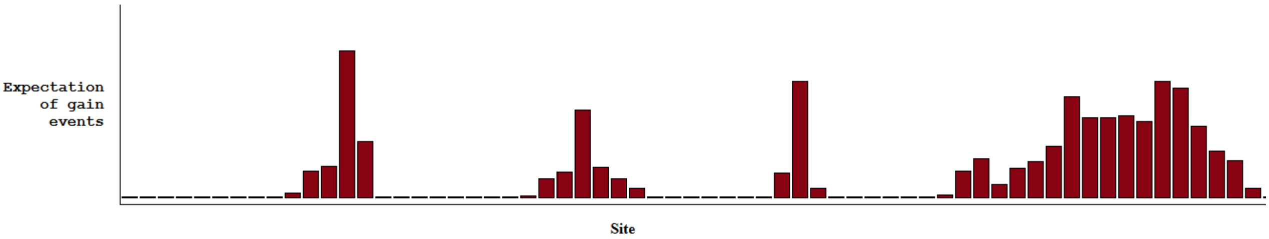
Meningitis gain and loss

Phyletic pattern color-coded by gain probability



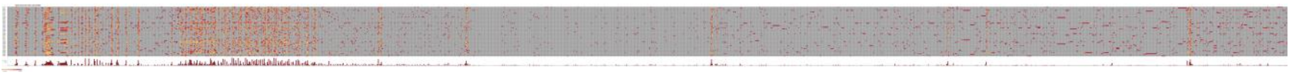


```
397AKRB01  1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
327MDXU01  1 1 1 1 1 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0
324MDXQ01  1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0
2FIXQ01    1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0
348AFGA01  1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0
323MDXR01  1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0
32FIWW01   1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
328MDXW01  1 1 1 1 1 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0
371AGNY01  1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0
350AGNZ01  1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0
29FIWS01   1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0
44FIZA01   1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0
```



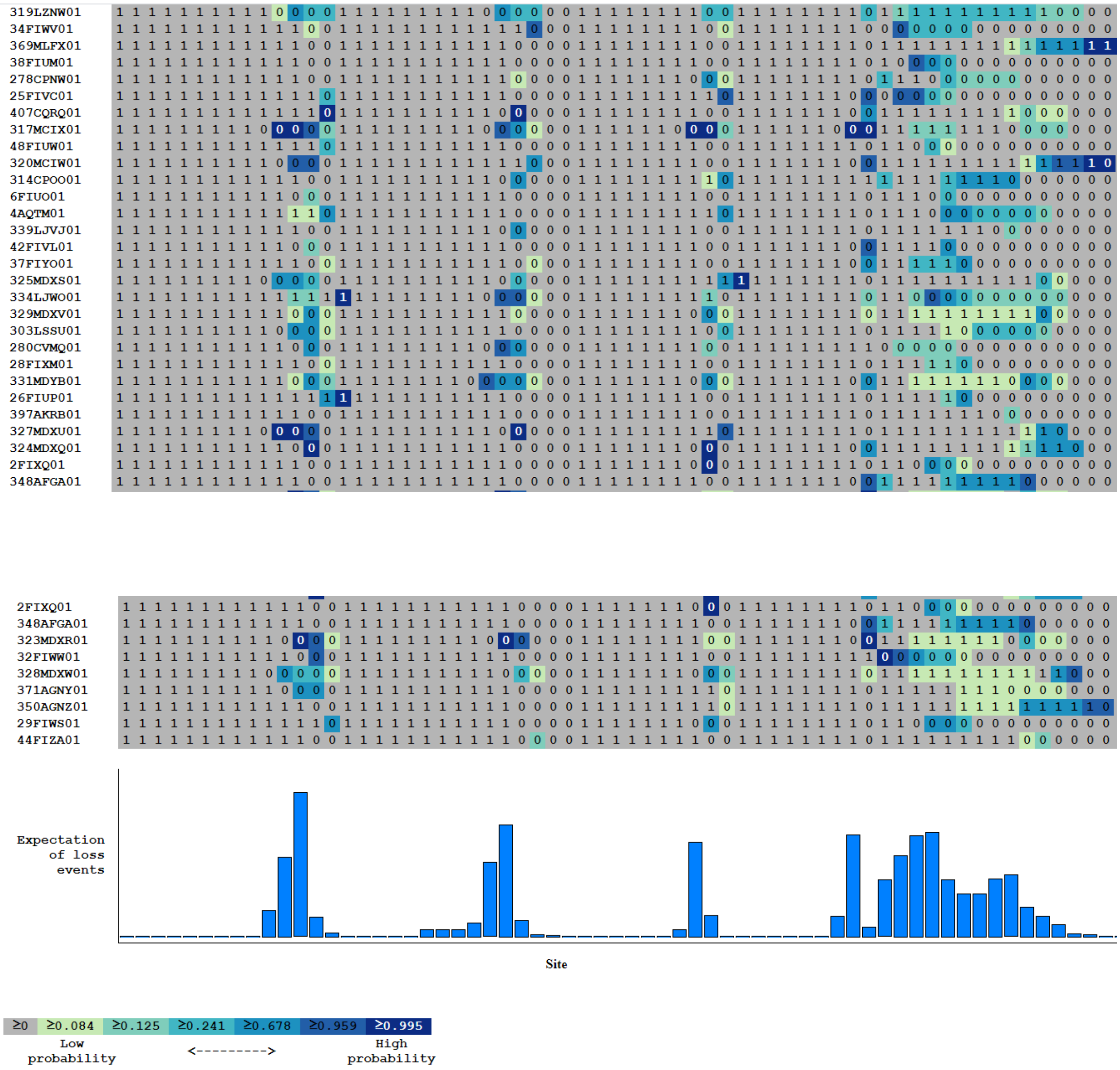
≥0	≥0.074	≥0.104	≥0.161	≥0.347	≥0.893	≥0.988

Low probability <---------> High probability



Phyletic pattern color-coded by loss probability

```
338LJVN01  1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
10FIUI01   1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0
3AQTO01    1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0
279LQQG01  1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
321LWGT01  1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0
337LJVT01  1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
433FCRD01  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0
21FIYL01   1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
283CPQN01  1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
7FIVV01    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0
23FIVR01   1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
36FIUE01   1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
281CVMS01  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
326MDXT01  1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0
12FIWM01   1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
47FIYG01   1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
368MLFY01  1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0
299NIFF01  1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0
333AGQB01  1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0
9FIXT01    1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0
330MDYA01  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0
13FIWR01   1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0
282CPRR01  1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
336LJWD01  1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0
2AQTP01    1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
347AFGD01  1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0
319LZNW01  1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0
34FIWV01   1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
369MLFX01  1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

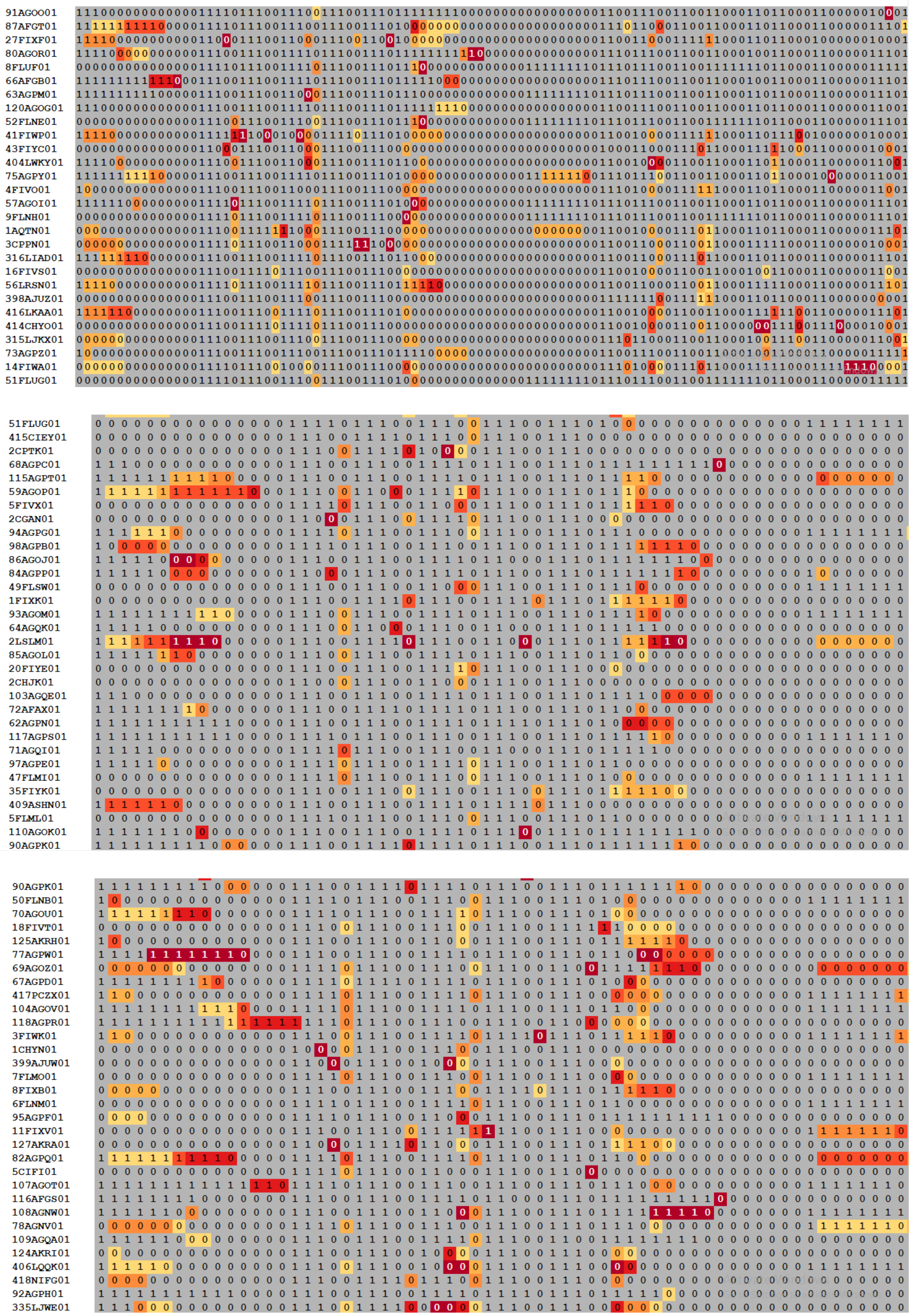


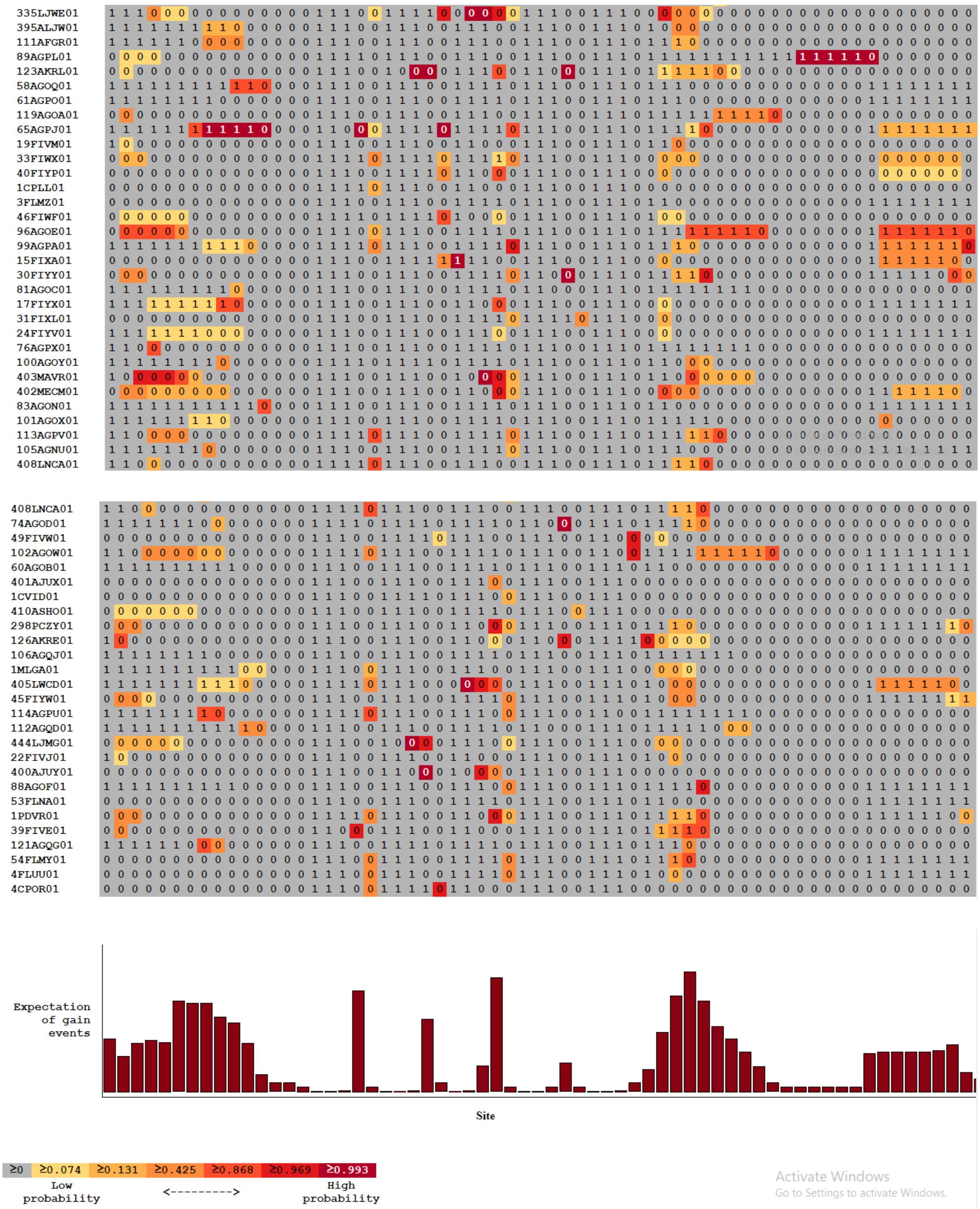
```
319LZNW01  1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0
34FIWV01   1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
369MLFX01  1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
38FIUM01   1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
278CPNW01  1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0
25FIVC01   1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
407CQRQ01  1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0
317MCIX01  1 1 1 1 1 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0
48FIUW01   1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0
320MCIW01  1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0
314CPOO01  1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0
6FIUO01    1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0
4AQTM01    1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0
339LJVJ01  1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
42FIVL01   1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0
37FIYO01   1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0
325MDXS01  1 1 1 1 1 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0
334LJWO01  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0
329MDXV01  1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0
303LSSU01  1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0
280CVMQ01  1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
28FIXM01   1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0
331MDYB01  1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0
26FIUP01   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0
397AKRB01  1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
327MDXU01  1 1 1 1 1 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0
324MDXQ01  1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0
2FIXQ01    1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0
348AFGA01  1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0
```

```
2FIXQ01    1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0
348AFGA01  1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0
323MDXR01  1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0
32FIWW01   1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
328MDXW01  1 1 1 1 1 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0
371AGNY01  1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0
350AGNZ01  1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0
29FIWS01   1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0
44FIZA01   1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0
```

Bacteremia gain and loss

Phyletic pattern color-coded by gain probability

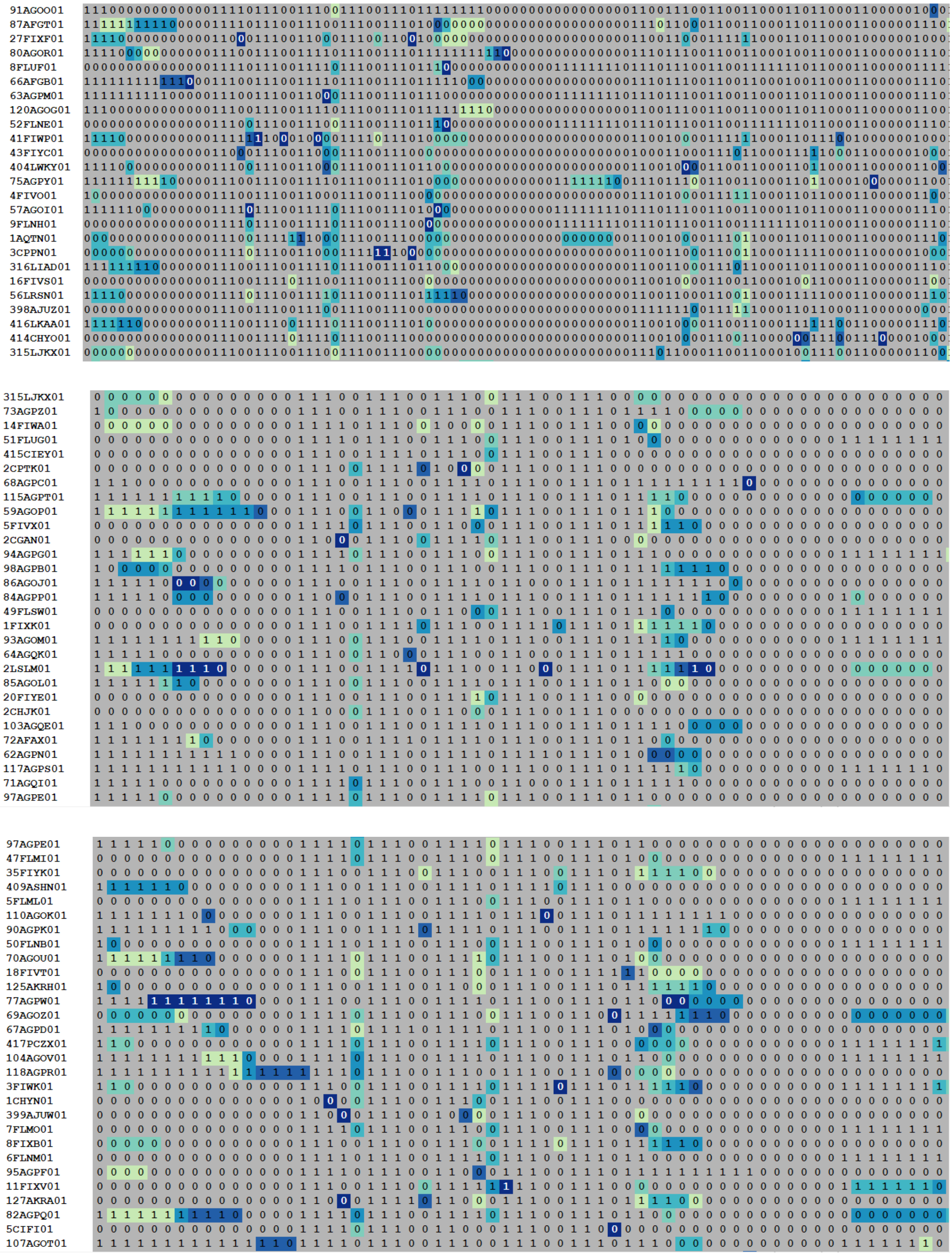


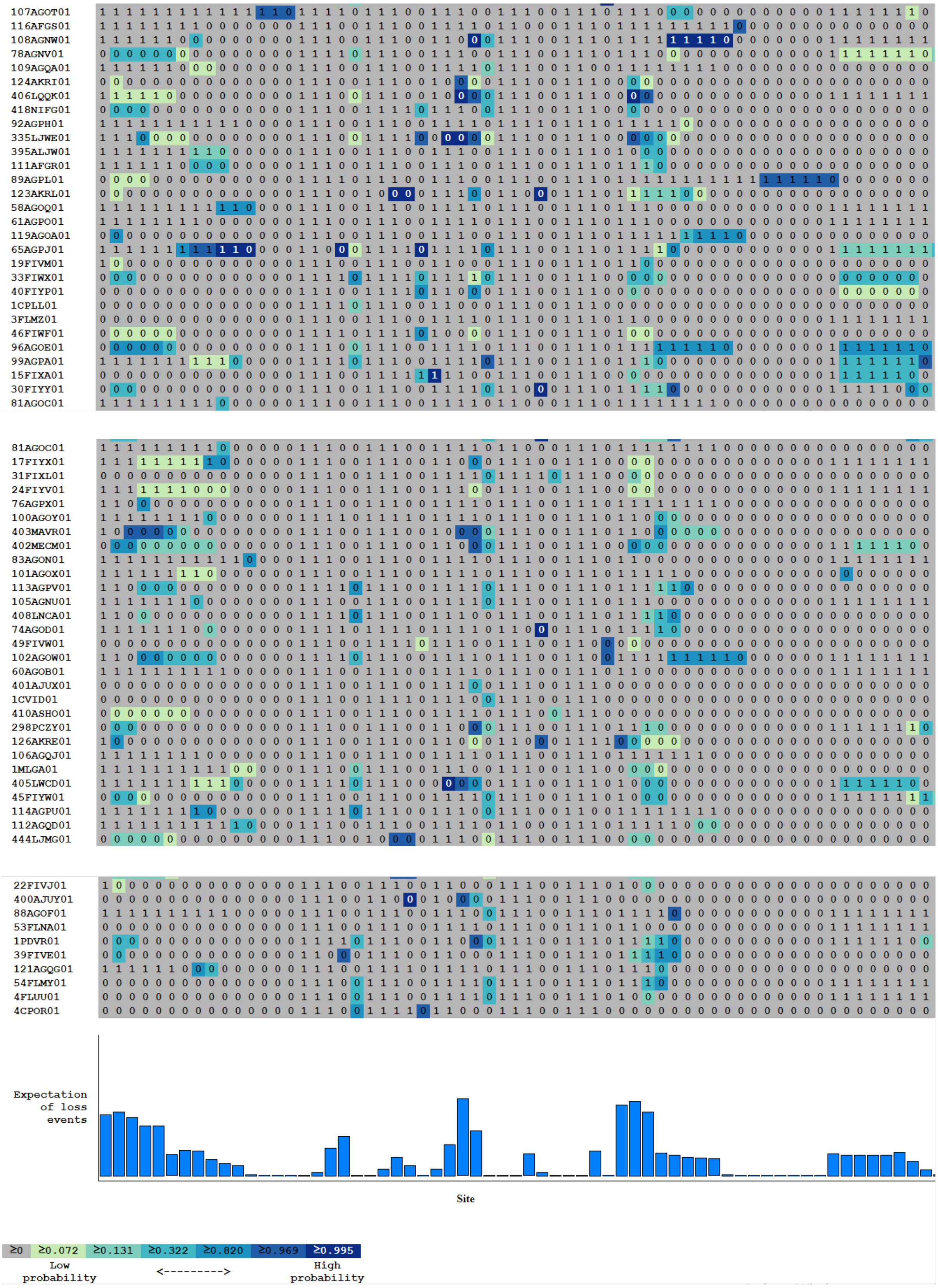


335LJWE01	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	1	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
395ALJW01	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
111AFGR01	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
89AGPL01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
123AKRL01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	0	0	0	0	1	1	1	0	0	1	1	0	0	0	1	1	1	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
58AGOQ01	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
61AGPO01	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
119AGOA01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
65AGPJ01	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	0	0	0	1	1	1	1	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
19FIVM01	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	0	0	0	1	1	1	0	0	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33FIWX01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1	1	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40FIYP01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	1	0	1	1	0	0	0	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1CPLL01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1	0	0	1	1	0	0	0	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3FLMZ01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
46FIWF01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1	1	0	1	0	0	0	0	1	1	1	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
96AGOE01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0
99AGPA01	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0
15FIXA01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	1	1	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0
30FIYY01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	1	0	1	1	0	0	0	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0
81AGOC01	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	1	0	1	1	0	0	0	1	1	1	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17FIYX01	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	0	0	0	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
31FIXL01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	1	0	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24FIYV01	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
76AGPX01	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
100AGOY01	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1	1	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
403MAVR01	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	0	0	0	0	1	1	1	0	0	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
402MECM01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	0	0	0	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0
83AGON01	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
101AGOX01	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
113AGPV01	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
105AGNU01	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
408LNCA01	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

408LNCA01	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
74AGOD01	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1	1	0	1	1	1	1	0	1	1	0	0	0	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
49FIVW01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
102AGOW01	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	1	0	1	1	1	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
60AGOB01	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
401AJUX01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1CVID01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
410ASHO01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
298PCZY01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	0	0	0	1	1	1	0	0	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0
126AKRE01	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	0	0	0	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
106AGQJ01	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1MLGA01	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
405LWCD01	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0
45FIYW01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
114AGPU01	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1	1	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
112AGQD01	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	1	0	1	1	0	0	0	1	1	1	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
444LJMG01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22FIVJ01	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	0	0	0	1	1	1	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
400AJUY01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	0	0	0	1	0	0	0	0	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
88AGOF01	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
53FLNA01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
1PDVR01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1	0	0	1	1	0	0	0	1	1	1	0	0	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0
39FIVE01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	1	1	0	0	1	1	0	0	0	1	1	1	0	0	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
121AGQG01	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	1	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
54FLMY01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
4FLUU01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
4CPOR01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	1	0	1	1	0	0	0	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

<u>**Phyletic pattern color-coded by loss probability**</u>




107AGOT01
116AFGS01
108AGNW01
78AGNV01
109AGQA01
124AKRI01
406LQQK01
418NIFG01
92AGPH01
335LJWE01
395ALJW01
111AFGR01
89AGPL01
123AKRL01
58AGOQ01
61AGPO01
119AGOA01
65AGPJ01
19FIVM01
33FIWX01
40FIYP01
1CPLL01
3FLMZ01
46FIWF01
96AGOE01
99AGPA01
15FIXA01
30FIYY01
81AGOC01
81AGOC01
17FIYX01
31FIXL01
24FIYV01
76AGPX01
100AGOY01
403MAVR01
402MECM01
83AGON01
101AGOX01
113AGPV01
105AGNU01
408LNCA01
74AGOD01
49FIVW01
102AGOW01
60AGOB01
401AJUX01
1CVID01
410ASHO01
298PCZY01
126AKRE01
106AGQJ01
1MLGA01
405LWCD01
45FIYW01
114AGPU01
112AGQD01
444LJMG01
22FIVJ01
400AJUY01
88AGOF01
53FLNA01
1PDVR01
39FIVE01
121AGQG01
54FLMY01
4FLUU01
4CPOR01
Expectation of loss events
Site
≥0
≥0.072
≥0.131
≥0.322
≥0.820
≥0.969
≥0.995
Low probability
<--------->
High probability

Appendix 4: Meningitis strains metadata

Assembly	WGS Sequencing Accession number	BioSample	Taxonomy (strain)	Serotype
GCA_000211875.2	AFGA00000000	SAMN00792759	*Streptococcus pneumoniae* GA17545	6B
GCA_000211915.2	AFGD00000000	SAMN00792778	*Streptococcus pneumoniae* GA41301	23F
GCA_000232025.2	AGNY00000000	SAMN00792792	*Streptococcus pneumoniae* GA44288	19A
GCA_000232045.2	AGNZ00000000	SAMN00762660	*Streptococcus pneumoniae* GA47281	19F
GCA_000233125.2	AGQB00000000	SAMN00792697	*Streptococcus pneumoniae* Netherlands15B-37	15B
GCA_000334655.1	AKRB00000000	SAMN02299524	*Streptococcus pneumoniae* PNI0006	N/A

GCA_000385775.1	AQTM00000000	SAMN02470834	*Streptococcus pneumoniae* 357	23F
GCA_000385755.1	AQTO00000000	SAMN02470835	*Streptococcus pneumoniae* 801	6
GCA_000385735.1	AQTP00000000	SAMN02470845	*Streptococcus pneumoniae* 845	6
GCF_001136245.1	CPNW00000000	SAMEA1020855	*Streptococcus pneumoniae* strain SN34677	6B
GCF_001085885.1	CPOO00000000	SAMEA1020830	*Streptococcus pneumoniae* strain SPN4876	6B
GCF_001130825.1	CPQN00000000	SAMEA1020037	*Streptococcus pneumoniae* strain 1014-00	6A
GCF_001329115.1	CVMS00000000	SAMEA1020187	*Streptococcus pneumoniae* strain SN27474	19F
GCF_900027575.1	FCRD00000000	SAMEA2239057	*Streptococcus pneumoniae* strain 2842STDY5644437	N/A
GCF_001103965.1	CPRR00000000	SAMEA1020711	*Streptococcus pneumoniae* strain 41_PMEN14	19F

GCF_001167925.1	CQRQ00000000	SAMEA970215	*Streptococcus pneumoniae* strain BHN647	3
GCF_001329135.1	CVMQ00000000	SAMEA1020205	*Streptococcus pneumoniae* strain SPN3402	6B
GCF_900055915.1	FIUO00000000	SAMEA867917	*Streptococcus pneumoniae* strain C13215X	1
GCF_900048875.1	FIUP00000000	SAMEA867918	*Streptococcus pneumoniae* strain C13181X	1
GCF_900063685.1	FIUE00000000	SAMEA867948	*Streptococcus pneumoniae* strain C11020	1
GCF_900050465.1	FIUI00000000	SAMEA867944	*Streptococcus pneumoniae* strain C14249	1
GCF_900050485.1	FIUM00000000	SAMEA867765	*Streptococcus pneumoniae* strain C14099X	1
GCF_900049715.1	FIVR00000000	SAMEA867837	*Streptococcus pneumoniae* strain B16827	10B

GCF_900053255.1	FIVV00000000	SAMEA867793	*Streptococcus pneumoniae* strain B16221X	12B
GCF_900052565.1	FIVC00000000	SAMEA867911	*Streptococcus pneumoniae* strain B14935	1
GCF_900061925.1	FIVL00000000	SAMEA867923	*Streptococcus pneumoniae* strain C16000X	1
GCF_900056895.1	FIUW00000000	SAMEA867834	*Streptococcus pneumoniae* strain B17333	1
GCF_900053845.1	FIWM00000000	SAMEA867795	*Streptococcus pneumoniae* strain B17731X	12F
GCF_900052595.1	FIWW00000000	SAMEA867909	*Streptococcus pneumoniae* strain B15188	23F
GCF_900054615.1	FIWR00000000	SAMEA867908	*Streptococcus pneumoniae* strain B15249	12B
GCF_900063735.1	FIWS00000000	SAMEA867852	*Streptococcus pneumoniae* strain B10622	15C

GCF_900049205.1	FIWV00000000	SAMEA867836	*Streptococcus pneumoniae* strain B16392	13
GCF_900055935.1	FIYO00000000	SAMEA867841	*Streptococcus pneumoniae* strain B15901	06C
GCF_900054635.1	FIYL00000000	SAMEA867906	*Streptococcus pneumoniae* strain C14215	06A
GCF_900049735.1	FIXM00000000	SAMEA867763	*Streptococcus pneumoniae* strain C14376X	25F/A
GCF_900052025.1	FIXQ00000000	SAMEA867913	*Streptococcus pneumoniae* strain B14721	35B
GCF_900051625.1	FIXT00000000	SAMEA867766	*Streptococcus pneumoniae* strain C14560X	4
GCF_900054625.1	FIYG00000000	SAMEA867916	*Streptococcus pneumoniae* strain C15085X	5
GCF_001581695.1	LJWO00000000	SAMN03964649	*Streptococcus pneumoniae* strain NTPn 44	N/A

GCF_900052045.1	FIZA00000000	SAMEA867854	*Streptococcus pneumoniae* strain B10027	09A
GCF_001581145.1	LJVJ00000000	SAMN03964619	*Streptococcus pneumoniae* strain NTPn 1	N/A
GCF_001581215.1	LJVN00000000	SAMN03964621	*Streptococcus pneumoniae* strain NTPn 3	N/A
GCF_001581295.1	LJVT00000000	SAMN03964627	*Streptococcus pneumoniae* strain NTPn 12	N/A
GCF_001581535.1	LJWD00000000	SAMN03964638	*Streptococcus pneumoniae* strain NTPn 30	N/A
GCF_001637405.1	LQQG00000000	SAMN04387653	*Streptococcus pneumoniae* strain CCUG 1350	6B
GCF_001715895.1	MCIX00000000	SAMN04337327	*Streptococcus pneumoniae* strain 29170-12F	12F
GCF_001578475.1	LSSU00000000	SAMN04496895	*Streptococcus pneumoniae* strain MTY32702340SN814	19A

GCF_001697005.1	LWGT00000000	SAMN04337322	*Streptococcus pneumoniae* strain 22522	N/A
GCF_001735995.1	LZNW00000000	SAMN04337326	*Streptococcus pneumoniae* strain 29170-6A	6A
GCF_001715865.1	MCIW00000000	SAMN04337323	*Streptococcus pneumoniae* strain 22421	N/A
GCF_001719775.1	MDXV00000000	SAMN04337308	*Streptococcus pneumoniae* strain 12985-14	14
GCF_001719975.1	MDXU00000000	SAMN04337310	*Streptococcus pneumoniae* strain 16599-9V	9V
GCF_001719815.1	MDXT00000000	SAMN04337311	*Streptococcus pneumoniae* strain 16599-15A	15A
GCF_001719885.1	MDXQ00000000	SAMN04337313	*Streptococcus pneumoniae* strain 18839-3	3
GCF_001719755.1	MDXR00000000	SAMN04337314	*Streptococcus pneumoniae* strain 18856	N/A

GCF_001719945.1	MDXS00000000	SAMN04337312	*Streptococcus pneumoniae* strain 18839-19A	19A
GCF_001982685.1	MLFX00000000	SAMN05912866	*Streptococcus pneumoniae* strain CCUG 35561	18C
GCF_001982715.1	MLFY00000000	SAMN05912978	*Streptococcus pneumoniae* strain CCUG 32672	14
GCF_002224185.1	NIFF00000000	SAMN07191007	*Streptococcus pneumoniae* strain TVM3	N/A
GCF_001719805.1	MDXW00000000	SAMN04337309	*Streptococcus pneumoniae* strain 12985-6A	6A
GCF_001719895.1	MDYA00000000	SAMN04337301	*Streptococcus pneumoniae* strain 7204	N/A
GCF_001719965.1	MDYB00000000	SAMN04337300	*Streptococcus pneumoniae* strain 7200	N/A

Appendix 5: Bacteremia strains metadata

Assembly	**WGS (Sequencing Accession Number)**	**BioSample**	**Taxonomy (strain)**	**Sero type**	**Collection date**	**Host disease**	**Geographical location**
GCA_000211895.2	AFGB00000000	SAMN00792760	*Streptococcus pneumoniae* GA17570	09V	3-Jan-01	Pneumonia	USA: Georgia
GCA_000194885.2	AFAX00000000	SAMN00792717	*Streptococcus pneumoniae* GA04375	19F	5-Feb-95	Bacteremia	USA: Georgia
GCA_000212515.2	AFGR00000000	SAMN00792680	*Streptococcus pneumoniae* GA47901	1	16-Aug-06	Pneumonia	USA: Georgia
GCA_000212535.2	AFGS00000000	SAMN00792663	*Streptococcus pneumonia*	19A	18-Mar-06	Bacteremia	USA: Georgia

			e GA47368				
GCA_000212555.2	AFGT00000000	SAMN00792779	*Streptococcus pneumoniae* GA41317	33F	23-Feb-04	Meningitis	USA: Georgia
GCA_000231945.2	AGNU00000000	SAMN00792729	*Streptococcus pneumoniae* GA11184	19F	19-Jan-99	N/A	USA: Georgia
GCA_000231965.2	AGNV00000000	SAMN00792669	*Streptococcus pneumoniae* GA47502	19A	5-Mar-06	Pneumonia	USA: Georgia
GCA_000231985.2	AGNW00000000	SAMN00792710	*Streptococcus pneumoniae* 4027-06	19A	2005	N/A	USA: Maryland
GCA_000232065.2	AGOA00000000	SAMN00792657	*Streptococcus pneumoniae* GA47033	06C	26-Dec-05	Bacteremia	USA: Georgia

GCA_000232085.2	AGOB00000000	SAMN00792788	*Streptococcus pneumoniae* GA43265	19A	11-May-05	Pneumonia	USA: Georgia
GCA_000232105.2	AGOC00000000	SAMN00792795	*Streptococcus pneumoniae* GA44452	19A	26-Apr-05	Pneumonia	USA: Georgia
GCA_000232125.2	AGOD00000000	SAMN00792682	*Streptococcus pneumoniae* GA49138	19F	7-Nov-06	Pneumonia	USA: Georgia
GCA_000232145.2	AGOE00000000	SAMN00792750	*Streptococcus pneumoniae* GA16531	06B	26-Jun-01	Bacteremia	USA: Georgia
GCA_000232165.2	AGOF00000000	SAMN00792776	*Streptococcus pneumoniae* 6901-05	19A	2005	N/A	USA: Connecticut

GCA_000232185.2	AGOG00000000	SAMN00792656	*Streptococcus pneumoniae* 7286-06	19A	2005	N/A	USA: Maryland
GCA_000232225.2	AGOI00000000	SAMN00792796	*Streptococcus pneumoniae* GA44500	19A	9-May-05	Bacteremia	USA: Georgia
GCA_000232245.2	AGOJ00000000	SAMN00792780	*Streptococcus pneumoniae* GA41410	19A	25-Mar-04	Pneumonia	USA: Georgia
GCA_000232265.2	AGOK00000000	SAMN00792684	*Streptococcus pneumoniae* GA49447	19A	21-Dec-06	Pneumonia	USA: Georgia
GCA_000232285.2	AGOL00000000	SAMN00792782	*Streptococcus pneumoniae* GA41538	06B	30-Apr-04	Pneumonia	USA: Georgia

GCA_000232305.2	AGOM00000000	SAMN00792754	*Streptococcus pneumoniae* 5787-06	19A	2005	N/A	USA: New York
GCA_000232345.2	AGOO00000000	SAMN00792764	*Streptococcus pneumoniae* GA18523	19A	29-Dec-01	Bacteremia	USA: Georgia
GCA_000232365.2	AGOP00000000	SAMN00792791	*Streptococcus pneumoniae* GA44194	19A	9-Feb-05	Bacteremia	USA: Georgia
GCA_000232385.2	AGOQ00000000	SAMN00792793	*Streptococcus pneumoniae* GA44378	23F	2-Apr-05	Pneumonia	USA: Georgia
GCA_000232325.2	AGON00000000	SAMN00792787	*Streptococcus pneumoniae* 6963-05	19A	2005	N/A	USA: Minnesota

GCA_000232405.2	AGOR00000000	SAMN00792797	*Streptococcus pneumoniae* GA44511	19A	19-May-05	Pneumonia	USA: Georgia
GCA_000232445.2	AGOT00000000	SAMN00792725	*Streptococcus pneumoniae* GA07643	4	22-Feb-98	Pneumonia	USA: Georgia
GCA_000232525.2	AGOX00000000	SAMN00792736	*Streptococcus pneumoniae* GA13338	14	26-Feb-99	N/A	USA: Georgia
GCA_000232465.2	AGOU00000000	SAMN00792730	*Streptococcus pneumoniae* GA11304	06B	7-Mar-99	Pneumonia	USA: Georgia
GCA_000232485.2	AGOV00000000	SAMN00792731	*Streptococcus pneumoniae* GA11426	19A	18-May-99	Pneumonia	USA: Georgia
GCA_000232505.2	AGOW00000000	SAMN00792733	*Streptococcus*	19F	23-Dec-99	Bacteremia	USA: Georgia

			pneumoniae GA11663				
GCA_000232545.2	AGOY00000000	SAMN00792738	*Streptococcus pneumoniae* GA13455	19F	31-Mar-99	Pneumonia	USA: Georgia
GCA_000232565.2	AGOZ00000000	SAMN00792739	*Streptococcus pneumoniae* GA13494	14	10-Apr-99	Septic arthritis, Septic knee	USA: Georgia
GCA_000232585.2	AGPA00000000	SAMN00792741	*Streptococcus pneumoniae* GA13637	18C	15-May-99	Bacteremia	USA: Georgia
GCA_000232605.2	AGPB00000000	SAMN00792744	*Streptococcus pneumoniae* GA13856	33F	15-Aug-99	Bacteremia	USA: Georgia
GCA_000232625.2	AGPC00000000	SAMN00792747	*Streptococcus pneumonia*	19F	7-Apr-00	Bacteremia	USA: Georgia

			e GA14798				
GCA_000232645.2	AGPD00000000	SAMN00792748	*Streptococcus pneumoniae* GA16121	19F	3-Nov-00	Pneumonia	USA: Georgia
GCA_000232665.2	AGPE00000000	SAMN00792749	*Streptococcus pneumoniae* GA16242	06B	16-Jan-01	Bacteremia	USA: Georgia
GCA_000232685.2	AGPF00000000	SAMN00792751	*Streptococcus pneumoniae* GA16833	19F	15-Feb-02	Pneumonia	USA: Georgia
GCA_000232705.2	AGPG00000000	SAMN00792752	*Streptococcus pneumoniae* GA17227	23F	14-Sep-00	Pneumonia	USA: Georgia
GCA_000232725.2	AGPH00000000	SAMN00792755	*Streptococcus pneumonia*	6A	20-Oct-00	N/A	USA: Georgia

			e GA17328				
GCA_000232765.2	AGPJ00000000	SAMN00792762	*Streptococcus pneumoniae* GA17971	06A	10-Apr-01	Bacteremia	USA: Georgia
GCA_000232785.2	AGPK00000000	SAMN00792766	*Streptococcus pneumoniae* GA19077	06A	9-Aug-02	Pneumonia	USA: Georgia
GCA_000232805.2	AGPL00000000	SAMN00792768	*Streptococcus pneumoniae* GA19451	19F	22-May-03	Cellulitis	USA: Georgia
GCA_000232825.2	AGPM00000000	SAMN00792777	*Streptococcus pneumoniae* GA41277	19A	8-Feb-04	Pneumonia	USA: Georgia
GCA_000232845.2	AGPN00000000	SAMN00792781	*Streptococcus pneumonia*	06A	2-Apr-04	Pneumonia	USA: Georgia

			e GA41437				
GCA_000232865.2	AGPO00000000	SAMN00792783	*Streptococcus pneumoniae* GA41565	19A	12-May-04	Pneumonia	USA: Georgia
GCA_000232885.2	AGPP00000000	SAMN00792784	*Streptococcus pneumoniae* GA41688	14	19-Jul-04	Pneumonia	USA: Georgia
GCA_000232905.2	AGPQ00000000	SAMN00792789	*Streptococcus pneumoniae* GA43380	19A	20-Sep-05	Pneumonia	USA: Georgia
GCA_000232925.2	AGPR00000000	SAMN00792661	*Streptococcus pneumoniae* GA47283	07F	24-Feb-06	Pneumonia	USA: Georgia
GCA_000232945.2	AGPS00000000	SAMN00792662	*Streptococcus pneumonia*	19A	16-Mar-06	Septic shock,	USA: Georgia

			e GA47360			Pneumonia	
GCA_000232965.2	AGPT00000000	SAMN00792664	*Streptococcus pneumoniae* GA47373	19F	16-Mar-06	Pneumonia	USA: Georgia
GCA_000232985.2	AGPU00000000	SAMN00792665	*Streptococcus pneumoniae* GA47388	19A	21-Mar-06	Septic shock, Pneumonia	USA: Georgia
GCA_000233005.2	AGPV00000000	SAMN00792666	*Streptococcus pneumoniae* GA47439	07F	4-Apr-06	Pneumonia	USA: Georgia
GCA_000233025.2	AGPW00000000	SAMN00792674	*Streptococcus pneumoniae* GA47688	19A	8-Jun-06	Pneumonia	USA: Georgia
GCA_000233045.2	AGPX00000000	SAMN00792677	*Streptococcus pneumonia*	19A	6-Jul-06	Pneumonia	USA: Georgia

			e GA47778				
GCA_000233065.2	AGPY00000000	SAMN00792681	*Streptococcus pneumoniae* GA47976	19F	11-Sep-06	Pneumonia	USA: Georgia
GCA_000233085.2	AGPZ00000000	SAMN00792686	*Streptococcus pneumoniae* GA52306	06C	11-Aug-07	Bacteremia	USA: Georgia
GCA_000233105.2	AGQA00000000	SAMN00792690	*Streptococcus pneumoniae* GA54644	19A	17-Aug-08	Pneumonia	USA: Georgia
GCA_000233165.2	AGQD00000000	SAMN00792675	*Streptococcus pneumoniae* GA47751	19A	24-Jun-06	Pneumonia	USA: Georgia
GCA_000233185.2	AGQE00000000	SAMN00792732	*Streptococcus*	19A	2005	N/A	USA: Maryland

			pneumoniae 5185-06				
GCA_000233225.2	AGQG00000000	SAMN00792655	*Streptococcus pneumoniae* 3063-00	19F	1999	Bacteremia	USA: Tennessee
GCA_000233265.2	AGQI00000000	SAMN00792724	*Streptococcus pneumoniae* GA07228	3	4-Jun-97	Bacteremia	USA: Georgia
GCA_000233285.2	AGQJ00000000	SAMN00792727	*Streptococcus pneumoniae* GA08780	09V	22-Dec-97	Bacteremia	USA: Georgia
GCA_000233305.2	AGQK00000000	SAMN00792769	*Streptococcus pneumoniae* GA19690	3	12-Jan-04	Pneumonia	USA: Georgia
GCA_000355985.1	AJUW00000000	SAMN02436862	*Streptococcus pneumoniae* PNI0197	N/A	N/A	N/A	N/A

GCA_000355965.1	AJUX00000000	SAMN02436798	*Streptococcus pneumoniae* PNI0159	N/A	N/A	N/A	N/A
GCA_000355945.1	AJUY00000000	SAMN02436799	*Streptococcus pneumoniae* PNI0164	N/A	N/A	N/A	N/A
GCA_000355925.1	AJUZ00000000	SAMN02436684	*Streptococcus pneumoniae* PNI0212	N/A	N/A	N/A	N/A
GCA_000334635.1	AKRA00000000	SAMN02299523	*Streptococcus pneumoniae* PNI0002	N/A	N/A	N/A	N/A
GCA_000334715.1	AKRE00000000	SAMN02299527	*Streptococcus pneumoniae* PNI0009	N/A	N/A	N/A	N/A
GCA_000334775.1	AKRH00000000	SAMN02299530	*Streptococcus pneumoniae* PNI0153	N/A	N/A	N/A	N/A

GCA_000334795.1	AKRI00000000	SAMN02299531	*Streptococcus pneumoniae* PNI0199	N/A	N/A	N/A	N/A
GCA_000334855.1	AKRL00000000	SAMN02299534	*Streptococcus pneumoniae* PNI0446	N/A	N/A	N/A	N/A
GCA_000348705.1	ALJW00000000	SAMN02436861	*Streptococcus pneumoniae* PCS8235	N/A	N/A	N/A	N/A
GCA_000385795.1	AQTN00000000	SAMN02470853	*Streptococcus pneumoniae* 2009	22F/A	Sep-08	N/A	Russia
GCA_000495335.1	ASHN00000000	SAMN02471309	*Streptococcus pneumoniae* BHN191	6B	1999	N/A	Sweden: Stockholm
GCA_000495395.1	ASHO00000000	SAMN02471308	*Streptococcus pneumoniae* BHN237	6B	2001	N/A	Sweden: Stockholm

GCA_001330795.1	CGAN00000000	SAMEA682579	*Streptococcus pneumoniae* strain M01_9995	19F	1999	N/A	Malaysia
GCA_001330675.1	CHJK00000000	SAMEA682560	*Streptococcus pneumoniae* strain spnIC203	6B	1994	N/A	Iceland: Reykjavik
GCA_001330695.1	CHYN00000000	SAMEA682563	*Streptococcus pneumoniae* strain spnIC192	6B	1993	Otitis media	Iceland: Eyjafjordur
GCA_001330055.1	CHYO00000000	SAMEA682310	*Streptococcus pneumoniae* strain Tw01_0057	19F	2000	N/A	China: Taiwan
GCA_001330075.1	CIEY00000000	SAMEA682312	*Streptococcus pneumoniae* strain	19F	2000	N/A	China: Taiwan

			Tw01_0059				
GCA_001329535.1	CIFI00000000	SAMEA1020785	*Streptococcus pneumoniae* strain LMG2290	6B	2008	Pneumonia	Peru: Lima
GCA_001150665.1	CPLL00000000	SAMEA1020637	*Streptococcus pneumoniae* strain LMG2230	6B	2007	Pneumonia	Peru: Lima
GCA_001115005.1	CPOR00000000	SAMEA1020747	*Streptococcus pneumoniae* strain LMG2302	6B	2009	Pneumonia	Peru: Lima
GCA_001087645.1	CPPN00000000	SAMEA1020584	*Streptococcus pneumoniae* strain LMG2311	6B	2009	Pneumonia	Peru: Lima
GCA_001098045.1	CPTK00000000	SAMEA1020797	*Streptococcus pneumonia*	6B	2009	Bacteremia	Peru: Lima

			e strain LMG3367				
GCA_001330935.1	CVID00000000	SAMEA682608	*Streptococcus pneumoniae* strain V01_99112	19F	1999	N/A	Viet Nam
GCA_900061915.1	FIVE00000000	SAMEA867781	*Streptococcus pneumoniae* strain A33973	1	1800/2014	N/A	Malawi
GCA_900050925.1	FIVJ00000000	SAMEA867780	*Streptococcus pneumoniae* strain A34030	1	1800/2014	N/A	Malawi
GCA_900063675.1	FIVM00000000	SAMEA867783	*Streptococcus pneumoniae* strain A34045	1	1800/2014	N/A	Malawi
GCA_900051305.1	FIVO00000000	SAMEA867806	*Streptococcus*	1	1800/2014	N/A	Malawi

			pneumoniae strain A28816				
GCA_900065845.1	FIVS00000000	SAMEA867801	*Streptococcus pneumoniae* strain A29037	10B	1800/2014	N/A	Malawi
GCA_900051605.1	FIVT00000000	SAMEA867798	*Streptococcus pneumoniae* strain A31131	12B	1800/2014	N/A	Malawi
GCA_900063725.1	FIVW00000000	SAMEA867777	*Streptococcus pneumoniae* strain D31621X	10B	1800/2014	N/A	Malawi
GCA_900050505.1	FIWA00000000	SAMEA867768	*Streptococcus pneumoniae* strain D25696X	14	1800/2014	N/A	Malawi
GCA_900048885.1	FIVX00000000	SAMEA867802	*Streptococcus*	12B	1800/2014	N/A	Malawi

			pneumoniae strain A29943				
GCA_900049215.1	FIWF00000000	SAMEA867807	*Streptococcus pneumoniae* strain A28640	14	1800/2014	N/A	Malawi
GCA_900052885.1	FIWK00000000	SAMEA867950	*Streptococcus pneumoniae* strain D28531	18A	1800/2014	N/A	Malawi
GCA_900054135.1	FIWP00000000	SAMEA867926	*Streptococcus pneumoniae* strain D30625	23F	1800/2014	N/A	Malawi
GCA_900052575.1	FIXA00000000	SAMEA867925	*Streptococcus pneumoniae* strain D30716	16F	1800/2014	N/A	Malawi
GCA_900053295.1	FIXB00000000	SAMEA867881	*Streptococcus*	23F	1800/2014	N/A	Malawi

			pneumoniae strain A33308				
GCA_900053285.1	FIWX00000000	SAMEA867878	*Streptococcus pneumoniae* strain D28166	18B	1800/2014	N/A	Malawi
GCA_900052615.1	FIXF00000000	SAMEA867770	*Streptococcus pneumoniae* strain D26316X	23F	1800/2014	N/A	Malawi
GCA_900058795.1	FIXK00000000	SAMEA867762	*Streptococcus pneumoniae* strain D38094X	25F	1800/2014	N/A	Malawi
GCA_900063745.1	FIXL00000000	SAMEA867877	*Streptococcus pneumoniae* strain D24847	25F	1800/2014	N/A	Malawi
GCA_900051635.1	FIXV00000000	SAMEA867800	*Streptococcus*	5	1800/2014	N/A	Malawi

			pneumoniae strain A29101				
GCA_900053875.1	FIYE00000000	SAMEA867930	*Streptococcus pneumoniae* strain D36051	06A	1800/2014	N/A	Malawi
GCA_900055565.1	FIYC00000000	SAMEA867924	*Streptococcus pneumoniae* strain D33275	06B	1800/2014	N/A	Malawi
GCA_900057095.1	FIYK00000000	SAMEA867905	*Streptococcus pneumoniae* strain D38023	6D	1800/2014	N/A	Malawi
GCA_900053905.1	FIYP00000000	SAMEA867785	*Streptococcus pneumoniae* strain A34562	09A	1800/2014	N/A	Malawi
GCA_900055245.1	FIYV00000000	SAMEA867799	*Streptococcus*	09A	1800/2014	N/A	Malawi

			pneumoniae strain A30277				
GCA_900055255.1	FIYW00000000	SAMEA867771	*Streptococcus pneumoniae* strain A34292	09A	1800/2014	N/A	Malawi
GCA_900056905.1	FIYX00000000	SAMEA867803	*Streptococcus pneumoniae* strain A29167	09A	1800/2014	N/A	Malawi
GCA_900058545.1	FIYY00000000	SAMEA867882	*Streptococcus pneumoniae* strain A33813	09A	1800/2014	N/A	Malawi
GCA_900088715.1	FLMI00000000	SAMEA4020771	*Streptococcus pneumoniae* isolate 246	N/A	1998	N/A	Germany
GCA_900088775.1	FLML00000000	SAMEA4021824	*Streptococcus*	N/A	2000	N/A	Germany

			pneumoniae isolate 762				
GCA_900088805.1	FLMO00000000	SAMEA4021842	*Streptococcus pneumoniae* isolate 20253	N/A	2004	N/A	Germany
GCA_900089065.1	FLMY00000000	SAMEA4022017	*Streptococcus pneumoniae* isolate 21295	N/A	2004	N/A	Germany
GCA_900089045.1	FLMZ00000000	SAMEA4022022	*Streptococcus pneumoniae* isolate 23543	N/A	2005	N/A	Germany
GCA_900088865.1	FLNA00000000	SAMEA4022019	*Streptococcus pneumoniae* isolate 21299	N/A	2004	N/A	Germany
GCA_900088885.1	FLNB00000000	SAMEA4025212	*Streptococcus*	N/A	2010	N/A	Germany

			pneumoniae isolate 46770				
GCA_900089105.1	FLNE00000000	SAMEA4025132	*Streptococcus pneumoniae* isolate 44017	N/A	2009	N/A	Germany
GCA_900088935.1	FLNH00000000	SAMEA4025166	*Streptococcus pneumoniae* isolate 43003	N/A	2010	N/A	Germany
GCA_900089485.1	FLNM00000000	SAMEA4027047	*Streptococcus pneumoniae* isolate 20605	N/A	2003	N/A	Germany
GCA_900092055.1	FLSW00000000	SAMEA4051576	*Streptococcus pneumoniae* isolate 27700	N/A	2006	N/A	Germany
GCA_900088995.1	FLUF00000000	SAMEA4025213	*Streptococcus*	N/A	2010	N/A	Germany

			pneumoniae isolate 46464				
GCA_900088975.1	FLUG00000000	SAMEA4025169	*Streptococcus pneumoniae* isolate 46048	N/A	2010	N/A	Germany
GCA_900088725.1	FLUU00000000	SAMEA4021833	*Streptococcus pneumoniae* isolate 893	N/A	2001	N/A	Germany
GCA_001584785.1	LIAD00000000	SAMN03946353	*Streptococcus pneumoniae* strain MT1	12F	2007	Meningitis	Canada: Manitoba
GCA_001585805.1	LJKX00000000	SAMN03946366	*Streptococcus pneumoniae* strain MT14	12F	2010	Meningitis	Canada: Manitoba
GCA_001585635.1	LJMG00000000	SAMN03946413	*Streptococcus*	12F	2006/2007	Septicemia	Spain: Madrid

			pneumoniae strain SN2206				
GCA_001581545.1	LJWE00000000	SAMN03964639	*Streptococcus pneumoniae* strain NTPn 31	N/A	2008	Meningitis	South Africa: Kwazulu-Natal
GCA_001560935.1	LKAA00000000	SAMN04101983	*Streptococcus pneumoniae* strain MTY1662 SN214	19A	2014	Pneumonia	Mexico: Monterrey
GCA_002016615.1	LNCA00000000	SAMN04259810	*Streptococcus pneumoniae* strain CCUG 63093	N/A	22-Jun-12	Bacteremia	Sweden: Vastra Gotaland, Gothenburg
GCA_001637485.1	LQQK00000000	SAMN04387839	*Streptococcus pneumoniae* strain	19A	12/7/1995	N/A	Sweden: Vastra Gotaland, Gothenburg

			CCUG 35180				
GCA_001545505.1	LRSN00000000	SAMN04440609	*Streptococcus pneumoniae* strain 225994	6B	6/8/2015	Pneumonia	Israel: Afula
GCA_001678985.1	LSLM00000000	SAMN04481652	*Streptococcus pneumoniae* strain CCUG 6798	3	1977-12	Pneumonia	Sweden: Vastra Gotaland, Gothenburg
GCA_001639345.1	LWCD00000000	SAMN04623576	*Streptococcus pneumoniae* strain CCUG 7206	N/A	1978	N/A	Sweden: Vastra Gotaland, Gothenburg
GCA_001642845.1	LWKY00000000	SAMN04859023	*Streptococcus pneumoniae* strain CCUG 63665	6A	12/12/2006	N/A	Sweden: Vastra Gotaland, Gothenburg

GCA_001856065.1	MAVR00000000	SAMN05366909	*Streptococcus pneumoniae* strain CCUG 36618	9V	8/20/1996	N/A	Sweden: Vastra Gotaland, Gothenburg
GCA_001870645.1	MECM00000000	SAMN05715961	*Streptococcus pneumoniae* strain CCUG 45673	19F	9/18/2001	N/A	Sweden:Uddevalla
GCA_001982705.1	MLGA00000000	SAMN05912981	*Streptococcus pneumoniae* strain CCUG 36800	7F	10/8/1996	N/A	Sweden:Gothenburg
GCA_002224195.1	NIFG00000000	SAMN07191006	*Streptococcus pneumoniae* strain B44415	N/A	2014	Septicemia	India: Vellore
GCA_002529515.1	PCZX00000000	SAMN07736520	*Streptococcus*	N/A	2012	Sepsis	India: Vellore

			pneumoniae strain CMC651				
GCA_002529495.1	PCZY00000000	SAMN07736519	*Streptococcus pneumoniae* strain CMC331	N/A	2012	Meningitis	India: Vellore
GCA_002718275.1	PDVR00000000	SAMN07811907	*Streptococcus pneumoniae* strain 199_17	4	2017	Bacteremia	Brazil: Sao Paulo

Appendix 6: Meningitis ortholog files part

cluster.vs.genome	2.FIXQ01	2.AQTP01	3.AQTO01	4.AQTM01	6.FIUO01	7.FIVV01	9.FIXT01	10.FIUI01	12.FIWM01	13.FIWR01	21.FIYL01	23.FIVR01	25.FIVC01	26.FIUP01	28.FIXM01	29.FIWS01	32.FIWW01	34.FIWV01	36.FIUE01
0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

0	1	0	1	1	0	1	1	1	1	0	1	1	1	1	1	1	0	1	0
0	0	0	0	1	0	1	1	0	0	0	1	0	1	1	0	1	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	0	0	0	1	0	1	1	1	1	1	1	1
1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0
1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

2	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
2	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
2	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
2	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
2	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
2	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
2	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**0**
2	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**
2	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
3	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
3	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
3	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
3	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
3	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
3	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
3	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
3	**1**	**1**	**1**	**1**	**1**	**0**	**1**	**0**	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**1**	**1**
3	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**
4	**1**	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**0**	**1**	**1**	**1**	**0**	**0**	**1**

4	**1**	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**0**	**0**	**1**	**1**	**1**	**0**	**0**	**1**
4	**0**	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**1**
4	**0**	**0**	**1**	**0**	**0**	**1**	**1**	**1**	**1**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**1**
4	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**1**
4	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**
4	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**
4	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
4	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
4	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
4	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
4	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
4	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
4	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
4	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
5	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
5	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
5	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
5	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
5	**1**	**1**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**

5	**0**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**0**	**0**	**1**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**
6	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
6	**0**	**0**	**0**	**1**	**0**	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**1**	**0**
6	**0**	**0**	**0**	**1**	**0**	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**1**	**0**
6	**0**	**0**	**0**	**1**	**0**	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**1**	**0**
6	**0**	**0**	**0**	**1**	**0**	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**1**	**0**
6	**0**	**0**	**0**	**1**	**0**	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**1**	**0**
6	**0**	**0**	**0**	**1**	**0**	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**1**	**0**
6	**0**	**0**	**0**	**1**	**0**	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**1**	**0**
6	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**1**	**0**
6	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**0**
6	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**0**
6	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
7	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
7	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
7	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
7	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
7	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**1**	**1**	**0**	**0**	**0**
8	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**

8	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
8	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
8	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
8	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**
9	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
9	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
9	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
9	**1**	**0**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**1**	**1**	**0**	**0**	**0**	**0**	**1**	**1**	**1**
9	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**

Appendix 7: Bacteremia ortholog files part

cluster/gene	1.CHYN01	1.PDVR01	1.MLGA01	1.AQTN01	1.CVID01	1.FIXK01	1.CPLL01	2.CPTK01	2.CHJK01	2.CGAN01	2.LSLM01	3.FLMZ01	3.FIWK01	3.CPPN01	4.CPOR01	4.FIVO01	4.FLUU01	5.FLML01	5.FIVX01
0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0
0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

1	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
1	**0**	**1**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**1**	**1**
1	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
2	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
2	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
2	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
2	**0**	**0**	**0**	**1**	**1**	**1**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**
2	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
3	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
3	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
3	**1**	**0**	**1**	**0**	**1**	**1**	**0**	**0**	**0**	**1**	**1**	**1**	**1**	**0**	**0**	**0**	**1**	**1**	**0**
3	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**1**	**1**	**0**	**0**	**0**	**1**	**0**	**0**
3	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
4	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
4	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
4	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
4	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**0**	**0**
4	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**
5	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
5	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
5	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**1**
5	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
6	**0**	**1**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**1**	**1**	**1**	**0**	**0**	**0**	**1**	**1**	**1**
6	**0**	**1**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**1**	**1**	**1**	**0**	**0**	**0**	**0**	**1**	**1**

6	**0**	**1**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**1**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**1**
6	**0**	**1**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**1**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**1**
6	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**1**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**1**
6	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
6	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
6	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
6	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
6	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
6	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
6	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
6	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
6	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
6	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
6	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
7	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**1**	**1**	**0**
7	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**1**	**1**	**0**
7	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**1**	**1**	**0**
7	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**1**	**1**	**0**
7	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**1**	**1**	**0**
7	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**1**	**1**	**0**
7	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**1**	**1**	**0**
7	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**1**	**1**	**0**
7	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
8	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**

8	**1**	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
8	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**0**	**1**	**1**	**1**	**0**
8	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
9	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
9	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**0**
9	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**1**	**1**	**0**
9	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
9	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
10	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
10	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
10	**1**	**1**	**0**	**1**	**1**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**1**	**0**	**1**	**1**	**0**	**0**	**0**
10	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**0**
11	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
11	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
11	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**
11	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**
11	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**
12	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
12	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
12	**0**	**1**	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**1**	**0**	**0**	**0**	**0**
13	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
13	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
13	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**
13	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**

13	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
14	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
14	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
14	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
14	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
14	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
14	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
14	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**