Parsing the Streptococcus pneumoniae virulome

Author: Emily Rudmann

Persistent link: http://hdl.handle.net/2345/bc-ir:108795

This work is posted on eScholarship@BC,
Boston College University Libraries.

Boston College Electronic Thesis or Dissertation, 2020

Copyright is held by the author, with all rights reserved, unless otherwise noted.

Parsing the *Streptococcus pneumoniae* virulome

Emily Rudmann

An Undergraduate thesis

submitted to the Faculty of

the department of biology

in partial fulfillment

of the requirements for the degree of

Bachelor of Science in Biology

Boston College
Morrissey College of Arts and Sciences

May 2020

© Copyright 2020 Emily Rudmann

PARSING THE *STREPTOCOCCUS PNEUMONIAE* VIRULOME

Emily Rudmann

Advisors: Federico Rosconi, Ph.D. and Tim van Opijnen, Ph.D.

Streptococcus pneumoniae is a prominent gram-positive commensal and opportunistic pathogen which possesses a large pan-genome. Significant strain-to-strain variability in genomic content drives the use of varied pathways and mechanisms to perform similar processes between strains, as our group has demonstrated in antibiotic stress response. Considering this variation, we employ a set of 30 *S. pneumoniae* clinical isolates and 6 well-studied strains, representative of 78% of total pan-genome diversity, with which to perform functional studies. This pan-genome set enables us to uncover common and divergent mechanisms used by *S. pneumoniae* strains to carry out crucial processes. Our group previously constructed transposon libraries in 22 of the 36 strains, representative of 73% of global diversity, and determined by *in vivo* Tn-seq in a mouse model the set of genes required by these strains to maintain successful infection in the host, or the virulome. Within the host, *S. pneumoniae* must perform a variety of functions; in this work, we sought to parse from the virulome the genes required specifically for nasopharyngeal adhesion, a crucial step in *S. pneumoniae* colonization and transmission, and often a precursor to invasive disease, as well as gene requirements for subversion of the macrophage. We performed *in vitro* attachment Tn-seq in the 22 pan-genome strains to D562 human nasopharyngeal epithelial cells, identifying thirteen factors that exhibit core or strain-dependent requirements for adhesion, and preliminarily validated a proposed universal requirement for survival of the macrophage by a killing assay using J774A.1 murine migratory macrophages. Considering that our pan-genome set includes both virulent and non-virulent isolates, this work provides a window into a holistic view of pneumococcal carriage and disease states by investigation of bacterial interactions with two host cell types which play integral roles in the establishment and progression of infection.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank everyone who guided and supported me in completing this thesis. Thanks to everyone within the van Opijnen lab for creating an exciting and enriching learning environment which has helped me grow and discover my passion for research during the past two and a half years. Specific to this project, thank you to Dr. Federico Rosconi for isolating the transposase for transposon mutant library construction, and to Jien Li for his help in constructing them in the 22 pan-genome strains, thank you to Bharathi Sundaresh and Suyen Espinoza for sequencing my samples, and thank you to Dr. Jon Anthony and Defne Surujon for their Tn-seq data processing pipelines. Most of all, thank you to my exceptional mentors Dr. Federico Rosconi and Dr. Tim van Opijnen for allowing me the opportunity to work on this project, and for their invaluable guidance and generosity with their time throughout this project.

1.0 INTRODUCTION

Since the designation of a new class of pathogen, the "opportunistic pathogen," in the late 1980s and early 90s, these shapeshifting microbes have commanded a large amount of the attention expended on infectious disease research. The emergence of this new, formidable classification coincided with the end of a grace period of about forty years of unbridled antibiotic usage, during which it was widely believed in the medical community that bacterial pathogens were largely treatable, and on the decline.[1] Despite the focus on bacterial opportunistic pathogens in recent decades, there still exist large gaps in our knowledge regarding their relation to the host in both commensal and pathogenic states. For example, it is known that colonization is possible due to the maintenance of an equilibrium between bacterial growth, production of virulence factors, and host immune defenses, but the specific interactions which underlie this state are poorly understood. Related to this are the unknowns regarding why human hosts, in interaction with the same bacterium, can display a wide range of responses, from little to no symptoms in some individuals to severe disease and death in others, and which particular combinations of virulence activities are utilized by potentially pathogenic vs. nonpathogenic colonizers.[1,2] Even among very well-studied pathogens, such as *Escherichia coli* and *Salmonella*, comprehensive characterization and linkage of genotype to phenotype is impeded by the fact that significant fractions of their genomes are comprised of unique genes about which functional predictions cannot be made.[1] Using human commensal and opportunistic pathogen *Streptococcus pneumoniae* as our model system, we aim to fill in some of these gaps. My project aims to parse, from the *in vivo* genetic requirements of both virulent and non-virulent *S. pneumoniae* strains, the genes involved in interaction with the nasopharyngeal epithelium (crucial for the commensal state), and with the first line of innate immune defense, macrophages (potentially influential in the host-pathogen interplay determining whether or not invasive infection occurs).

1.1 *Streptococcus pneumoniae* DESCRIPTION AND EPIDEMIOLOGY

Streptococcus pneumoniae is a gram-positive commensal inhabitant of the nasopharynx, with estimates of carriage prevalence ranging from 10 to 90% among healthy individuals.[3] While this carriage state is, for the most part, asymptomatic, occasionally accompanied at its most disruptive by a runny nose, colonization is immensely clinically important because it is a major reservoir for human transmission of *S. pneumoniae*, and it is thought to be a necessary first step in the initiation of invasive disease.[1] As an opportunistic pathogen, *S. pneumoniae* can take advantage of host immune defects imposed by chronic diseases, immunosuppression, ongoing or prior viral infections (for example with HIV or influenza), old age, or immune immaturity, and shift from the commensal to the pathogenic state if able to infiltrate host compartments outside of the nasopharynx).[2,1] In this case, *S. pneumoniae* can incite both local infection, i.e. respiratory pneumonia or otitis media in the ear, and systemic infections such as bacteremia and meningitis if able to infiltrate the blood or meninges.[1] Closely linked with this diverse range of host environments that *S. pneumoniae* can inhabit is the existence of at least 98 capsular serotypes, defined by their distinct capsular polysaccharide compositions and thus distinct antigenic properties.[1, 4] Certain serotypes have been found to be more likely to cause invasive disease than others.[4] *S. pneumoniae* crucially also possesses a very large pan-genome, with individual strains differing by as much as 10% of their genomic content; this variation has been intimately linked to the development of antibiotic resistance by different mechanisms by our group and others, as well as to myriad other elements of *S. pneumoniae* physiology. [1, 5]

It is estimated that almost 1.2 million deaths worldwide occur each year due to pneumococcal-derived lower respiratory infections, with the vast majority of those afflicted either under the age of 4 or over 70.[1] The Center for Disease Control and Prevention has quantified the more than 2 million infections that make up the yearly pneumococcal burden of

disease in the United States as costing the healthcare system $4 billion.[6] With the rise of antibiotic resistance among many strains, these numbers will only increase until a productive pipeline for new drug development can be implemented.

1.2 ANTIPNEUMOCOCCAL THERAPIES: THE RISE OF ANTIBIOTIC RESISTANCE

Vaccines formulated to recognize *S. pneumoniae* capsular serotypes have proven successful. PCV-7, introduced in 2000, was able to induce MHCII-mediated-T-cell-dependent responses even in infants by covalently linking carbohydrates to peptide antigens from seven serotypes which were responsible for over 83% of the invasive antibiotic-resistant infections in children at that time.[1, 6] PCV-13, introduced in 2010, added 6 additional serotypes, leading to a combined reduction in pneumococcal infections in children by more than 90% by 2019.[6] While this advancement is highly encouraging, one major problem is its highly limited scope; the new vaccine still excludes 77 known *S. pneumoniae* serotypes, and "capsule switching" is a natural phenomenon that is known to occur due to the natural competence of *S. pneumoniae*. Furthermore, although the valent vaccines were highly successful in the United States, the pneumococcal serotypes which plague many developing nations are a distinct set which only slightly overlap with that of the US, which has rendered these prophylactics much less effective around the world, especially when compounded with their high cost.[1] Within the United States and across the world, some infections still require antibiotic treatment.

The 2019 CDC Antimicrobial Resistance Threat Report indicates that each year in the United States alone, more than 2.8 million antibiotic-resistance infections arise, leading to the deaths of over 35,000 people. Almost one third of these resistant pathogens exhibit multi-drug resistance.[6] Antibiotic-resistant *S. pneumoniae* has been designated a "serious threat," despite the

successes of the two pneumococcal conjugate vaccines during the past two decades.[6] Acquisition of resistance mechanisms by pneumococcal strains can arise by horizontal evolution, (i.e. horizontal gene transfer), or vertical evolution, in which *de novo* mutations are responsible. Selective pressure by antibiotics is a near-constant process occurring within the human body and in livestock, when administered antibiotics reach the sites of *S. pneumoniae* infection at sublethal levels, allowing for genomic changes to accumulate which might improve the bacterium's ability to resist or overcome the drug's detrimental effect.[14] Thus, it is reasonable to assume that new clinically relevant strains will emerge which will necessitate the use of novel treatment and prevention strategies.

Importantly, our group and others have extensively characterized the response of a bacterium to antibiotic stressors as a genome-wide program. This not only includes the interaction of the drug and its direct target in the organism, but "propagates into various biochemical, metabolic, and regulatory processes of the cell," and thus is highly genetic-background dependent.[14, 5] In order to anticipate and prevent resistance, as well as design new therapeutic agents against bacterial pathogens without stimulating resistance development, it is highly important to understand the complex network of genetic interactions that underlie both bacterial stress responses and virulence mechanisms. This understanding must additionally consider the high genomic variability present between different strains of pathogens with large pan-genomes in order to be biologically relevant.

1.3 TN-SEQ AS A POWERFUL FUNCTIONAL GENOMICS INTERROGATION TOOL

One extremely versatile and comprehensive method of determining bacterial requirements for surviving in a particular condition or for successful interaction with eukaryotic

host cells at the genome-wide level is Tn-seq.[7] This technique relies on the construction of highly saturated transposon-insertion mutant libraries in strains of interest, and high-throughput sequencing for the relative frequency of mutants grown in a particular stress-inducing environment. Mutants that significantly decrease in frequency, or disappear, from the population after incubation in the stressful condition are suspected to have interrupted genes that are important for survival in this condition. Using this technique, our lab has conclusively identified a set of genes essential to growth in all studied strains (core-essentialome) and a set of genes only essential for growth in some strains (strain-dependent- essentialome). Similar approaches have been performed in other bacterial species, for example in *Pseudomonas aeruginosa*.[8] We have also used Tn-seq to determine the genetic requirements for bacterial interactions with particular host cells that are required for infection, or make up part of the virulome. Both genes that are essential for bacterial survival in all conditions, and genes that are required for successful infection of a host--especially those required by all strains of a species--represent attractive therapeutic targets.[9]

1.4 PARSING *S. pneumoniae* VIRULOME

Our lab has determined the set of genes required by *S. pneumoniae* for successful infection of the nasopharynx and lungs via *in vivo* Tn-seq in a mouse model, or the virulome, at the pan-genome level. Moreover, we determined these requirements in both low- and high- level virulent isolates, shedding light on the commensal versus infective behavior of *S. pneumoniae*. The goal of my project is to parse this set of genes into those required for attachment to the nasopharyngeal epithelia, a crucial step in establishing colonization of the host and stepping stone to invasive infection or transmission to new hosts, and those involved in subversion of host innate immune defenses, particularly by macrophages. By using Tn-seq and considering the *S.*

pneumoniae pan-genome, the identification of a core set of genes required by *S. pneumoniae* for such crucial components in the process of inciting a successful infection will yield extremely attractive targets for therapeutic agents and biologicals that reliably combat *S. pneumoniae* infection, no matter the genetic background of the infecting agent.

2.0 BACKGROUND

2.1 PAN-GENOME AND THE NETWORK STRUCTURE OF GENOMES

The pan-genome is defined as "the entire genetic repertoire of a given phylogenetic clade [which] encodes for all possible lifestyles carried out by its organisms."[10] As such, pan-genome analyses can focus on any taxonomic level or subgroup, such as the domain Bacteria, although they are more commonly done at the species level.[11] Studying a bacterial species at the pan-genome level can yield invaluable insights into the role of the environment as an architect of genetic diversity, highlighted in a recent study analyzing the pan-genome of *Bacillus cereus* and *Staphylococcus aureus*, using strains isolated from humans, soil, Earth-based built environments, and the International Space Station.[12] Recently, the pan-genome analysis has also been pinpointed as a potential new tool for the more precise classification, or even re-classification of bacterial species, which has historically relied at the genotypic level on somewhat arbitrary thresholds of percentage DNA-DNA hybridization or sequence similarity of the 16S rRNA gene, and which is made difficult at baseline by bacterial capabilities like horizontal gene transfer.[13] For our purposes, approaching the study of *S. pneumoniae* through a pan-genome lens allows us to appreciate the common and divergent genetic pathways which underlie phenotypes of this opportunistic pathogen to inform future treatment strategies.

The pan-genome can be divided into different categories. The core genome is composed of genes present in all strains, and typically is "responsible for the basic aspects of the biology of [the species] and its major phenotypic traits."[10] What we term the accessory genome, or genes present in some strains but not all, "might encode supplementary biochemical pathways and functions that are not essential for growth but which confer selective advantages, such as

adaptations to different niches, antibiotic resistance, or colonization of a new host."[10] The third group is unique genes, or genes that are present in only one strain.

It is particularly important to consider the genetic background, or genome, of a bacterium when interrogating any phenotype, because genetic responses to stimuli occur in a network-dependent manner. Genes are not "isolated units, but they interact with each other forming a network, [in which the] connections that shape this network can represent protein-protein interactions, epistatic relationships or transcriptional regulatory interactions."[14] An interaction between two genes in one species might not be present in another species that possesses those same genes, as revealed by an analysis of genetic interaction networks of *S. cerevisiae* and *S. pombe*.[15, 16] This indicates that networks are uniquely shaped by all of their gene members. Thus, it is not possible to assume that a gene will act in the same manner across strains, and awareness of the complete genetic background in a strain is crucial to understanding or predicting how that gene will act. As previously mentioned, our group and others have characterized the genome-wide response of a bacterium to antibiotic stress, revealing that "acute [stress] at the antibiotic's target" is one element of a systemic reaction involving diverse pathways and "[resulting] in selective pressures acting throughout the genome."[5, 14] For example, two daptomycin-susceptible strains of *S. pneumoniae*, TIGR4 and Taiwan-19F, share only half of the genes that make up each strain's program of response to the antibiotic, and the other half is specific to each strain. Interestingly, "the distribution of the functional categories of the responsive genes [is also] significantly different between the two strains," indicating a broader schism in response strategy, rather than simply granular differences in gene selection.[5, 14] Future study of any organism possessing a substantial pan-genome necessitates consideration of this highly consequential genetic diversity, including the determination of a feasible and representative manner in which to study species with high numbers of known genomic variants.

2.1.2 *S. pneumoniae* pan-genome

The average pneumococcal genome consists of 2100 genes, and the genome is approximately 2 megabases long. Globally, thousands of unique strains have been sequenced, and our group has identified a pan-genome of ~4300 gene clusters, with average strain-to-strain genomic content varying on average by 220 genes, or ~10% of an individual genome, as mentioned. This broad pan-genome provides *S. pneumoniae* with "an expanded set of genes to draw from for its own adaptation and evolutionary success," especially when different strains interact directly in environments conducive to exchange of genetic content, such as biofilms in the nasopharynx.[17] Horizontal gene transfer between cohabitating strains can basically generate new strains, whose novel genetic interaction networks may prove "beneficial to outcompete colonizers, evade host immunity, and escape human interventions, such as vaccines and antibiotics," and donor strains of genetic material include not only the diverse phylogeny of *S. pneumoniae*, but additionally co-residents of the nasopharynx *S. mitis*, *S. pseudopneumoniae*, *S. oralis*, and *S. infantis*.[17] Our group approaches the study of the large *S. pneumoniae* pan-genome through the usage of a carefully curated selection of 36 strains encompassing a majority of pan-genome diversity, allowing us to work with a representative model set at a manageable scale.

2.1.3 Pan-genome strain collection

A 2015 study by Dr. Aldert L. Zomer and team at Radboud University Medical Center in Nijmegen, Netherlands, presented a phylogenetic classification of 349 *S. pneumoniae* strains isolated from patients with bacteremic pneumococcal infection between the years 2001 and 2011.[18] From this phylogenetic tree, our group selected 15 pairs of closely-related strains, optimizing our selection such that these 15 pairs were scattered throughout the phylogeny to the greatest degree possible, in order to cover maximum genetic diversity. This set of 30, along with

6 well-studied strains (D39, TIGR4, Taiwan-19F, BHN97, CT22F, and GA22F), make up the complete pan-genome collection, with which we perform functional studies. The collection includes 16 capsule serotypes, both virulent and avirulent strains, and contains ~78% of pan-genomic content, making it highly representative of global genetic diversity, and thus an excellent model set to study *S. pneumoniae* biology from a pan-genome perspective. An in-house clustering method developed by our group on these pan-genome strains revealed the existence of ~4300 total clusters in the pan-genome, with ~1500 clusters representing the core genome. Given the average genome content of 2100 genes, this leaves ~600 genes per strain belonging to the accessory or unique genomes.

Although the genes that comprise the core genome are present in all isolates, they are not necessarily essential for *S. pneumoniae* growth. Our group determined this set of genes, termed the essentialome, using Tn-seq in rich medium on 22 pan-genome strains, representative of 73% of pan-genome diversity. Due to the low transformation efficiency of the remaining 14 strains we could not construct saturated mutant libraries in them, so they could not be included in the essentialome determination **(Table 1)**. By this approach we found that the essentialome consists of genes which are required for growth in any condition and independent of genetic background (core essentialome), as well as genes for which growth requirements were different among the strains (strain-dependent essentialome). We determined that the *S. pneumoniae* essentialome consists of 396 clusters, 214 of them belonging to the core essentialome and 188 clusters which were identified as essential in one or more strains but not all.

Given this identification from the *S. pneumoniae* pan-genome of core- and strain-dependent- requirements to satisfy the most basic physiological needs of a bacterium and our prior work revealing both genetically conserved and divergent responses to antibiotic stress, it is likely that other processes, such as specific interactions with the host, present the same division

of genes universally required or genes differentially required among isolates. Interactions that are highly interesting to probe are those involved with *S. pneumoniae* mechanisms of virulence; the set of genes involved in maintaining a successful infection also known as the virulome. Our group has also determined the *S. pneumoniae* virulome via *in vivo* Tn-seq infection experiments in a mouse model using the same 22 pan-genome strains, however, it is not yet known which specific biological function each of the genes fulfill inside the host. Genes required for successful infection could be involved in a variety of processes, including attachment to the nasopharyngeal epithelia, nutrient scavenging, evasion of phagocytosis and killing by host immune cells like macrophages and neutrophils, and paracellular invasion of new host compartments.[19] Parsing the virulome into functional groupings such as this will not only augment understanding of *S. pneumoniae* biology and comprehensively link genotype to phenotype, but additionally provide an extensive, contextualized set of attractive targets for future development of therapeutic agents. Comparing the requirements of virulent and nonvirulent strains may additionally yield insights into the gene requirements and pathways underpinning commensal vs. pathogen behavior.

Table 1. Strains and transposon library characteristics

Strain	Serotype	Organ of isolation	Country of isolation	Transformation efficiency	Total libraries saturation (%)	Included in final attachment Tn-seq analysis?
PG01	23F	Blood	Netherlands: Nijmegen	High	75	Yes
PG02	23F	Blood	Netherlands: Nijmegen	High	69	Yes
PG04	14	Blood	Netherlands: Nijmegen	High	80	Yes
PG06	33A/F	Blood	Netherlands: Nijmegen	High	62	No
PG09-Cm*	19F	Blood	Netherlands: Nijmegen	High	31	Yes
PG12	6B	Blood	Netherlands: Nijmegen	High	50	Yes
PG13	6C	Blood	Netherlands: Nijmegen	High	55	No
PG15	19A	Blood	Netherlands: Nijmegen	High	62	Yes
PG16	19A	Blood	Netherlands: Nijmegen	High	66	Yes
PG18	4	Blood	Netherlands: Nijmegen	Low	39	No
PG21	7F	Blood	Netherlands: Nijmegen	Medium	33	Yes
PG22	7F	Blood	Netherlands: Nijmegen	Medium	29	Yes
PG24	3	Blood	Netherlands: Nijmegen	Low	24	No
PG25	22A/F	Blood	Netherlands: Nijmegen	High	64	No
PG27	35B	Blood	Netherlands: Nijmegen	High	53	No
PG28	35B	Blood	Netherlands: Nijmegen	High	66	Yes
PG29	8	Blood	Netherlands: Nijmegen	Low	53	Yes
TIGR4-AC316	4	Blood	Norway: Konsvinger	High	63	Yes
D39-AC1770	2	Missing	Missing	High	60.7	No
Taiwan19F-14	19F	Cerebrospinal fluid	Taiwan	High	65	Yes
BHN97	19F	Nasopharynx	Sweden: Stockholm	High	76	No
CT-22F	22F	Blood	USA: Connecticut	High	63	No

*Transposon library in strain PG09 was constructed in a chloramphenicol resistance cassette (Cm) mutant of WT PG09 (inserted in an intergenic region); parental strain has low transformation efficiency. The last column refers to whether or not the strain was included in the ultimate analysis of *in vitro* D562 nasopharyngeal epithelial attachment Tn-seq results after quality control (QC) analyses were undertaken.

2.2 NASOPHARYNGEAL ATTACHMENT

S. pneumoniae can inhabit the human nasopharynx as a commensal, in which state this niche acts as a reservoir of the species, facilitating host-to-host transmission by shedding of the nasopharyngeal mucosa, or it can breach host boundaries at the nasopharynx and spread throughout the host, making carriage "the first step toward invasive disease."[19] The nasopharynx also acts as a reservoir of *S. pneumoniae* genetic material, as previously discussed, given that biofilms are highly conducive to gene exchange; the "colonization [state can select for and] drive evolution of virulence factors that cause disease."[19] In either state, pre-virulent or virulent, the ability of *S. pneumoniae* to successfully colonize the nasopharynx is a critical step in its survival, access to new hosts, and attainment of a foothold from which to potentially inflict invasive disease; reduction in commensal carriage by PCV administration to infants and toddlers has reduced pneumococcal disease in their communities by herd immunity, and if all benign *S. pneumoniae* colonization relationships were eradicated, there would be little to no potential for invasive disease.[20, 21] Mechanisms that facilitate *S. pneumoniae* survival in the nasopharynx, then, belong to the attractive set of potential drug targets that comprise virulence factors.

Establishing colonization of the nasopharynx requires *S. pneumoniae* to "attach to the epithelial lining, grow on the nutrient-limited mucosal surface, evade the host immune response, and transmit to a susceptible host."[2] Attachment is the first step in this process, and "span[s] three broad categories: transient association with mucus, weak association with host carbohydrates, and strong association with host surface proteins," with the third category clearly representing the most effective long-term strategy.[1, 2] *S. pneumoniae* possesses several well-known mechanisms for establishing this strong association, as well as several proteins known to function as adhesins; likely the pan-genome holds some novel proteins of interest that are additionally involved in host protein association.

2.2.1 Phases of attachment

Attachment to the nasopharyngeal epithelium is a dynamic process. The thick polysaccharide capsule is a critical virulence factor which allows *S. pneumoniae* to travel through the dense nasopharyngeal mucus layer to reach the epithelium and evade phagocytosis by immune cells, however it is a hindrance to protein-protein interactions between *S. pneumoniae* and host cells, because it shields membrane-associated proteins from the extracellular environment.[2] This usage of the net negative capsular charge to travel through the positively-charged mucus toward the epithelium and resist sloughing at the top layer represents the transient phase of adhesion.[2]

The neuraminidase NanA catalyzes the desialylation of host epithelial cells, inducing exposure of host glycoconjugate receptors, and both enabling *S. pneumoniae* binding at these new, unobstructed sites on host epithelia (for example by pneumococcal exoglycosidase BgaA to specific host carbohydrate residues post sialic acid-removal), and providing the bacteria with a source of nutrients.[2, 1] This set of interactions comprises the second phase of adhesion: weak association with host carbohydrates.

To allow for protein-based adhesion to occur, *S. pneumoniae* can undergo rapid phase variation, or "spontaneous, reversible phenotypic variation in...encapsulation, and pilus expression," expressing "fewer capsular polysaccharides and more teichoic acid [in order to] colonize the nasopharynx...more efficiently," or "more capsular polysaccharides and fewer teichoic acids" to increase virulence at the expense of attachment.[22] Cells in this phase also display more known surface adhesins than in the more heavily encapsulated phase, such as PspC, which binds secretory component on the polymeric immunoglobulin receptor, lipoprotein pneumococcal surface adhesin (PsA), a binding partner for epithelial cadherin, pneumococcal adherence and virulence factor A (PavA), which binds to fibronectin, and NADH oxidase (nox),

which binds extracellular matrix component laminin alpha-5.[2, 1] Pili-associated adhesins, such as RrgA, which is "required...in both the presence and absence of pili" are upregulated as well.[2] Pili are "long organelles able to extend beyond the polysaccharide capsule," and have been found to be encoded, in *S. pneumoniae*, in an operon "belonging to the microbial surface cell recognition adhesion matrix molecule...family of adhesins."[23] This constitutes the strong phase of adhesion.

My project aims to uncover novel *S. pneumoniae*-encoded proteins required for attachment at the pan-genome level, so this critical step in colonization and subsequent virulence can be comprehensively understood.

2.3 INNATE IMMUNE DEFENSE AGAINST *S. PNEUMONIAE*

The colonization state of opportunistic pathogens like *S. pneumoniae* is poorly understood, but plainly relies on a dynamic balance between bacterial growth and immune activation. Innate immune activation against *S. pneumoniae* infection has been well characterized to include macrophages initially, and secondarily, neutrophils.

2.3.1 Macrophages

S. pneumoniae colonizing a host will initially be met by resident phagocytic macrophages in the nasopharynx. The bacterium's main protective mechanism against phagocytosis is its polysaccharide capsule, however this is not always effective. When phagocytosed by macrophages, *S. pneumoniae* is subjected to a battery of stressors such as lysozyme, antimicrobial peptides, proteases, lipases, glycanases, metal stress, and oxidative stress by both reactive oxygen species, or ROS, and reactive nitrogen species, or RNS.[24, 25] The efficacy of microbicidal stress by ROS is hindered, however, by the fact that *S. pneumoniae* naturally possesses a variety of mechanisms to reduce oxidative stress, likely because oxidative stress is a

natural byproduct of its own metabolism, in the form of hydrogen peroxide. These include three regulated iron-importing ABC transporters, the manganese superoxide dismutase *sodA,* which removes superoxide anions, the NADH oxidase *nox*, which converts O_2 to H_2O (preventing formation of the superoxide anion), a manganese permease *psaA*, and a putative glutathione reductase *psaD*, the latter two which curb hydrogen peroxide production, and *spxB* which "prevents ATP depletion during oxidative stress" but is also a main producer of hydrogen peroxide during pneumococcal metabolism.[1, 25] Despite the fact that the varied mechanisms of oxidative stress resistance encoded in the pneumococcal genome likely act primarily to allow *S. pneumoniae* to "withstand the ROS generated by their own metabolism…[their existence likely makes the bacteria] relatively more resistant to the ROS, produced by macrophages...than bacteria that lack these adaptations."[25] *S. pneumoniae* possesses less robust mechanisms of resistance to RNS-induced stress, which can evolve from byproducts of ROS interacting with NO, and can be "potent anti-microbicidal molecules," but are countered by the *S. pneumoniae* pneumococcal surface protein C (*pspC*), which reduces NO production, S-nitrosoglutathione reductase *adhC*, and *clpP,* which has been generally implicated in RNS resistance.[25] If macrophages are overwhelmed by the invading bacteria and cannot establish the equilibrium characteristic of commensal status, the recruitment of neutrophils to the nasal cavity will occur within several hours to several days of introduction of *S. pneumoniae.*[1]

If *S. pneumoniae* migrates to the lungs, the first responders from the immune system will be alveolar macrophages, with neutrophils potentially recruited later. Alveolar macrophages possess the advantage of the scavenger receptor MARCO, or macrophage receptor with collagenous structure, which "binds to *S. pneumoniae* and promotes...opsonization."[1] Importantly, neutrophils are not always required to clear alveolar infection; if infectious dose is low, airway epithelial cells and alveolar macrophages in conjunction can clear the infection, via

release of protective inflammatory mediators Il-1, TNF-alpha, IL-18, complement proteins, surfactant protein-D, and AMPs.[1] If alveolar macrophages can handle the burden of immune response to *S. pneumoniae* infection, invasive infection can be prevented while avoiding the tissue damage that is often characteristic of robust neutrophil activation.[26] Avoidance of neutrophil recruitment also eliminates the risk of *S. pneumoniae* "escap[ing] to the bloodstream by slipping out of openings that allow neutrophil influx," which is a significant risk factor for the development of invasive disease such as bacteremia.[1]

2.3.2 Neutrophils

If interaction of *S. pneumoniae* with macrophages does progress to the stage of infection, for example due to high bacterial load that macrophages are unable to deal with, neutrophils (also phagocytes) are recruited. Neutrophils possess some of the same microbicidal mechanisms as macrophages and some unique ones; overall, neutrophils are much more successful in killing *S. pneumoniae* bacteria once phagocytosed. Within the cytoplasmic granule of the neutrophil, lysozyme is released to digest bacterial peptidoglycan cell walls, "generat[ing] fragments for Nod2-dependent signaling."[2] *S. pneumoniae* counters this by inhibiting opsonization, or the molecular marking of a bacterium for phagocytosis, by capsular blockage of cell wall antigens that innate immune cells could bind, and "deacetylating the N-acetylglucosamine residues in its peptidoglycan to protect [the cell wall] from host lysozyme cleavage."[1] *S. pneumoniae* is also subjected to stress within the neutrophil by "production of toxic oxygen radicals, proteolytic enzymes, myeloperoxidase, defensins, and other bactericidal peptides."[27] Neutrophils are also capable of killing bacteria that have not been phagocytosed by releasing their own genomic DNA along with a variety of antimicrobial granule proteins, including elastase, which will form a net-like structure known as the DNA-based neutrophil extracellular trap. Critically, neutrophil

activation is also the initiation of a much more robust and multifactorial immune response give that neutrophils additionally release granule components, cytokines, and chemokines, which will all result in the recruitment of other immune cells such as monocytes, dendritic cells, and T-cells.[27]

2.3.2 *S. pneumoniae in vivo* requirements

Our group, in collaboration with Dr. Jason Rosch of St. Jude's Children's Hospital in Tennessee, has determined the *S. pneumoniae* virulome in macrophage- and neutrophil- depleted mice by Tn-seq. This work revealed the requirement of a major facilitator superfamily, or MFS, transporter gene, SP_1587, by all pan-genome strains, to survive in the presence of macrophages. This gene has previously been identified by our group as required for infection of the lung in strain TIGR4. Another interesting gene identified was SP_1469, the *nox* enzyme (NADH oxidase protein), which has known functions in adherence to epithelial cells and resistance of oxidative stress, and which was identified in the immune-modified *in vivo* experiments as required for survival in the presence of macrophages and neutrophils in some strains.

Although the microbicidal strategy of neutrophils has been well characterized as more effective than macrophages against *S. pneumoniae*, *ex vivo* characterization of the specific host-pathogen interactions that underlie this is very difficult due to the short lifespan of neutrophils in cell culture and the lack of an established cell line. Macrophages, in contrast, have been domesticated for *ex vivo* culturing in many variants, including two that make them specifically attractive to study in the context of this project. The first is a cell line of alveolar macrophages, MH-S, which is clearly useful for characterization of the immune system-pathogen interaction that occurs first in local infection of the lung. The second is a non-alveolar line, J774A.1, in which activation of high production of both ROS and RNS has been demonstrated in previous

work.[28] Elucidation of the factors within the *S. pneumoniae* virulome required for survival of host macrophages will also represent ideal drug targets to promote restriction of the infection to a local, rather than systemic one, as well as to limit host injury by immune over activation. For these reasons, my project focuses on *S. pneumoniae* interactions with both alveolar and non-alveolar macrophages. My project aims to investigate the role of the MFS transporter and the *nox* gene in *S. pneumoniae* interaction with these macrophages, specifically with regards to oxidative stress resistance, in order to determine *S. pneumoniae* mechanisms for survival of the crucial first line of defense by the innate immune system.

3.0 HYPOTHESES

3.1 *S. pneumoniae* NASOPHARYNGEAL ATTACHMENT REQUIREMENTS

Within our pan-genome collection, strains exhibit variation in baseline attachment level to epithelial cells in tissue culture. One simple measure of this is the percentage of the inoculum that is recovered from monolayers after an incubation period; by this measure we have found that average percentage attachment to human nasopharyngeal cell line D562 varies within the 36-strain collection from below 1% to just above 30% (**Figure 1**). We hypothesize, given the demonstrated variability in genetic programs utilized to perform similar physiological functions between genomically distinct microorganisms, that this variation in attachment percentage between pan-genome strains is both due to their different genetic backgrounds and evidence of diversity in the set of genes used by different strains to attach.[5, 15, 16] From my experimental approach, characterization of the genetic requirements in 22 of the pan-genome strains to attach to D562 monolayers by Tn-seq, we expect to find both redundancy among strain requirements (core attachment genes), as well as unique genes required only by some strains (strain-dependent attachment genes).

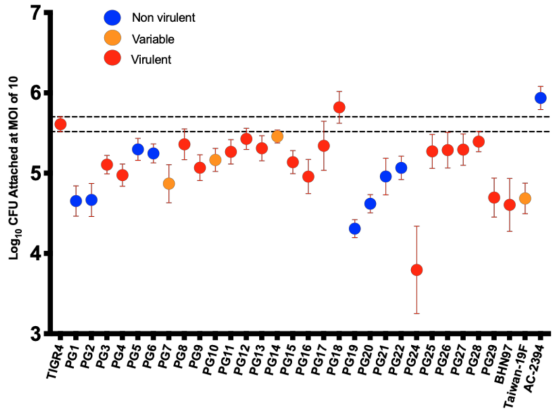

Figure 1. Attachment level (in CFU) to D562 epithelial cells varies between *S. pneumoniae* pan-genome strains. Data collected by Dr. Federico Rosconi and Lauren Havens.

3.2 *S. pneumoniae* INNATE IMMUNE SURVIVAL REQUIREMENTS

As described, *in vivo* experiments by our group challenging mouse cohorts at either full, macrophage-depleted, or neutrophil-depleted states of innate immune integrity with pan-genome strains in the lungs indicated a universal requirement of the MFS transporter gene, SP_1587, to survive macrophage attack, and a differential requirement of the *Nox* gene, SP_1469, to survive both macrophage and neutrophil attack (**Figure 2a-b)**. One potential explanation to account for differing requirements of *Nox* between strains could be that different genetic backgrounds stimulate recruitment of differing numbers of neutrophils, or that different animals possessed different numbers of resident macrophages inherently at the time of infection. Our group eliminated this as a likely possibility by quantifying the number of migrating neutrophils and macrophages by flow cytometry in mice challenged with two strains showing differing requirements for *Nox* (TIGR4 and PG12), in WT and knockout mutants of both genes, and found no significant difference between mice challenged with any of the six variants. We thus hypothesize that the MFS transporter is a universal component of the *S. pneumoniae* program of response to macrophage-induced stress. We also hypothesize that the *Nox* enzyme is required by some strains to aid in the resistance to both macrophage- and neutrophil-induced oxidative stress. Secondary to the latter hypothesis, we speculate that the addition of exogenous cysteine- and manganese-rich molecules, (both implicated in *S. pneumoniae* oxidative stress response), will aid in *S. pneumoniae* ability to survive oxidative stress by the macrophage.[25] To test these hypotheses, my experimental approach was to perform *in vitro* killing assays of PG12 and TIGR4 WT vs. their mutant derivatives in the MFS transporter and *Nox* genes, by the MH-S alveolar macrophage cell line as well as the J774A.1 non-alveolar macrophage line. The latter was to be tested with and without ROS/RNS activation.

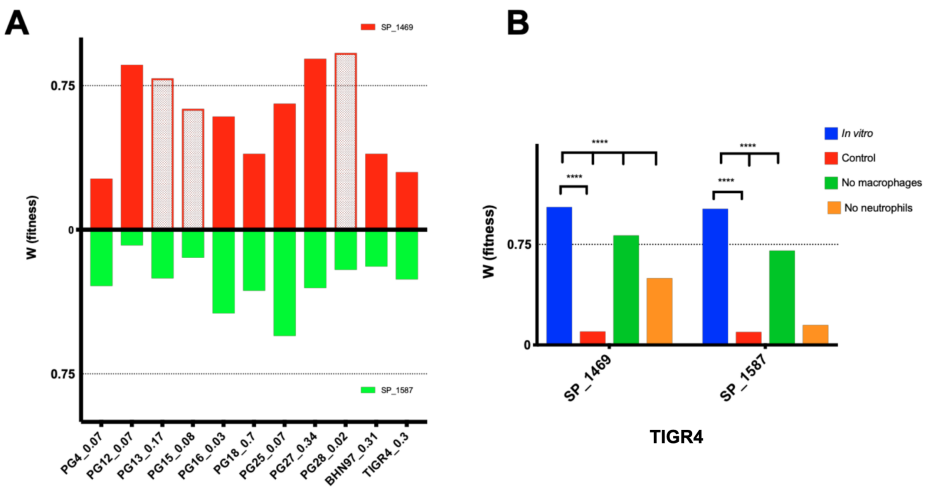

Figure 2. ***In vivo*** **Tn-seq experiments in the lungs reveals a strain-dependent requirement of SP_1469 and a universal requirement of SP_1587 by** ***S. pneumonaie*** **to survive (a), and** ***in vivo*** **immune-modified lung infection experiments by these two mutants in TIGR4 reveals the requirement of both genes to survive in the presence of macrophages (b).** A high dosage/volume ($1x10^8$ CFU in 40 µl) of each strain were intranasally inoculated in female Swiss Webster mice of 7 weeks old for each *in vivo* experiment. Lungs were collected at the humane end-point or after seven days. Control animals in the immune-depletion comparison experiments (**b)** retained full immune integrity, and *in vitro* data was obtained from growth in SDMM. These data were collected by Dr. Federico Rosconi, Dr. Jason Rosch, and Bharathi Sundaresh.

4.0 METHODS

4.1 NASOPHARYNGEAL CELL ATTACHMENT TN-SEQ

4.1.1 Transposon library construction

Transposon insertion libraries in 22 pan-genome strains were constructed as described in van Opijnen, Lazinski, and Camilli's 2014 *Current Protocols in Molecular Biology* paper, with a few modifications which are detailed below.[29] The first was an improved Mariner transposase purification protocol, which scales up the volume of the second culture of *E. coli* pMalC9 to 1000 mL from 80 mL, calls for induction with IPTG in the presence of ethanol at 24 Celsius rather than 37, using several cycles of sonoporation over ice to lyse cells, and alters some elution volumes and centrifugation steps in the purification steps. Our modified protocol also includes a purity analysis using SDS-PAGE electrophoresis of samples isolated throughout the protocol. These modifications serve to increase both the amount of active enzyme obtained from purification, and the consistency with which this result was obtained. The second modification was the usage of a maxi-prep protocol for genomic DNA isolation, via the Qiagen Genomic-tip 100/G kit, which enabled the isolation of high-quality and -quantity (up to 100 μg) DNA. The third modification was a volume scale-up in the transformation protocol, from 1 mL cultures to 4 mL (and proportional increases in all other reagents), enabling the attainment of more transformants. Finally, we were able to construct highly saturated transposon mutant libraries in some strains with naturally low transformation efficiency, such as PG9, by constructing the libraries in previously transformed mutants with a Chloramphenicol cassette inserted into a neutral intergenic region of the genome, allowing for the PG9-Cm mutant to transform more readily with transposed DNA.

4.1.2 Basic culturing of D562 nasopharyngeal epithelial cells

D562 human nasopharyngeal epithelial cell line was cultured in T150 tissue culture flasks, at 37℃with 5% CO_2. Nitrogen-frozen aliquots previously prepared by our group were used to begin growth, in complete Eagle's Minimum Essential Medium (+ 10% FBS, 1% Penicillin-Streptomycin). After growth to 70-80% confluence, cells were passaged by removing and discarding media, washing adherent cells twice with PBS without calcium and magnesium ions, incubating flasks for 10 minutes at 7℃with 5% CO_2 with 6 mL Trypsin EDTA, and then quenching and resuspending detached cells with fresh complete EMEM and seeding new flasks in a ¼ dilution (5 mL out of 20 total of this cell suspension, plus 20 mL fresh complete medium). Estimated live cell count in newly seeded flasks was performed by mixing 40 µL cell suspension with 10 µL Trypan Blue Stain, visualizing 10 µL of this on a hemocytometer under a microscope, and utilizing hemocytometer.org to calculate total number of seeded cells, after inputting the appropriate dilution parameters. T150 flasks with 25 mL EMEM seeded at a ¼ ratio typically grow to 70-80% confluence within 7 days, and medium renewal is required on both the third and fifth day after passaging.

Medium renewal was performed by first examining cell morphology and confluence of flasks under the microscope, and recording relevant observations. Using a seropipette, media was then removed from the flask, and all but 5 mL discarded, while the 5 mL was saved in a conical tube and centrifuged for 5 minutes at 1000 rpm to capture floating cells; supernatant was removed by aspiration. Pelleted cells were then resuspended in 25 mL fresh EMEM complete medium and returned to flask, after doubly washing adherent cells with 5 mL DPBS 1X containing ions. (All media was pre-warmed to 37℃in a water bath for ~1 hour prior to epithelial cell work).

4.2.3 Attachment assay protocol

48 hours prior to the start of the experiment, one 70-80% confluent flask of D562 cells was passaged, and cells were seeded in 1.2 mL aliquots, at ½ ratio, to 12-well plates, and set to grow at 37℃with 5% CO_2. On the day of the experiment, satisfactory wells showed a radial distribution of confluence in the epithelial monolayer, with 90-100% confluence in the center of the well, and lower degrees of confluence toward the edges of the well. 8 mL of THY with 5 µL/mL oxyrase was inoculated with 600 µL of pooled libraries in pan-genome strains (each strain was assayed in three separate experiments, with different library pool combinations matching new and old libraries to maximize saturation). These cultures were incubated for 1.5 hours, with a desired final OD_{600} of 0.3-0.5. Prior to inoculation of epithelial monolayers with bacteria, wells were washed twice with fresh plain EMEM medium, and bacteria were centrifuged for 7 minutes at 4000 rpm, washed with 8 mL DPBS 1X containing ions, resuspended in 600 µL DPBS 1X, and diluted into EMEM + FBS 10%. Wells were inoculated with an 800 µL dilution of EMEM + FBS 10% in triplicate for each strain, in which approximately 45 µL were composed of PBS suspension. 12-well plates were centrifuged for 5 minutes at 500 g to localize all mutants close to the epithelial monolayer, and then plates were incubated at 37℃with 5% CO_2 for 80 minutes. The remaining volume of EMEM + FBS 10% bacterial dilution was used in serial dilution, for the quantification of *S. pneumoniae* cells in the inoculum. The remainder of the PBS suspension (450 µL) was centrifuged down, the supernatant was discarded, and the pellet was frozen and later used to isolate inoculum, or Time Point 1 (T1) genomic DNA. After the 80-minute incubation, all media was collected from wells and replicates of the same strain were pooled. 10 µL was collected from each to perform serial dilutions and quantify *S. pneumoniae* cells in this population, and the remainder of the media was centrifuged down to plate the cells on TSA plates with 5% sheep's blood for overnight growth and

subsequent gDNA isolation. This population of cells was designated as Time 2 Non-Attached Bacteria (T2-NA). In the 12-well plates, containing the epithelial monolayer and any attached bacteria, Trypsin-EDTA incubation and quenching was used to detach both eukaryotic and prokaryotic cells. The same serial dilution and blood agar plating protocol was then undertaken for these Time 2 Attached Bacteria (T2-A), with the one modification that 15 µL of cell suspension was collected and pooled from one replicate of every strain, in order to perform eukaryotic cell counting using the hemocytometer as described above (**Figure 3**). After overnight growth of diluted bacteria on blood agar plates, CFU counting was performed to estimate the size of all bacterial populations.

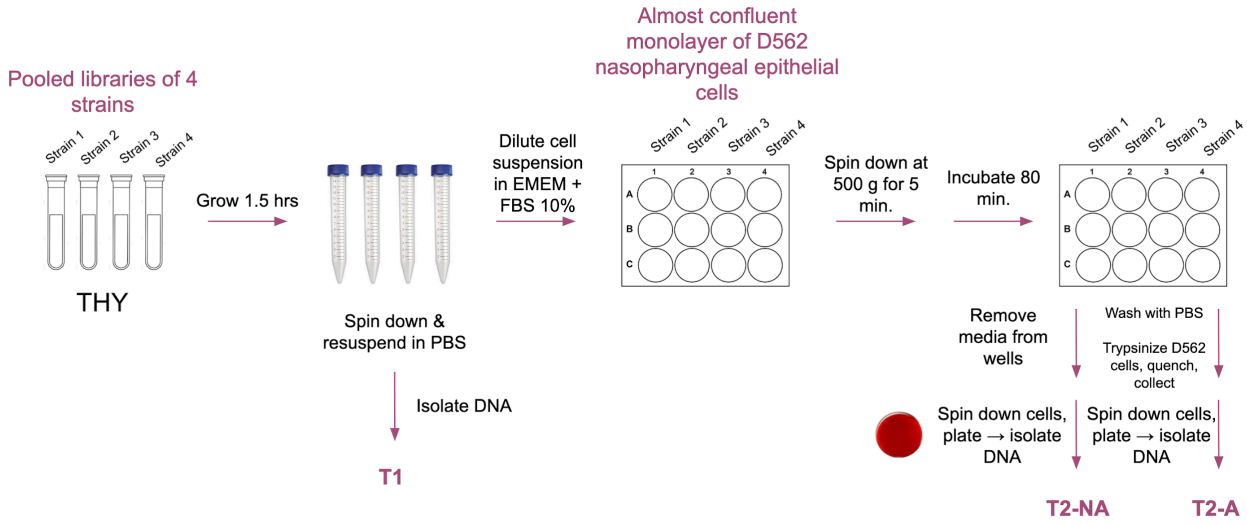

Figure 3. Schematic overview of D562 attachment Tn-seq assay. Serial dilutions at T1, T2-NA, and T2-A and eukaryotic cell counting as described above are not depicted.

4.2.4 Sample preparation for Illumina sequencing

Genomic DNA from T1, T2-NA, and T2-A described above was isolated using the DNeasy Blood & Tissue kit. Sample preparation was performed essentially as described in the 2014 *Current Protocols* paper, and constituted seven steps: MmeI digestion, dephosphorylation,

adapter ligation, PCR with labeled primers, electrophoresis of total PCR Product and band isolation from the gel, and finally aggregation of all samples per run into one mixture with equal concentration of each sample (in ng), Qubit measurement, and Tape-Station size checking. MmeI digestion occurred during a 2.5 hour incubation, enabling excision of the Himar minitransposon along with 16 base pairs of *S. pneumoniae* genome flanking its insertion on either side. Dephosphorylation was performed with phosphatase Quick CIP to ensure that excised linear fragments did not self-ligate. In the adapter ligation, all of the digested sample was incubated overnight at 16℃with 1.5 µL of adapter (1 of 48 unique barcode sequences; 1a-d through 12a-d), ATP, and T4 DNA ligase. The PCR reaction next amplified adapter-ligated fragments using a universal primer from the adapter end, and one of four specific labeled primers (BC01-04) from the other end, to complete the double, unique labeling of each sample (multiplexing) and allow for Illumina sequencing in parallel. The following steps enabled extraction of only the desired 180-bp fragments composed of the transposon, the flanking regions allowing localization of transposon insertion in the genome, and adapter and primer labels, and then preparation of the master sample (containing ~80 individual strain-pool-time point samples, or ~8-9 strains) for each of three runs. Illumina sequencing was performed by PhD candidates Bharathi Sundaresh and Suyen Espinoza, in our group.

4.2.5 Data analysis

Illumina sequence reads were de-multiplexed and mapped to their proper genomes using a proprietary data-cleaning pipeline called Aerobio developed by Dr. Jon S. Anthony, in our group. In order to achieve this, three .csv files were created, the first (SampleSheet.csv) detailing the BC01-04 primer sequences used, the second (Exp-SampleSheet.csv) detailing the strains used and samples sequenced, as well as their associated adapter barcodes, and the third

(ComparisonSheet.csv) indicating the time point represented by each sample and the expansion of each library pool to enable fitness calculation for each individual experiment. The output of the pipeline is the fitness values for all genes that had insertions in the transposon libraries, separated by strain. (Fitness was calculated as previously described in van Opijnen, Bodi, and Camilli 2009). The next part of the analysis was performed in Python and R in four discrete steps for the first sequencing run, and was subsequently consolidated into a pipeline by PhD candidate Defne Surujon in our group; the second two sequencing runs utilized this pipeline. First, an estimation of the likelihood of a bottleneck in each strain was calculated in Python. Next, in R, genes were filtered out from the total set which either: had previously been identified by our group as essential, had a total $n < 4$ in all three experiments for a strain, or which did not have fitness data, due to either not being present at T1, or not having any insertions in the libraries due to poor saturation or an absence of TA sites. Finally, to determine which genes were determined as significant in attachment, two methods were employed. The first was a comparison of all genes against the expected fitness value for each gene, for which we used the median fitness value of the total fitness distribution after filtering. The second was a comparison of the fitness in the Attached population versus the Non-Attached population. Genes flagged as significant between Attached and Non-Attached populations within a strain were then compared with *in vivo* nasopharynx infection data, and compared across all strains to identify sets of universally required and differentially required genes for attachment.

4.2 MACROPHAGE KILLING ASSAYS

4.2.1 Basic culturing of J774A.1 non-alveolar macrophages

J774A.1 murine macrophage cell line was cultured in T75 tissue culture flasks, at 37℃ with 5% CO_2. The original nitrogen stock was initially thawed in the incubator for several minutes, then resuspended in 10 mL Dulbecco's Modified Eagle Medium, and centrifuged to remove cryoprotectant DMSO. All media above the pellet was aspirated, and cells were resuspended in fresh complete DMEM (+ 10% FBS, 1% Penicillin-Streptomycin) to seed two flasks with 15 mL cell suspension in DMEM each. These initial flasks were grown to 70-80% confluence, and then passaged by cell scraping in 8 mL fresh complete DMEM after removal of media and doubly washing adherent cells with 5 mL PBS lacking calcium and magnesium ions; 1 mL was then resuspended in 14 mL fresh complete DMEM twice to seed two new flasks at a ⅛ dilution. For the first five passages, nitrogen stocks were also prepared of remaining scraped cells by centrifuging 5 mL of cells for 5 minutes at 1000 rpm and resuspending in 400 μL DMEM plain, 500 μL FBS, and 100 μL DMSO. Estimated live cell count was determined using the same method described above. Flasks typically grow to ~7-8 × 10^6 cells/flask at 70-80% confluence. When passaging at ⅛ dilution, flasks reach confluence approximately 5 days after seeding, with media renewal required once 3 days after seeding.

Medium renewal was performed in the same manner described above, with the modifications that appropriate volumes for the T75 flask were used, and DMEM media was used. (All media was pre-warmed to 37℃ in a "bath" of Lab Armour Beads for ~2 hours prior to macrophage work).

4.2.2 Killing assay protocol

This protocol was adapted from Drevets et. al 2015, in Basic Protocol 2 of their 2015 *Current Protocols* paper, "Measurement of Bacterial Ingestion and Killing by Macrophages".[30]

24 hours prior to the start of the experiment, media was changed in an almost 70-80% confluent flask of J774A.1 macrophages and replaced with DMEM + FBS 10% (lacking antibiotics), after triple washes with DPBS 1X containing ions. On the day of the experiment, 7.5 mL THY with 5 µL/mL oxyrase was inoculated with 2 starter cultures (THY with 15% glycerol) for each strain of interest. Starters were previously prepared according to the protocol in the 2014 van Opijnen, Lazinski, and Camilli *Current Protocols* paper. Bacterial cultures were grown for 1.5 hours, with a desired final OD_{600} of 0.3-0.5. OD_{600} of all cultures was standardized, and cultures were then centrifuged for 7 minutes at 4000 rpm in conical tubes, resuspended to wash in 7 mL DPBS 1X and centrifuged again under the same conditions, and finally resuspended in DMEM + FBS 10% at a dilution of 1×10^6 cells/10 µL. During the last 20 minutes of bacterial incubation, media was removed from a 70-80% confluent flask of macrophages, adherent cells were triply washed with PBS without calcium and magnesium ions, and adherent cells were resuspended in 10 mL fresh DMEM + FBS 10% by cell scraping. Total cell count was estimated using the hemocytometer as previously described. Resuspended cells were centrifuged for 10 minutes at 250 x *g* and 4℃, and resuspended in DMEM + FBS 10% at a dilution of 1×10^6 cells/100 µL. In Eppendorf tubes, 3 multiplicities of infection (MOI) for each tested strain were then set up (ideally in triplicate) with a final volume of 400 µL: 1:1 ratio of *S. pneumoniae* cells to macrophages (1×10^6 *S. pneumoniae* : 1×10^6 macrophages), 10:1 ratio of *S. pneumoniae* to macrophages, and 0.1:1 *S. pneumoniae* cells to macrophages. Cultures were mixed well by pipetting, and 10 µL of each condition was collected to quantify inocula by serial dilution in DPBS 1X. Eppendorf tubes were then incubated on a Labquake shaker rotating at ~8 rpm and, at

37℃, for 20 minutes. After this first incubation period, during which it is assumed that macrophages have been able to phagocytose bacteria but not kill them, tubes were centrifuged for 8 minutes at 250 x *g* and 4℃, supernatant was removed, and pelleted cells were gently resuspended in 400 µL. This wash was repeated twice more, with the final resuspension in 500 µL, to remove any extracellular bacteria which had not been phagocytosed. 10 µL of the final 500 µL was collected from each condition, and used for serial dilution in sterile water, in order to lyse macrophages and release phagocytosed bacteria for quantification of the total phagocytosed population. Eppendorf tubes were turned to 37℃ to incubate for 120 more minutes, after which time period macrophage lysis and quantification of remaining intracellular bacteria by serial dilution was performed in the same manner as before. After overnight growth of bacteria on TSA plates with 5% sheep's blood, CFU counts were performed to estimate the size of the bacterial population at each time point, and a ratio of intracellular bacteria remaining after the two hour incubation period to the initial population of phagocytosed bacteria was obtained for each strain and MOI (**Figure 4).**

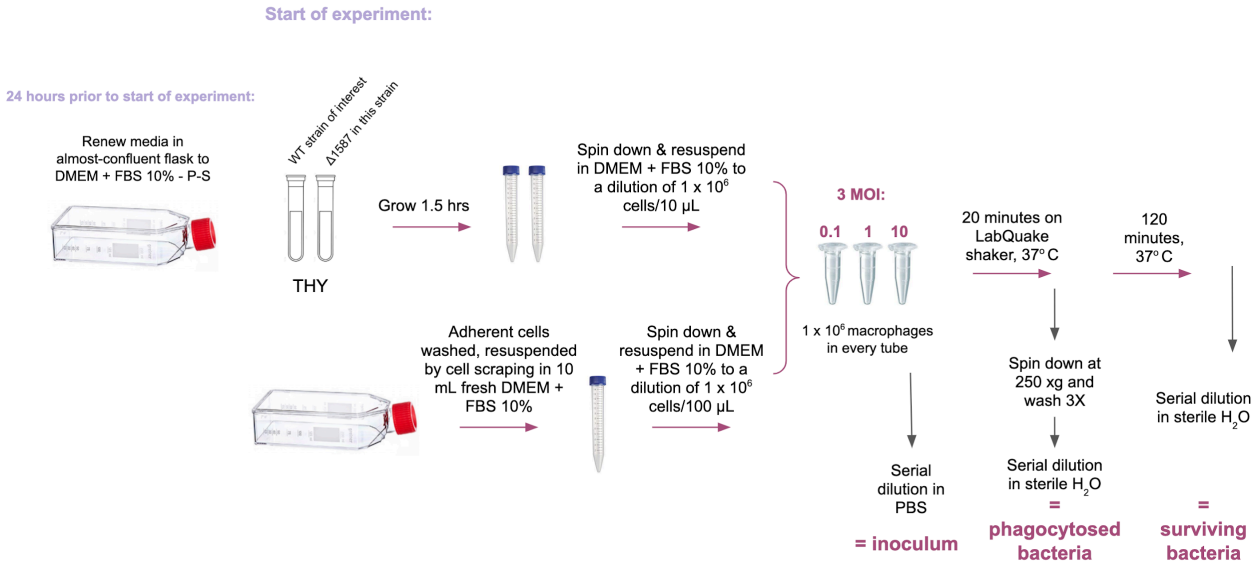

Figure 4. Schematic overview of J774A.1 macrophage killing assay. P-S refers to the Penicillin-Streptomycin antibiotic mixture typically added to cell culture growth medium to prevent contamination.

4.2.3 Oxidative stress activation in J774A.1 non-alveolar macrophages

Activation of J774A.1 macrophages to produce high quantities of reactive oxygen and nitrogen species (ROS and RNS, respectively), by stimulation with lipopolysaccharide (LPS) from gram-negative bacterial outer membranes and interferon gamma (IFN-γ) from murine immune cells has been described by Szliszka et al. Briefly, in their assay, macrophages were seeded at a density of 1×10^6cells/mL in 96-well plates, and grown in the presence of LPS (200 ng/mL) and IFN-γ (25 U/mL) for 24 hours, after which nitric oxide (NO) production was measured by spectrophotometric absorbance of 96-well plate supernatants mixed with the Griess reagent.[28] The Griess reagent is a mixture of naphthylethylenediamine dihydrochloride suspended in water and sulphanilamide in phosphoric acid, which can react with nitrite to form a purple azo product which exhibits absorbance at 546 nm.[31] For induction of ROS, phorbol 12-myristate 13-acetate (PMA) has been used in a CellROX oxidative stress assay, in which one measure of oxidative stress can be the oxidation of chemiluminescent cyclic diacylhydrazide luminol.[32] Our existing killing assay will be modified to include activation of J774A.1 macrophages by adapting these methods, along with addition of cysteine- and manganese-rich molecules.

5.0 RESULTS

5.1 DATA PROCESSING

5.1.1 Attachment Tn-seq results from the 22 pan-genome strains assayed were analyzed by three methods to identify untrustworthy data

After filtering from Illumina sequencing results essential genes, genes with insufficient frequency in either T2 population to carry significance in our estimation, and genes which were not present at T1 as described, the total number of genes for which confident fitness predictions could be made for each strain was quantified. Given that the average *S. pneumoniae* genome is 2,100 genes, roughly ~300 essential genes in each strain are removed, and it is probable that some genes will be excluded due to library saturation or lack of TA sites, we consider a successful experiment in a strain to include data from ~1,500 genes. This threshold was reached by attachment Tn-seq experiments in TIGR4 and T19F, performed previously by Dr. Federico Rosconi, and by the majority of strains included in the first sequencing run, known as "Run01," while subsequent sequencing runs Run02 and Run03 exhibited a great degree of variation (**Fig. 5a)**.

While a lack of confident data in many genes in the genome is indicative of technical errors in the attachment assay or the sequencing sample preparation by imperfect reagents or imperfect execution, alone this statistic cannot conclusively relay whether or not the data is trustworthy. Another, more informative measure of the efficacy of the experiment is to plot the distribution of fitness values for all genes remaining after the filtering steps in both T2 populations (Attached and Non-Attached) for every strain. The expected distribution is characterized by the most frequent fitness value among genes being W =1, given that most genes

in the genome are not involved in either attachment to nasopharyngeal epithelium or in special requirements (i.e. apart from the essential genes) for growth in EMEM, with a smaller but substantial group of genes exhibiting $0 < W < 1$, and another small group exhibiting $1 < W < 2$. The range is thus expected to be roughly $W = 0$ to $W = 2$, with the median value being $W = 1$. Previous Tn-seq experiments by the lab, in conditions such as SDMM (a compositionally similar medium to EMEM), have established this expected distribution. Initially plotted, these distributions in Runs 02 and 03 (and, more mildly, in Run01) showed unexpected diversity in their median fitness values, with many populations exhibiting low median fitness values. In order to standardize and enable comparison of gene fitness between strains and between populations of the same strain, all fitness values were normalized to the median value of their distribution, correcting the median values to roughly $W = 1$ (**Fig. 5b)**.

However, post-normalization, many distributions in Run02 and Run03 strains still exhibited unexpected characteristics, displaying one or both of two main problems: the first, a range of fitness values which is too wide (extending far beyond $0 < W < 2$), and the second, frequency of fitness values which are too great for $W \neq 1$. To address the first problem, two sets of fitness cutoffs were determined using qualities of the ranges of TIGR4 and Run01 distributions, one more stringent relative to the other, and one more lenient. The more stringent cutoff, represented in pink in **Figure 5b**, excludes any strains which possess more than 7 data points outside of the range [-0.662, 3.483]. The upper limit in this set is one standard deviation above the average maximum value of the TIGR4 and Run01 distributions, and the lower limit is one standard deviation below the average minimum value of these model distributions. The more lenient cutoff, represented in lilac in the same figure, excludes strains which possess more than 7 data points outside of the range [-1.028, 4.235], which denotes the absolute minimum and maximum fitness values from the TIGR4 and Run01 set.

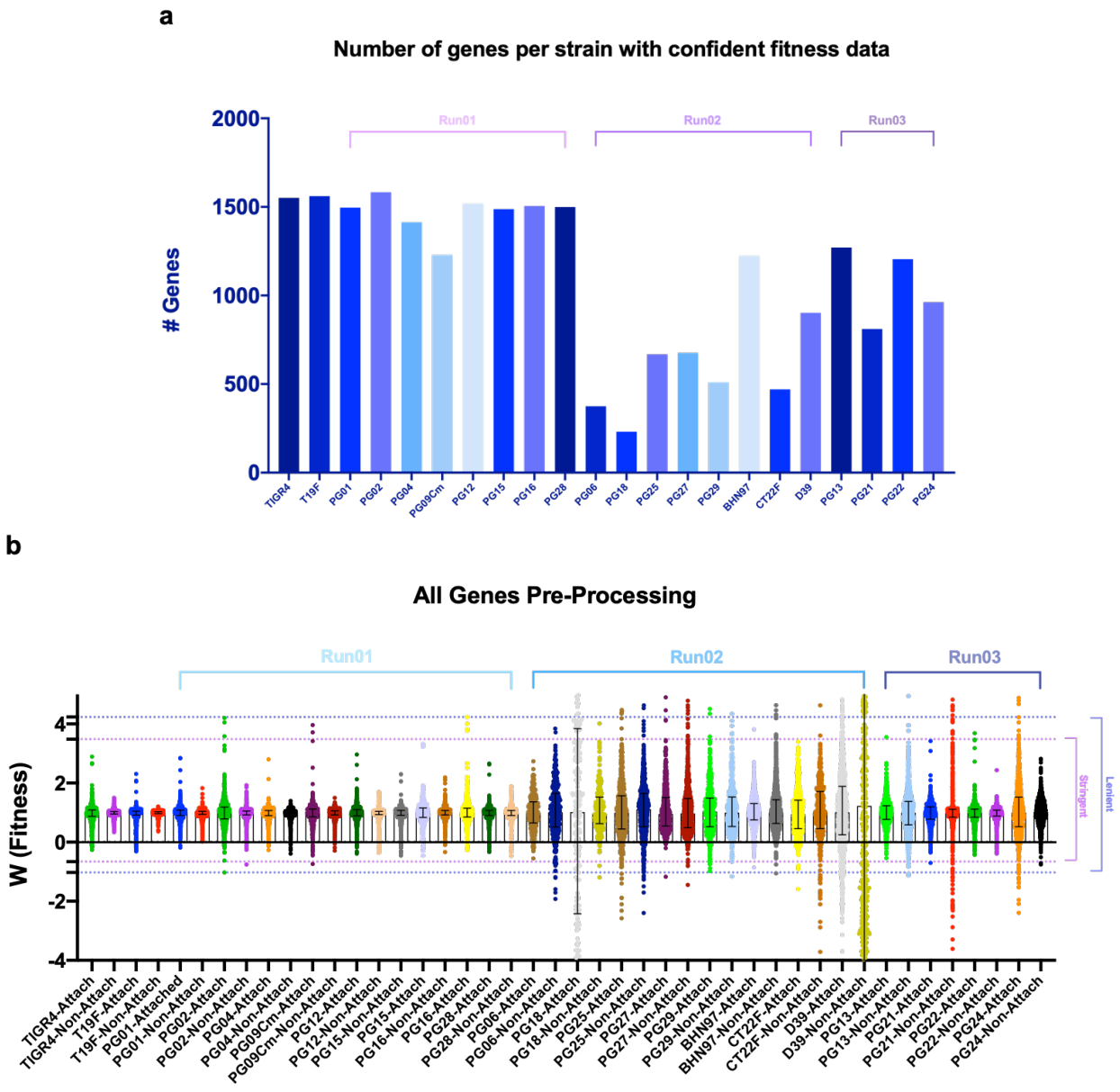

Figure 5. The number of genes per strain for which confident fitness data was obtained (a), as well as the normalized distribution of fitness values for all genes with confident fitness data within each strain in both Attached and Non-Attached populations (b), indicate that experiments and/or sequencing preparation for the strains comprising sequencing Run01 were more technically successful than Run02 or Run03. The number of genes with confident fitness data was obtained after filtering out essential genes and genes for which n < 4 for that mutant in Attached or Non-Attached populations in the three experiments performed for each strain (**a**). Fitness values shown above are all normalized to the median fitness value of the raw set for each strain in their Attached and Non-Attached populations, respectively. Pre-normalization, median fitness values for strain-population combinations ranged from -0.2645 (D39-Non-Attached) to 1.1652 (PG13-Non-Attach), with an average of 0.7355 ± SD 0.3255. Post-normalization (shown above), median fitness values range from 0.9380 (CT22F-Attached) to 1.230 (D39-Non-Attached), with an average of 1.0063 ± SD 0.0423. Cutoffs to remove strains with problematic range distributions included a more stringent option (removes any strains which have greater than 7 data points above W = 3.483 or below W = -0.662; one standard deviation above the average maximum value from T4/Run01, and one standard deviation below the average minimum, respectively), and a more lenient option (removes any strains which have greater than 7 data points above W = 4.235 or below W = -1.028; the absolute maximum value from T4/Run01, and the absolute minimum value from T4/Run01, respectively). 738 data points currently outside the axis limits (**b**). TIGR4 and T19F data were previously collected by Dr. Federico Rosconi.

Using either cutoff, 7 strains were excluded (PG18, PG21, PG24, PG25, PG27, CT22F, D39), and from the lenient to the stringent cutoff, 3 more were excluded (PG06, PG13, BHN97). In order to validate the removal of these 7 strains and decide whether the lenient or stringent cutoff was more biologically relevant, four "control" genes were chosen in which to plot fitness values for a strain-by-strain comparison, with the average values of the TIGR4 and Run01 set used as the marker for expected data **(Fig. 6a-d)**. Three of the genes (SP0346, SP0349, and SP0350 in old TIGR4 locus tags) were designated as "controls" because they are all capsule genes, about which it is known that inactivation leads to higher attachment levels; thus W_Attached > W_Non-Attached reliably in these mutants. Our group has robustly demonstrated this phenotype in attachment assays investigating unencapsulated derivative of TIGR4, strain 2394, and it has additionally been shown by many others **(Fig. 1)**.[22, 35] The fourth gene (SP1650) was chosen due to its universal requirement by *S. pneumoniae* to grow in medium lacking manganese (which EMEM does); thus W_Attached and W_Non-Attached were both expected to be low. In all four genes, fitness values in strains which lacked the T2 mutant frequency of 4 that we require for statistical confidence clustered very dissimilarly to T4/Run01, validating the use of this threshold. Based on the relative similarity of strains PG22, PG29, and PG21 to T4/Run01 averages, and the dissimilarity (as well as strong indication of skewed values, such as W = 0 uniquely in one or two strains, or W = 40) of strains PG06, PG13, BHN97, PG18, PG24, PG25, PG27, CT22F, and D39, to T4/Run01 averages, a final set of 13 strains was chosen **(Fig. 7; Table 1)**.

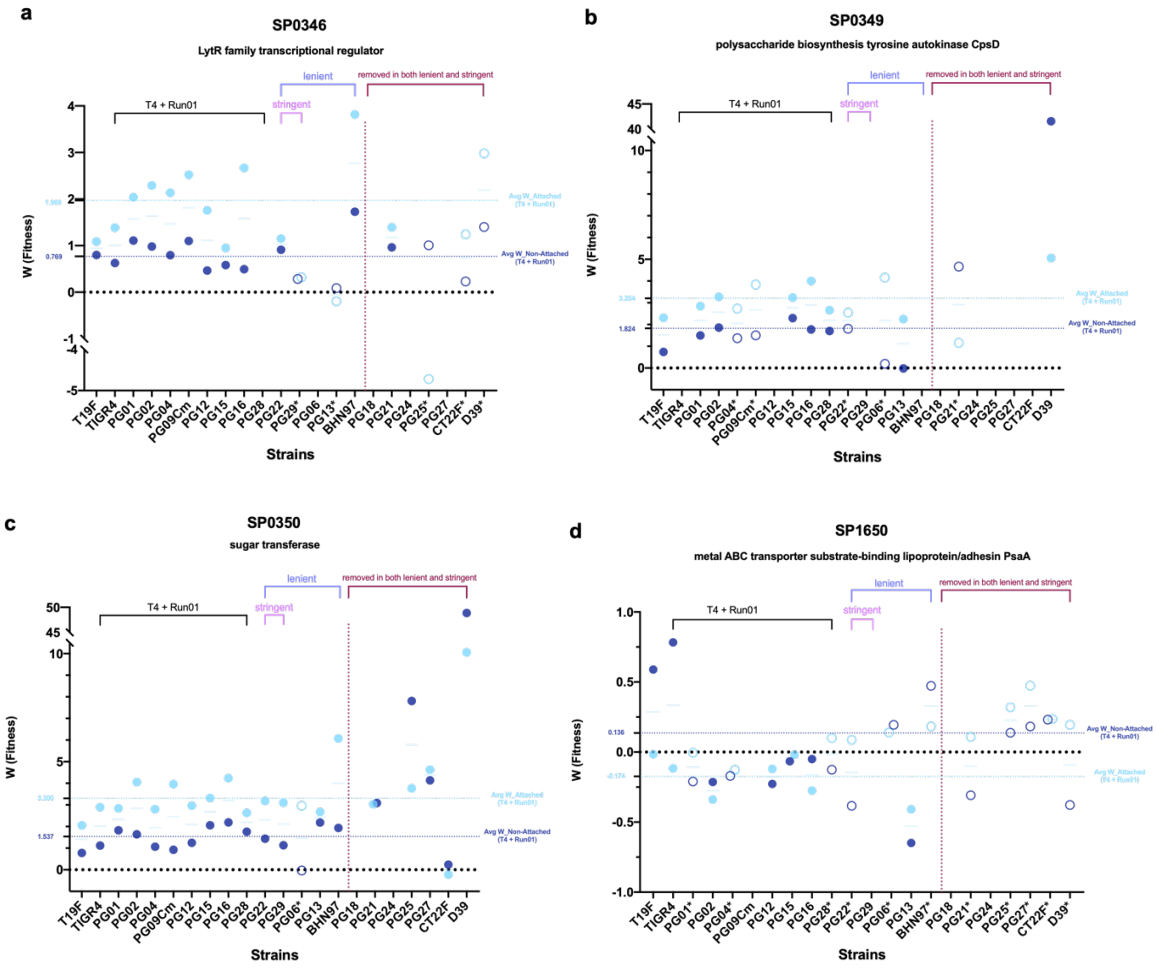

Figure 6. Strain-by-strain fitness comparisons in four "control" genes validate the exclusion of six of the seven removed strains from attachment analyses with the exception of PG21, and indicate that the stringent cutoff should be used. SP0346, SP0349, and SP0350 (**a-c**) are genes within the capsule operon of *S. pneumoniae* strain TIGR4, with the expected phenotype of higher W_Attached than W_NonAttached, and SP1650 (**d**) is a gene required for growth in EMEM, with the expected phenotype of low W_Attach and W_NonAttached. Fitness values designated by an open circle represent mutants which showed frequency < 4 in T2 populations.

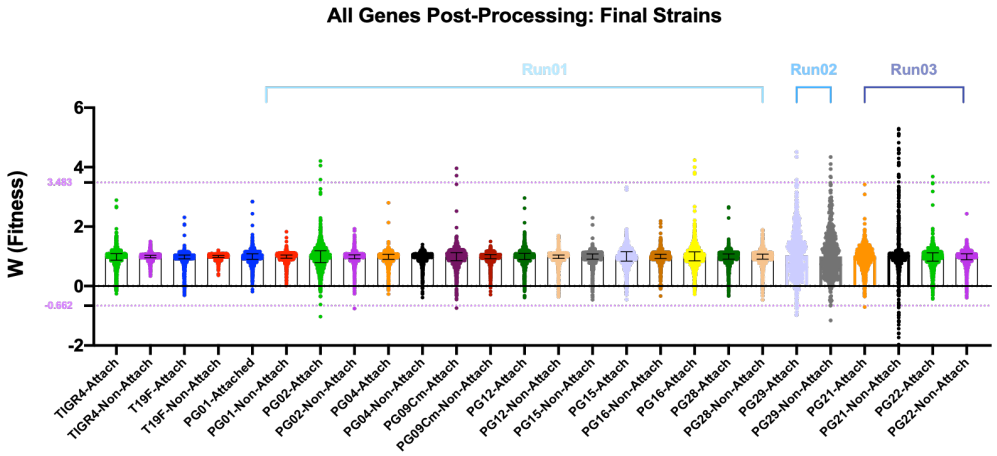

Figure 7. Final strains included for attachment analysis include 3 strains that did not belong to Run01 or previous experiments by Dr. Federico Rosconi. Genes for which fitness values fall outside of the stringent fitness cutoff were not considered.

5.2 REDUNDANT AND DIVERSE GENE REQUIREMENTS FOR NASOPHARYNGEAL ATTACHMENT WERE IDENTIFIED BETWEEN STRAINS

5.2.1 Genes which showed a statistically- and biologically- significant difference between Attached and Non-Attached fitness were largely Attachment-Implicated

To identify genes of interest in any of the 13 strains considered, a multiple T test with correction for multiple comparisons by false discovery rate 5% was performed in GraphPad Prism, comparing W_Attached and W_Non-Attached for each gene. Genes with a significant difference between the two populations in a given strain were designated a "discovery." When mining the full dataset containing 22 strains for "discoveries," 628 were found, whereas when considering only the 13 strains vetted for similarity to T4/Run01 in control genes, 617 were found. This further validated that the 9 strains excluded during data processing should have been excluded, given that only 11 discoveries, or 1.75% of the total number, were lost. Within this set of 617, what we consider to exhibit "biological significance," or a difference in W_Attached and W_Non-Attached by at least 0.2, was identified in 509 genes.

The 509 genes were then split into three groups: the first separated out any genes for which W_Non-Attached < 0.6, indicating that the gene is likely required for growth in EMEM. The relative fitness of W_Attached to W_Non-Attached in these mutants is unlikely to yield biologically relevant results, because these mutants exhibit a significant defect in the experimental growth medium. (Although a biologically significant growth defect would technically be $W \leq 0.8$, the cutoff of 0.6 was chosen because two capsule genes appeared in the Attachment-Hindering group with W_Non-Attached in the range $0.6 < W < 0.7$). With the remaining genes, those for which W_Attached > W_Non-Attached, denoting genes which hinder attachment when present (such as the capsule genes), and genes for which W_Attached <

W_Non-Attached, denoting genes which are implicated in attachment to nasopharyngeal epithelial cells, were further separated. Of the 509 total biologically significant discoveries, 74 belonged to the potential EMEM-required group, with 59 of these unique genes, 129 belonged to the Attachment-Hindering group, with 94 of these unique, and 306 belonged to the Attachment-Implicated group, harboring 234 unique genes.

A preliminary look at the functional categories represented by these genes was obtained by obtaining COG ID's and corresponding top-level COG categories for each gene with a TIGR4 homolog (old TIGR4 locus tags were used to find functional predictions for the genes from the Integrated Microbial Genes and Genomes database (IMG)) **(Figure 8)**. This analysis revealed a relatively diverse set of functions represented in the Attachment-Implicated genes **(a)**, while Attachment-Hindering Genes showed outsize representation for cell wall/membrane/envelope biogenesis proteins (the category to which capsule genes belong) **(b)**, and Potential EMEM-Required genes showed almost ¼ of genes belonging to amino acid transport and metabolism **(c)**, indicative of the need for all bacteria to extract nutrients from the environment.

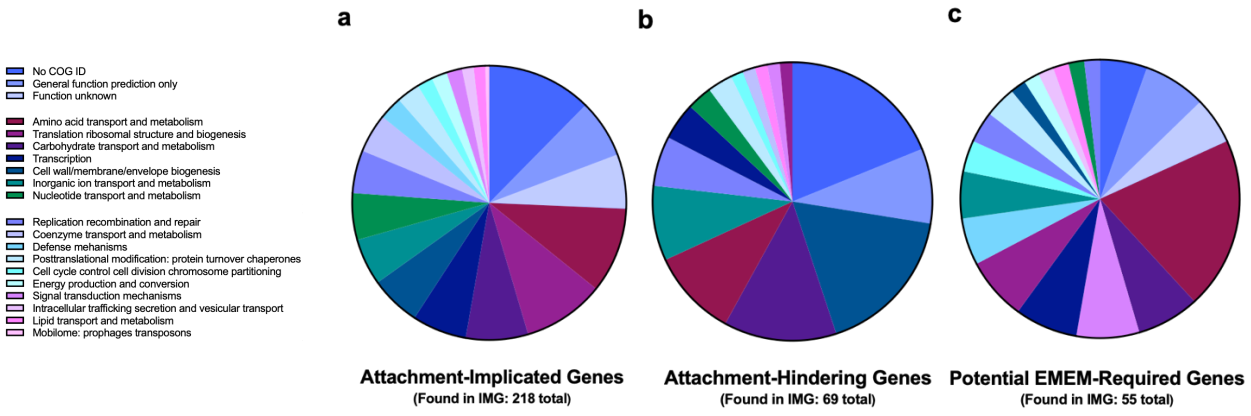

Figure 8. A functional analysis of the proportion of TIGR4-homologous genes within three experimentally relevant groups belonging to the different top-level COG categories revealed a relatively diverse set of functions represented in the Attachment-Implicated group, and a likely enrichment for cell/wall/membrane/envelope biogenesis in the Attachment-Hindering group. Old TIGR4 locus tags were used to find genes in the Integrated Microbial Genes and Genomes database (IMG), and COG IDs were used to identify top-level COG category affiliation. In total, the Attachment-Implicated (W_A < W_NA) group contained 234 unique genes (218 of which possessed a TIGR4 homolog) (**a**), the Attachment-Hindering group (W_A > W_NA) contained 94 unique genes (69 of which possessed a TIGR4 homolog) (**b**), and the Potential EMEM-Required group (W_NA < 0.6) contained 59 unique genes (55 of which possessed a TIGR4 homolog) (**c**). This analysis was performed not adjusting for the proportion of the TIGR4 genome made up of these functional categories, i.e. it was not an enrichment analysis.

5.2.2 Parsing the virulome

While analysis of all genes which exhibited statistically and biologically significant phenotypes in the attachment assay is useful in terms of understanding the assay itself, to accomplish our stated goal of parsing the virulome into functional groupings, the genes that are most important to study are those which exhibit a corresponding fitness defect between the *in vitro* attached population and the *in vivo* nasopharynx infection model. To assess the extent of correlation between the *in vitro* nasopharyngeal attachment data collected in this project with previously collected *in vivo* data by Dr. Federico Rosconi, the average W_Attached from all genes covered by the 13 strains considered in the *in vitro* analysis was plotted against the average W_Nasopharynx of these genes **(Figure 9)**.

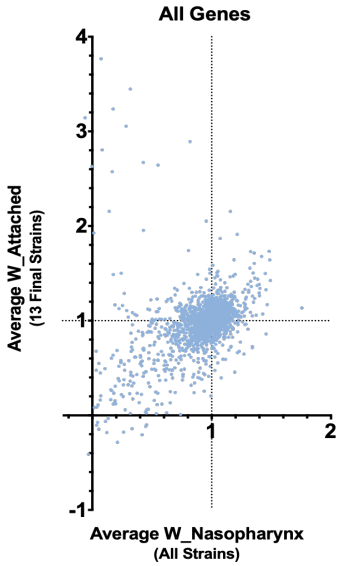

Figure 9. Average fitness for all genes represented in the final set of 13 strains, in the population of bacteria attached to D562 nasopharyngeal epithelial cells *in vitro*, compared with the average fitness of these genes from all pan-genome strains assayed *in vivo* in the nasopharynx by Dr. Federico Rosconi exhibit expected clustering of fitness values for most genes at 1. The genes most promising to consider in order to parse from the virulome factors involved in nasopharyngeal attachment cluster in the bottom left corner of the graph, where W < 0.75 in the nasopharynx and W < 0.8 *in vitro*.

As expected, most genes cluster at W = 1 for both the *in vitro* and *in vivo* conditions. The genes which exhibit low W_Nasopharynx but high W_Attached likely belong to the capsule, because while an interrupted capsule *in vitro* enables higher levels of attachment as previously

discussed, *in vivo* the loss of the capsule confers a significant disadvantage to *S. pneumoniae* in interaction with the immune system.[36] The genes which are most promising in the context of parsing nasopharyngeal attachment requirements from the virulome are the genes which cluster where W_Nasopharynx < 0.75 and W_Attached < 0.8.

5.2.3 Core attachment-implicated genes were identified among the 13 considered strains

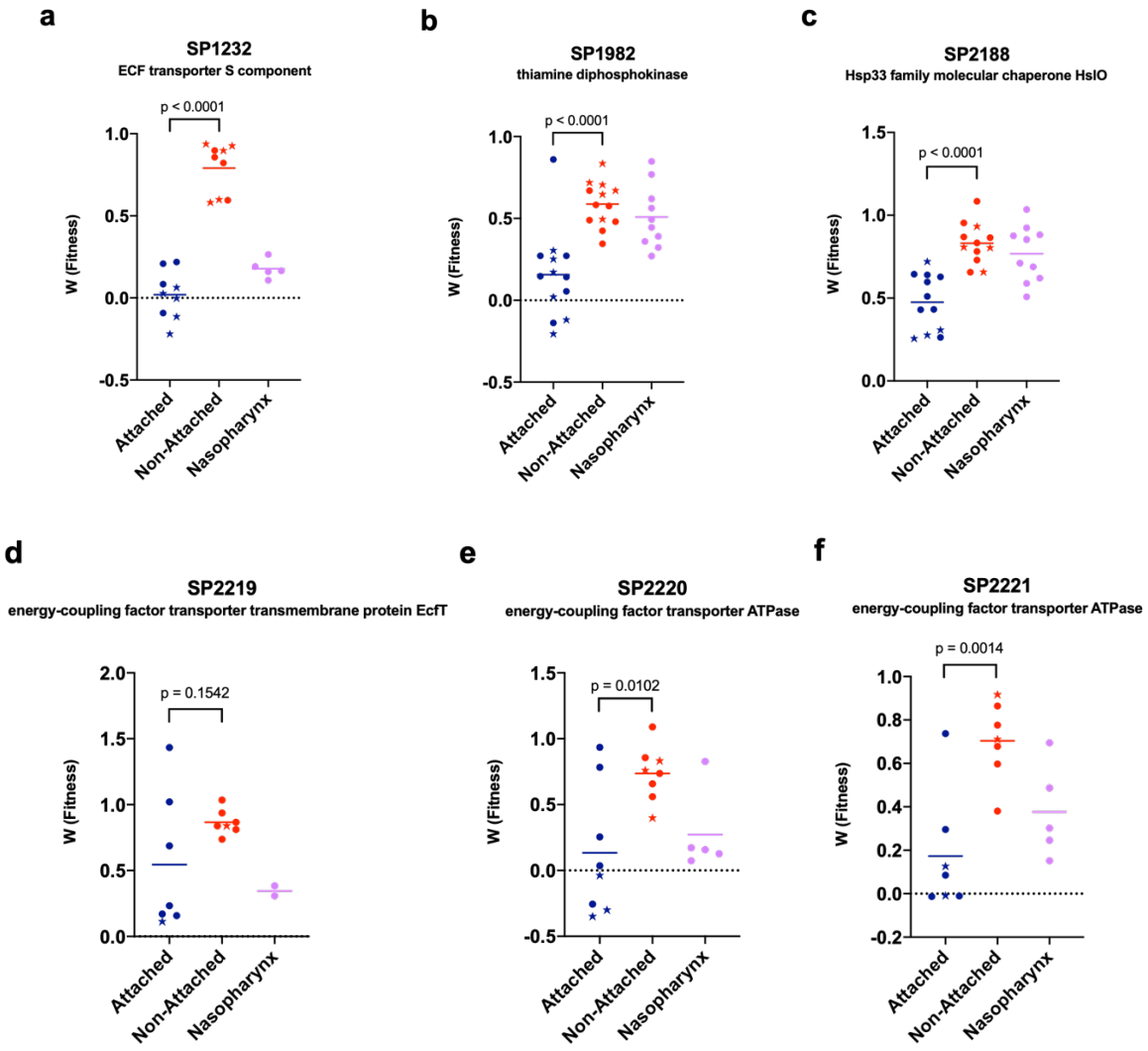

Figure 10. Genes for which W_Attached correlates with W_Nasopharynx, and for which the range of W_Attached is ~0.5 and/or W_Nasopharynx values are all ≤ 0.75 are considered to be potential core attachment-implicated genes. SP1232 and SP1982 are involved in vitamin transport or metabolism (**a-b**), SP2188 is involved in oxidative stress response **(c)**, and SP2219, SP2220, and SP2221 are part of an energy-coupling factor (ECF) transporter that likely transports cobalt (**d-f**). Data points represented by a star were flagged as "Discoveries" by our multiple T-test analysis with correction for multiple comparisons by false discovery rate 5%, indicating a statistically and biologically significant difference between W_Attached and W_Non-Attached for these genes in marked strains. Statistical difference between W_Attached and W_Non-Attached groups for each gene as reported above was calculated via Welch's t test in GraphPad Prism.

Among the Attachment-Implicated genes which correlated with *in vivo* nasopharynx data, one subgroup is genes which exhibit relatively the same fitness defect across all 13 strains in W_Attached and in W_Nasopharynx. If all but one or two outlying W_Nasopharynx data points cluster below 0.75, and/or the range of W_Attached values is ~0.5 for a gene, it is considered to be part of the potential core attachment-implicated genes. Six genes that were investigated in this project exhibited a core attachment phenotype: two genes encoding proteins involved in vitamin transport or metabolism (SP1232, the S component of an ECF transporter for pantothenate, and SP1982, thiamine diphosphokinase), one gene encoding a chaperone protein involved in the oxidative stress response (SP2188), and three genes which encode another ECF transporter, likely for cobalt (SP2219, SP2220, SP2221) **(Figure 10a-f)**.

The other subgroup within the Attachment-Implicated set consists of genes which exhibit a strain-dependency in attachment requirement. These genes exhibit a wide range in W_Nasopharynx including some for which $0.75 \leq W$, and/or a range of ~1 in the W_Attached values among the 13 strains considered *in vitro*. Five genes that were initially investigated in this project demonstrated a strain dependent attachment phenotype: one gene which has previously been identified as attachment-implicated (SP1274), two genes which encode enzymes involved in the reduction of oxidative stress (SP1469, NADH oxidase, a demonstrated adhesin, and SP0766, superoxide dismutase), one gene encoding part of an ABC transporter permease for zinc (SP2170), and one gene which catalyzes the formation of aspartate from phosphoenolpyruvate (SP1068, phosphoenolpyruvate carboxylase) **(Figure 11a, c, e, g, i)**. For strain-dependent attachment genes, we also investigated the extent to which fitness data from the *in vitro* assay undertaken in this project correlated with *in vivo* nasopharynx data on a strain-by-strain basis. The general positive trend shown in **b**, **d**, **f, h**, and **j** in **Figure 11** indicate the reproduction of strain-specific phenotypes from the *in vivo* experiment in the *in vitro* assay.

5.2.4 Strain-dependent attachment-implicated genes were identified among the 13 considered strains

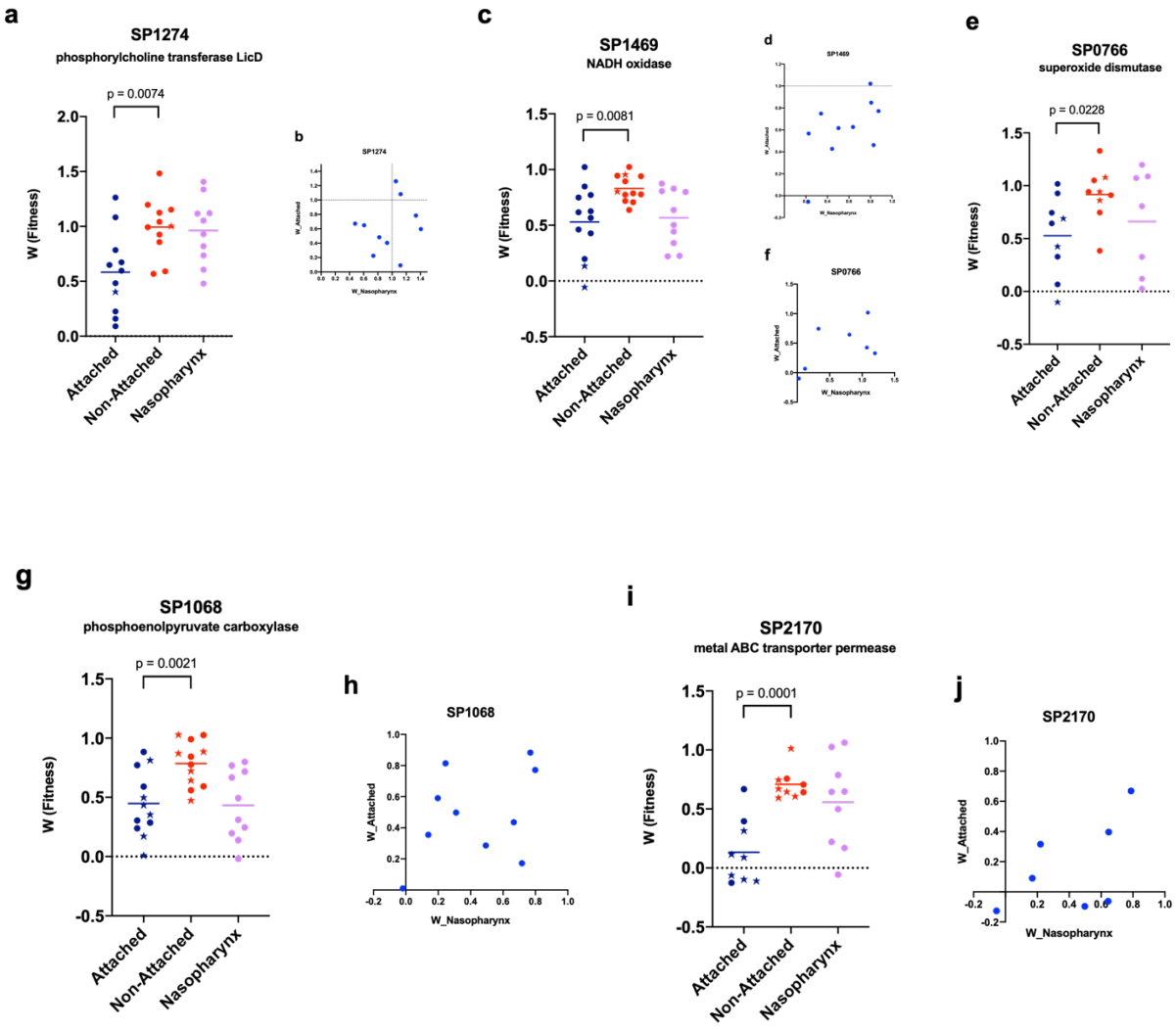

Figure 11. Genes for which W_Attached correlates with W_Nasopharynx, and for which the range of W_Attached is ~1 and/or W_Nasopharynx values ≤ 0.75-1 are considered to be potential strain-dependent attachment-implicated genes. SP1274 and SP1469 have been previously characterized as adhesins **(a, c)**.[19, 33] SP1469 and SP0766 are implicated in the *S. pneumoniae* oxidative stress response (**c-e**). SP1068 is involved in catalyzing a reaction adding bicarbonate to phosphoenolpyruvate to form oxaloacetate, and SP2170 is an ABC transporter permease that is associated with zinc transport (**f**, **h**). The graphs correlating W_Attached and W_Nasopharynx values on an XY plane indicate similar *in vitro* and *in vivo* requirements by the 13 analyzed strains for these genes (**b, d, f, h, j**).

Two other genes that form a phosphotransferase system (PTS) transporter demonstrate a more tenuous strain-dependent attachment requirement; they were identified as genes of particular interest due to the proposed role of PTS transporters in the *S. pneumoniae* oxidative stress response in strain D39.[34] SP0282 **(Fig. 12a)** exhibits a more clear strain-dependency than SP0283 **(Fig. 12b)**, but W_Attached for the 13 considered strains *in vitro* closely matches *in vivo* data in both genes as demonstrated in **(Fig. 12c)**. Interestingly, while these two genes were flagged as discoveries in the Attachment-Implicated group due to statistically- and biologically-significant defects in the W_Attached compared with W_Non-Attached in two different strains (T19F for the former gene and PG16 for the latter), in strains TIGR4 and PG01, designated with asterisk shape in **(c)**, the gene appears to have a hindering effect on attachment indicated by their clustering for W_Attached and W_Nasopharynx around W > 1 in one or both conditions.

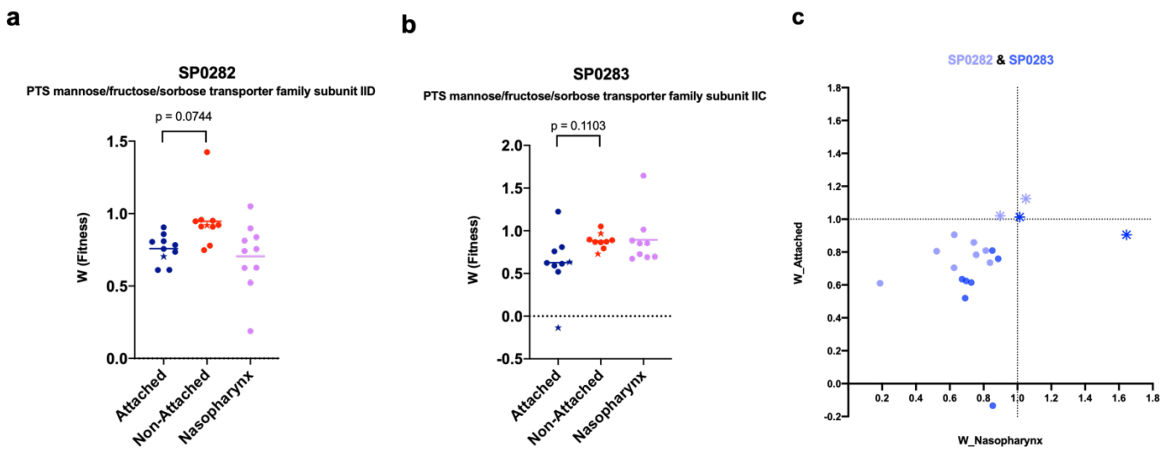

Figure 12. The phosphotransferase system (PTS) transporter made up by SP0282 and SP0283 exhibits a strain-dependent phenotype for attachment requirements in SP0282, and a less clear requirement in SP0283 (**a-b**). **Strains exhibit similar phenotypes *in vitro* as *in vivo*, (c), and it is clear that the four asterisk-shaped data points with W > 1 (representing fitness values for each gene in strains TIGR4 and PG01) reveal an attachment-hindering phenotype of these genes in these two genetic backgrounds.** When TIGR4 and PG01 data is removed from W_Attached and W_Non-Attached populations, the statistical difference between W_Attached values and W_Non-Attached sets is p = 0.0120 in SP0282, and p = 0.0659 in SP0283.

5.3 A PRELIMINARY J774A.1 KILLING ASSAY REVEALED A POTENTIAL TIGR4 REQUIREMENT OF SP1587 TO SURVIVE MACROPHAGE STRESS

One preliminary killing assay of WT TIGR4 and its Δ1587 (MFS transporter) mutant derivative in J774A.1 migratory monocyte/macrophage cell line revealed that the absence of this gene imposed a significant defect in the ability of *S. pneumoniae* to replicate and even survive within the phagosome. Expansion was calculated by dividing the CFU after 2 hours of intracellular incubation by the initial CFU of phagocytosed bacteria. Whereas the WT bacteria replicated 2-3 times in 2 hours (enough time for 3 doublings in pure liquid culture, which would result in an expansion of 8), the mutant bacterial population failed to fully replicate once in the MOI of 1 and 10 conditions, and was successfully killed in roughly half of macrophages within the MOI of 0.1 condition **(Fig. 13a-b)**.

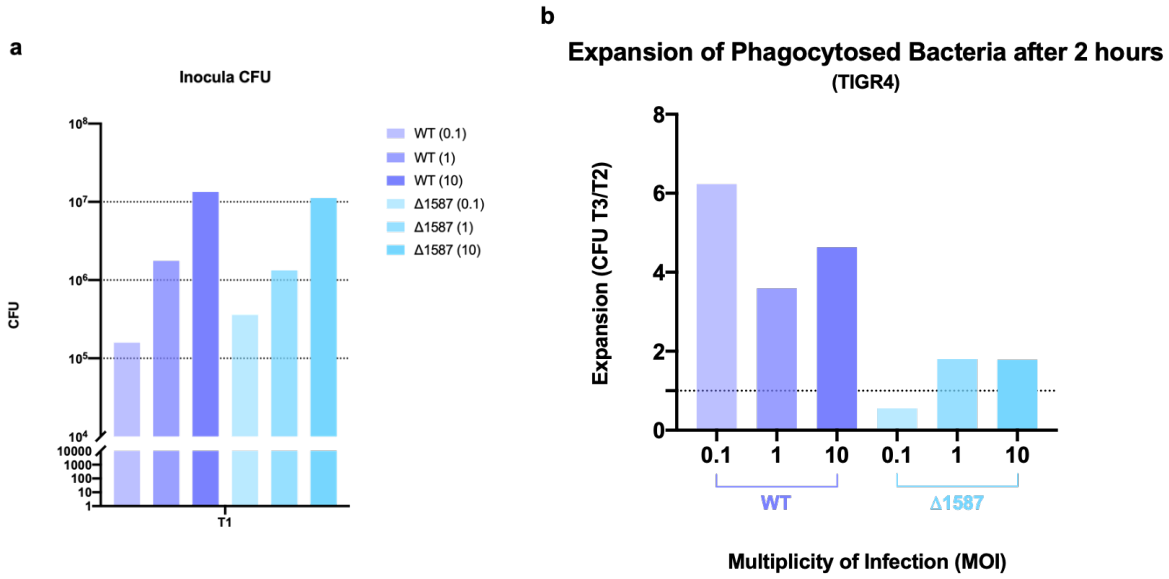

Figure 13. J774A.1 macrophages were incubated with TIGR4 cells at a close approximation of an MOI of 0.1, 1, and 10, for WT and Δ1587 mutant conditions (a), and the stark difference in expansion within the phagosome of the WT vs. Δ1587 mutant after 2 hours indicates a potential requirement by *S. pneumoniae* strain TIGR4 to withstand macrophage-induced stress (b). This experiment was only performed once, and thus does not report multiple replicates and cannot make statistical conclusions.

6.0 DISCUSSION

6.1 *S. pneumoniae* NASOPHARYNGEAL ATTACHMENT REQUIREMENTS

One major goal of this project was to identify novel core- and strain-dependent- factors used by *Streptococcus pneumoniae* to attach to the nasopharyngeal epithelium. Investigation of a subset of genes identified as relevant in the *S. pneumonaie* attachment interaction with D562 human nasopharyngeal cells revealed six potential core attachment-implicated genes among the 13 considered strains, and seven potential strain-dependent attachment-implicated genes. Given these preliminary results, we can state with confidence that diversity in genetic background does imply some level of diversity in gene usage for attachment (i.e. that all strains of *S. pneumoniae* do not employ a single, uniform set of genes to perform this function). Based on preliminary investigation of the extent to which *in vitro*- and *in vivo*-determined requirements of strain-dependent genes correlate within each strain, we can also conclude that our *in vitro* attachment experimental design reproduces broad phenotypic trends from *in vivo* results, affirming that it is a good model to use in order to parse requirements for nasopharyngeal infection **(Fig. 11 b, d, f, h, j; Fig. 12c)**. A strong positive correlation between *in vitro*- and *in vivo*-determined requirements which differ among strains is also indicative of the substantial degree to which the establishment and success of an infection relies on initial adherence to the nasopharyngeal epithelium.

Of the small set of attachment-implicated proteins identified already in the literature, this work confirmed phenotypes in several important genes, and did not show that other described genes are required for adhesion in our experimental set-up. In either case, however, *in vitro* data largely correlated with *in vivo* phenotypes determined by our group. In this work, we found the NADH oxidase protein/*nox* enzyme (SP1469) as required in a strain-dependent manner for

attachment, which correlates with a previous study characterizing the "moonlighting" function of this enzyme as an adhesin to A549 human alveolar basal epithelial cell line in *S. pneumoniae* unencapsulated serotype 2 derived R6 strain and serotype 3 strain WU2 **(Fig. 11c-d)**.[33] We thus demonstrate that this gene is implicated in *S. pneumoniae* attachment throughout the airway epithelia. We also found a strain-dependent requirement of gene phosphorylcholine transferase LicD (SP2174), which has been previously characterized as highly expressed during colonization of airway epithelia by *S. pneumoniae* biofilms, and more specifically involved in the surface expression of a binding partner for platelet activating factor receptor (PAF-R) **(Fig. 11a)**.[37, 33] Another characterized attachment protein, lipoprotein pneumococcal surface adhesin PsA (SP1650), binds epithelial cadherin and did appear in our *in vitro* assay with extremely low W_Attached, however its requirement for growth in EMEM and corresponding extreme-defect phenotype in W_Non-Attached for all strains but TIGR4 and T19F obscured the attachment phenotype **(Fig. 6d)**.[2] We hypothesize that the differing W_Non-Attached results between the previously performed TIGR4 and T19F experiments and the recent experiments in pan-genome strains performed for this work are due to EMEM inconsistency. Our data also affirm the phenotype, demonstrated previously by our group and many others, that non-encapsulated strains and mutants in capsule genes possess a relative advantage in attachment over encapsulated strains *in vitro* (while *in vivo*, any attachment advantage is offset by the disadvantage of interacting with the immune system while lacking a crucial protective feature) **(Fig. 1; Fig. 6 a-c)**.[22, 35]

PspC/CbpA has been implicated in binding to the secretory component on the polymeric immunoglobulin receptor.[2] This work did not find a requirement for this gene among any of the 13 *in vitro*-considered strains, however strains PG25 and PG27 exhibited a biologically significant fitness defect in different PspC domain-containing proteins *in vivo*. Thus, it is

possible that this gene is required only in two out of the 22 pan-genome strains which could be assayed by Tn-seq, neither of which were represented in this analysis and thus could not be corroborated. Pilus-associated adhesin RrgA has been previously implicated in adhesion *in vitro* to A549 respiratory epithelial cells and in *in vivo* nasopharynx infection, both in TIGR4.[38, 39] Our work did not find a requirement for this gene *in vivo* or *in vitro*. The *in vivo* result was ascertained using a transposon mutagenesis system that was a precursor to Tn-seq in the early 2000s; potentially after many modifications to the method, including the updated protocols enabling construction of much more highly saturated libraries which we used in this recent work, this result no longer appears due to relative statistical insignificance. Considering the *in vitro* result, possibly this gene is required differentially between the lungs and the nasopharynx. Finally, neither our *in vitro* nor *in vivo* results corroborated the requirement of pneumococcal adherence and virulence factor PavA, which has been described as binding to the extracellular matrix component fibronectin.[33] With regards to our *in vitro* results, it is likely that the extracellular matrix is not perfectly reproduced in cell culture, thus removing the substrate for this protein. Either of the *in vivo* inconsistencies could also be due to differences in the strain of mouse used by our group vs. others, or due to the bacterial dose inoculated into the animals (our group used 1 x 10^8 CFU, relatively high).

6.1.1 Proposed model of *S. pneumoniae* attachment

Considering the 13 attachment-implicated genes identified in this analysis, we propose a simple model for *S. pneumoniae* requirements to attach to the nasopharyngeal epithelium and/or survive in this attached population. The genes within this set fall into three broad categories: vitamin transport and metabolism, metal transport, and oxidative stress response. An ECF transporter S component (SP1232) which imports pantothenate and a thiamine diphosphokinase

(SP1982) which functions in the thiamine salvage pathway appear as potential core requirements, which we hypothesize is due to a local deficiency in these nutrients at the epithelial interface, where a high density of bacteria and eukaryotic cells are all in competition for them.[43] Three genes which encode subunits of another ECF transporter (SP2219, SP2220, SP2221) also appear as potential core; a previous work analyzing the transcriptome of D39 and the SwissProt annotations suggest that SP2219, SP2220 and SP2221 form a cobalt transporter system.[43] Phosphorylcholine transferase LicD (SP1274), a phosphorylcholine metabolism protein previously demonstrated to be attachment-implicated (phosphorylcholine is an adhesin), appeared to be required strain-dependently.[45, 33] The potential strain-dependent set also includes phosphoenolpyruvate carboxylase, SP1068, which utilizes phosphoenolpyruvate for the production of aspartate (again, possibly indicative of local nutrient scarcity).[44] A zinc transporter, SP2170, also appears to be required strain-dependently. One potential role for this transporter is in supplying zinc for the chaperone protein (SP2188) discussed below, which binds zinc in its inactive state.[43] The remainder of the attachment-implicated genes all have demonstrated or predicted roles in oxidative stress response. NADH oxidase (SP1469), and superoxide dismutase (SP0766), both in the strain-dependent group, are well-characterized oxidative stress response proteins; the former reduces free oxygen to water and the latter catalyzes the partitioning of the superoxide radical into O_2 and hydrogen peroxide. As described, NADH oxidase also "moonlights" as an adhesin, whereas no such role has been described for superoxide dismutase.[33] The potential core-required Hsp33 family chaperone protein HslO (SP2188) is described as activated under oxidative stress, and associates with zinc atoms in the inactive state as stated.[43] The mannose/fructose/sorbose PTS transporter is encoded by SP0282 and SP0283, which, when removing outlying data points PG01 and TIGR4, display a potential strain-dependent requirement (the former) and a nonsignificant trend of W_Attached < W_Non-Attached (the

latter); PTS transporters were found to be upregulated in a transcriptional analysis of D39 under oxidative stress.[34] This system inhibits escape of imported sugars from the cell by phosphorylation (sourcing the phosphate group from phosphoenolpyruvate), and may be activated in response to local glucose deficiency or increased carbohydrate need during the oxidative stress response **(Fig. 14)**.

We hypothesize that *S. pneumoniae* must be able to competitively import and synthesize vitamins and other nutrients due their local depletion at the epithelial interface, as well as withstand significant oxidative stress in this microenvironment in order to survive and thus remain in the attached population.

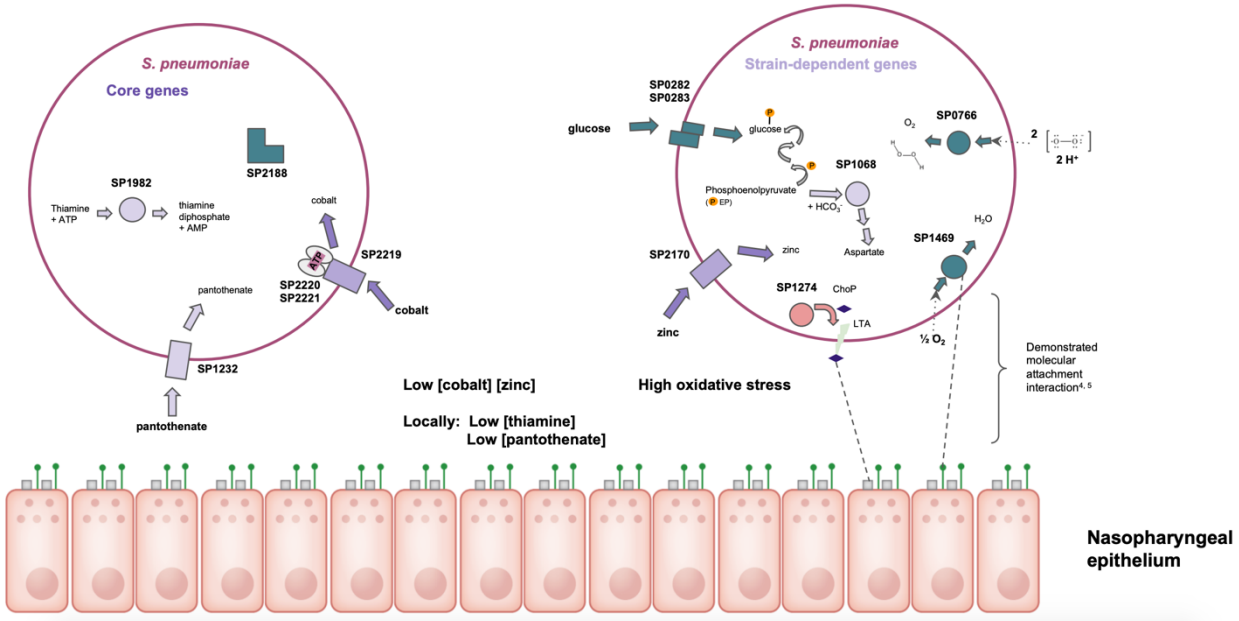

Figure 14. Proposed model of *S. pneumoniae* core and strain-dependent genetic requirements for survival at the interface of the nasopharyngeal epithelium based on attachment-implicated genes identified in this project. Transporters were represented as rectangles spanning the *S. pneumoniae* membrane, enzymes were represented as circles, and the chaperone protein was represented as an L-shaped block. LTA refers to lipoteichoic acid and ChoP refers to phosphorylcholine. Oxidative stress-related proteins are represented in green, vitamin/nutrient transport and metabolism proteins are represented in light purple, and metal transporters are represented in darker purple. The cartoon epithelial cells are adapted from Figure 1 of "Mechanisms of Bacterial Colonization of the Respiratory Tract." [2]

6.2 *S. pneumoniae* INNATE IMMUNE SURVIVAL REQUIREMENTS

One important goal which this project sought to address was to corroborate, from *in vivo* results in immune cell-depleted mice, the universal requirement of an MFS transporter (SP1587) by *S. pneumoniae* for survival of phagocytosis by the macrophage. Although statistically confident conclusions with regards to this goal cannot be made, preliminary investigation of this requirement revealed a promisingly stark disparity in expansion of the TIGR4Δ1587 mutant population compared to the WT TIGR4 population within the phagosome of migratory macrophage cells belonging to cell line J774A.1 **(Fig. 13)**. As previously demonstrated in an *in vivo* infection model by strain D39, these data confirm that WT *S. pneumoniae* is able to replicate within the phagosome due to the innate ability of the bacterium to survive most macrophage-induced stress.[40] Future work, if these results are replicated, will enable the identification of a specific factor that underlies this innate ability.

6.3 LIMITATIONS AND FUTURE DIRECTIONS

6.3.1 Nasopharyngeal attachment

One limitation of assaying relative attachment fitness in single gene mutants, as we did by Tn-seq, is that adhesion of *Streptococcus pneumoniae* to host epithelium is likely a multifactorial process, employing multiple protein-protein interactions at a given time. Thus, it is possible that disrupting the function of solely one gene within a set of attachment-implicated genes does not always result in low attachment phenotype. Due to this limitation of Tn-seq, the role of some gene products as adhesins may be obscured by seemingly moderate fitness defects. An additional limitation in the use of *ex vivo* corroboration is the lack or attenuation of important *in vivo* components of the nasopharyngeal environment apart from the epithelium, such as the extracellular matrix and the nasal mucosa.

The most important measure of the validity and biological relevance of these data will be single-mutant validation experiments in selected genes. We propose future validation of attachment phenotypes identified by Tn-seq in three potential core attachment-implicated genes (SP1232, SP2219, SP2188), as well as three potential strain-dependent attachment-implicated genes (SP0766, SP1068, SP2170). If Tn-seq determined phenotypes are validated, this will allow us to consider this assay as a biologically relevant model and lend confidence toward fitness predictions in the entire set of genes which were included in this work. One future analysis to be performed on the data obtained from this work is a functional enrichment analysis to determine which functional gene groups are overrepresented in the attachment-implicated gene population. To address the hypothesized local nutritional deficiencies at the epithelial interface, attachment Tn-seq could be repeated with (for example) pantothenate- and thiamine- supplemented EMEM, with the expectation that mutants in these vitamin transport and metabolism pathways would no

longer exhibit an attachment defect if the nutrient scarcity is remedied. Another important step in investigation of interesting gene products for which a demonstrated inhibitory drug or antibody (for membrane or cell-wall-expressed proteins) already exits will be to perform the attachment assay after incubating *S. pneumoniae* with the inhibitory agent; an attenuation of adhesion to epithelium would provide a strong indication that the gene is an adhesin.

6.3.2 Macrophage survival

A limitation inherent in the experimental design of this killing assay is the difficulty in ensuring that no extracellular bacteria remain after the first engulfment incubation period. Previous work by our group has revealed some *S. pneumoniae* drawdown into the pellet even at the low relative centrifugal force used in this assay. We attempt to minimize the presence of non-engulfed bacteria by performing this centrifugation and washing step three times, however it is possible that extracellular bacteria remain and distort CFU counts for surviving bacteria after the two-hour incubation. (Given the results of this preliminary assay with regards to mutant expansion, it is unlikely that many extracellular bacteria remain after centrifugation, but still a possibility). In the context of this work, a significant limitation is that we cannot make any confident or statistically powerful conclusions, due to the lack of replicate experiments because of the pandemic-imposed time constraint.

To further explore this interesting result, we propose that replicate experiments be performed in TIGR4 as well as PG12 and its MFS transporter (SP1587) mutant derivative. This killing assay will also be performed with prior activation of the J774A.1 macrophage cell line to produce ROS and RNS, under which conditions NADH oxidase (SP1469) mutants can also be assayed. Finally, all WT and mutant strains will be assayed for survival after engulfment by the

alveolar macrophage cell line MH-S as well, to validate and uncover the extent of *S. pneumoniae* requirements of these genes to survive attack by two innate immune cells.

6.4 CONCLUSIONS

The major goal of this project was to identify the specific role of factors within the virulome (the set of genes required by *Streptococcus pneumoniae* to execute successful infection of a host at the pan-genome level) by correlating the genetic requirements of pan-genome strains to attach to nasopharyngeal epithelial cells and survive interaction with circulating macrophages *in vitro* with this *in vivo*-determined set. We have successfully parsed 13 factors from the *in vivo* set which exhibit corresponding fitness defects in the *in vitro* nasopharyngeal environment, including six potential core requirements and seven potential strain-dependent requirements, two of which have been previously characterized as possessing attachment roles. Pending validation of our method as proposed, we have produced a comprehensive dataset likely harboring further novel gene requirements for nasopharyngeal attachment by considered strains, including virulent and non-virulent isolates. Analysis of trends and disparities between the two sets can yield new insights regarding the genetic underpinnings of the colonization state versus disease state. Our preliminary *in vitro* investigation of a macrophage-interaction requirement demonstrated a promisingly strong correlation with *in vivo* results. In sum, we have uncovered novel and medically relevant pneumococcal biology, and made significant strides toward parsing the *Streptococcus pneumoniae* virulome comprehensively.

BOSTON COLLEGE
Morrissey College of Arts & Sciences

The Thesis of: Emily Rudmann

Title: Parsing the *Streptococcus pneumoniae* virulome

Submitted to the Department of Biology in the Morrissey College of Arts & Sciences has been read and approved by the Advisor:

_____
Advisor (Print Name)

_____
Advisor (Signature)

_____
Date

7.0 REFERENCES

[1] Wilson, Brenda A., et al. "Bacterial Pathogenesis: A Molecular Approach." Fourth Edition. Amazon, ASM Press, 2019.

[2] Siegel, S. J., & Weiser, J. N. (2015). Mechanisms of Bacterial Colonization of the Respiratory Tract. *Annual Review of Microbiology*, *69*(1), 425–444. https://doi.org/10.1146/annurev-micro-091014-104209

[3] Lowth, M. (2015). Pneumococcal disease. *Practice Nurse*, *45*(10). https://doi.org/10.5005/jp/books/12557_20

[4] Song, J. Y., Nahm, M. H., & Moseley, M. A. (2013). Clinical implications of pneumococcal serotypes: Invasive disease potential, clinical presentations, and antibiotic resistance. In *Journal of Korean Medical Science* (Vol. 28, Issue 1, pp. 4–15). https://doi.org/10.3346/jkms.2013.28.1.4

[5] van Opijnen, T., Dedrick, S., & Bento, J. (2016). Strain Dependent Genetic Networks for Antibiotic-Sensitivity in a Bacterial Pathogen with a Large Pan-Genome. *PLoS Pathogens*, *12*(9). https://doi.org/10.1371/journal.ppat.1005869

[6] *Antibiotic resistance threats in the United States, 2019*. (2019). https://doi.org/10.15620/cdc:82532

[7] Van Opijnen, T., Bodi, K. L., & Camilli, A. (2009). Tn-seq; high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature Methods* 6(10): 767-772.

[8] Poulsen, B. E., Yang, R., Clatworthy, A. E., White, T., Osmulski, S. J., Li, L., Penaranda, C., Lander, E. S., Shoresh, N., & Hung, D. T. (2019). Defining the core essential genome of Pseudomonas aeruginosa. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(20), 10072–10080. https://doi.org/10.1073/pnas.1900570116

[9] Clatworthy, A. E., Pierson, E., & Hung, D. T. (2007). Targeting virulence: A new paradigm for antimicrobial therapy. In *Nature Chemical Biology* (Vol. 3, Issue 9, pp. 541–548). Nature Publishing Group. https://doi.org/10.1038/nchembio.2007.24

[10] Vernikos, G., Medini, D., Riley, D. R., & Tettelin, H. (2015). Ten years of pan-genome analyses. *Current Opinion in Microbiology*, *23*, 148–154. https://doi.org/10.1016/J.MIB.2014.11.016

[11] Lapierre, P., & Gogarten, J. P. (2009). Estimating the size of the bacterial pan-genome. *Trends in Genetics*, *25*(3), 107–110. https://doi.org/10.1016/J.TIG.2008.12.004

[12] Blaustein, R. A., McFarland, A. G., Ben Maamar, S., Lopez, A., Castro-Wallace, S., & Hartmann, E. M. (2019). Pangenomic Approach To Understanding Microbial Adaptations

within a Model Built Environment, the International Space Station, Relative to Human Hosts and Soil. *MSystems*, *4*(1). https://doi.org/10.1128/msystems.00281-18

[13] Caputo, A., Fournier, P. E., & Raoult, D. (2019). Genome and pan-genome analysis to classify emerging bacteria. In *Biology Direct* (Vol. 14, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s13062-019-0234-0

[14] Ortiz-Marquez J.C., Zhu Z., Surujon D., Wood S., Rosconi F., and van Opijnen T. (2020). A pan-genomic perspective on the emergence, maintenance and predictability of antibiotic resistance. Nature springer - in press.

[15] Tong, A. H. Y., Lesage, G., Bader, G. D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G. F., Brost, R. L., Chang, M., Chen, Y., Cheng, X., Chua, G., Friesen, H., Goldberg, D. S., Haynes, J., Humphries, C., He, G., Hussein, S., … Boone, C. (2004). Global Mapping of the Yeast Genetic Interaction Network. *Science*, *303*(5659), 808. https://doi.org/10.1126/science.1091317

[16] Roguev, A., Bandyopadhyay, S., Zofall, M., Zhang, K., Fischer, T., Collins, S. R., Qu, H., Shales, M., Park, H. O., Hayles, J., Hoe, K. L., Kim, D. U., Ideker, T., Grewal, S. I., Weissman, J. S., & Krogan, N. J. (2008). Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science*, *322*(5900), 405–410. https://doi.org/10.1126/science.1162609

[17] Hiller, N. L., & Sá-Leão, R. (2018). Puzzling Over the Pneumococcal Pangenome. *Frontiers in Microbiology*, *9*. https://doi.org/10.3389/fmicb.2018.02580

[18] Cremers, A. J. H., Mobegi, F. M., De Jonge, M. I., Van Hijum, S. A. F. T., Meis, J. F., Hermans, P. W. M., Ferwerda, G., Bentley, S. D., & Zomer, A. L. (2015). The post-vaccine microevolution of invasive Streptococcus pneumoniae. *Scientific Reports*, *5*. https://doi.org/10.1038/srep14952

[19] Weiser, J. N., Ferreira, D. M., & Paton, J. C. (2018). Streptococcus pneumoniae: Transmission, colonization and invasion. In *Nature Reviews Microbiology* (Vol. 16, Issue 6, pp. 355–367). Nature Publishing Group. https://doi.org/10.1038/s41579-018-0001-8

[20] Fleming-Dutra, K. E., Conklin, L., Loo, J. D., Knoll, M. D., Park, D. E., Kirk, J., Goldblatt, D., Whitney, C. G., & O'Brien, K. L. (2014). Systematic review of the effect of pneumococcal conjugate vaccine dosing schedules on vaccine-type nasopharyngeal carriage. In *Pediatric Infectious Disease Journal* (Vol. 33, Issue SUPPL. 2). https://doi.org/10.1097/INF.0000000000000083

[21] Yang, A., Cai, F., & Lipsitch, M. (2019). Herd immunity alters the conditions for performing dose schedule comparisons: an individual-based model of pneumococcal carriage. *BMC Infectious Diseases*, *19*(1). https://doi.org/10.1186/s12879-019-3833-6

[22] Li J, Zhang JR. (2009). Phase variation of Streptococcus pneumoniae. *Microbial Spectrum*, 7(1):GPP3-0005-2018. https://doi.org/10.1128/microbiolspec

[23] Barocchi, M. A., Ries, J., Zogaj, X., Hemsley, C., Albiger, B., Kanth, A., Dahlberg, S., Fernebro, J., Moschioni, M., Masignani, V., Hultenby, K., Taddei, A. R., Beiter, K., Wartha, F., von Euler, A., Covacci, A., Holden, D. W., Normark, S., Rappuoli, R., & Henriques-Normark, B. (2006). A pneumococcal pilus influences virulence and host inflammatory responses. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(8), 2857–2862. https://doi.org/10.1073/pnas.0511017103

[24] Flannagan, R. S., Heit, B., & Heinrichs, D. E. (2015). Antimicrobial mechanisms of macrophages and the immune evasion strategies of Staphylococcus aureus. In *Pathogens* (Vol. 4, Issue 4, pp. 826–868). MDPI AG. https://doi.org/10.3390/pathogens4040826

[25] Aberdein, J. D., Cole, J., Bewley, M. A., Marriott, H. M., & Dockrell, D. H. (2013). Alveolar macrophages in pulmonary host defense-the unrecognized role of apoptosis as a mechanism of intracellular bacterial killing. In *Clinical and Experimental Immunology* (Vol. 174, Issue 2, pp. 193–202). https://doi.org/10.1111/cei.12170

[26] Pechous, R. D. (2017). With friends like these: The complex role of neutrophils in the progression of severe pneumonia. In *Frontiers in Cellular and Infection Microbiology* (Vol. 7, Issue MAY). Frontiers Media S.A. https://doi.org/10.3389/fcimb.2017.00160

[27] Bordon, J., Aliberti, S., Fernandez-Botran, R., Uriarte, S. M., Rane, M. J., Duvvuri, P., Peyrani, P., Morlacchi, L. C., Blasi, F., & Ramirez, J. A. (2013). Understanding the roles of cytokines and neutrophil activity and neutrophil apoptosis in the protective versus deleterious inflammatory response in pneumonia. In *International Journal of Infectious Diseases* (Vol. 17, Issue 2). https://doi.org/10.1016/j.ijid.2012.06.006

[28] Szliszka, E., Kucharska, A. Z., Sokół-Łętowska, A., Mertas, A., Czuba, Z. P., & Król, W. (2013). Chemical composition and anti-inflammatory effect of ethanolic extract of Brazilian green propolis on activated J774A.1 macrophages. *Evidence-Based Complementary and Alternative Medicine*, *2013*. https://doi.org/10.1155/2013/976415

[29] van Opijnen, T., Lazinski, D.W. and Camilli, A. (2014), Genome-Wide Fitness and Genetic Interactions Determined by Tn-seq, a High-Throughput Massively Parallel Sequencing Method for Microorganisms. Current Protocols in Molecular Biology, 106: 7.16.1-7.16.24. doi:10.1002/0471142727.mb0716s106

[30] Drevets, D. A., Canono, B. P., & Campbell, P. A. (2015). Measurement of Bacterial Ingestion and killing by macrophages. *Current Protocols in Immunology*, *2015*, 14.6.1-14.6.17. https://doi.org/10.1002/0471142735.im1406s109

[31] Tsikas, D. (2007). Analysis of nitrite and nitrate in biological fluids by assays based on the Griess reaction: Appraisal of the Griess reaction in the l-arginine/nitric oxide area of research. *Journal of Chromatography B,* 851 (1-2) pp: 51-70. DOI: 10.1016j.jchromb.2006.07.054

[32] Khan, P., Idrees, D., Moxley, M. A., Corbett, J. A., Ahmad, F., Von Figura, G., Sly, W. S., Waheed, A., & Hassan, M. I. (2014). Luminol-based chemiluminescent signals: Clinical

and non-clinical application and future uses. In *Applied Biochemistry and Biotechnology* (Vol. 173, Issue 2, pp. 333–355). Humana Press Inc. https://doi.org/10.1007/s12010-014-0850-1

[33] 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. *PLoS ONE*, *8*(4). https://doi.org/10.1371/journal.pone.0061128

[34] Bortoni, M. E., Terra, V. S., Hinds, J., Andrew, P. W., & Yesilkaya, H. (2009). The pneumococcal response to oxidative stress includes a role for Rgg. *Microbiology*, *155*(12), 4123–4134. https://doi.org/10.1099/mic.0.028282-0

[35] Marks, L. R., Iyer Parameswaran, G., & Hakansson, A. P. (2012). Pneumococcal interactions with epithelial cells are crucial for optimal biofilm formation and colonization in vitro and in vivo. *Infection and Immunity*, *80*(8), 2744–2760. https://doi.org/10.1128/IAI.00488-12

[36] Brooks, L. R. K., & Mias, G. I. (2018). Streptococcus pneumoniae's virulence and host immunity: Aging, diagnostics, and prevention. *Frontiers in Immunology*, *9*(JUN). https://doi.org/10.3389/fimmu.2018.01366

[37] Pettigrew, M. M., Marks, L. R., Kong, Y., Gent, J. F., Roche-Hakansson, H., & Hakansson, A. P. (2014). Dynamic changes in the Streptococcus pneumoniae transcriptome during transition from biofilm formation to invasive disease upon influenza A virus infection. *Infection and Immunity*, *82*(11), 4607–4619. https://doi.org/10.1128/IAI.02225-14

[38] Nelson, A. L., Ries, J., Bagnoli, F., Dahlberg, S., Fälker, S., Rounioja, S., Tschöp, J., Morfeldt, E., Ferlenghi, I., Hilleringmann, M., Holden, D. W., Rappuoli, R., Normark, S., Barocchi, M. A., & Henriques-Normark, B. (2007). RrgA is a pilus-associated adhesin in Streptococcus pneumoniae. *Molecular Microbiology*, *66*(2), 329–340. https://doi.org/10.1111/j.1365-2958.2007.05908.x

[39] Hava, D. L., & Camilli, A. (2002). Large-scale identification of serotype 4 Streptococcus pneumoniae virulence factors. *Molecular Microbiology*, *45*(5), 1389–1406. https://doi.org/10.1046/j.1365-2958.2002.t01-1-03106.x

[40] Ercoli, G., Fernandes, V. E., Chung, W. Y., Wanford, J. J., Thomson, S., Bayliss, C. D., Straatman, K., Crocker, P. R., Dennison, A., Martinez-Pomares, L., Andrew, P. W., Moxon, E. R., & Oggioni, M. R. (2018). Intracellular replication of Streptococcus pneumoniae inside splenic macrophages serves as a reservoir for septicaemia. *Nature Microbiology*, *3*(5), 600–610. https://doi.org/10.1038/s41564-018-0147-1

[41] Manzoor, I., Shafeeq, S., & Kuipers, O. P. (2015). Transcriptome analysis of Streptococcus pneumoniae D39 in the presence of cobalt. *Genomics Data*, *6*, 151–153. https://doi.org/10.1016/j.gdata.2015.08.033

[42] Gámez, G., Castro, A., Gómez-Mejia, A., Gallego, M., Bedoya, A., Camargo, M., & Hammerschmidt, S. (2018). The variome of pneumococcal virulence factors and regulators. *BMC Genomics*, *19*(1), 1–18. https://doi.org/10.1186/s12864-017-4376-0

[43] Chen, I. M. A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., Huntemann, M., Varghese, N., White, J. R., Seshadri, R., Smirnova, T., Kirton, E., Jungbluth, S. P., Woyke, T., Eloe-Fadrosh, E. A., Ivanova, N. N., & Kyrpides, N. C. (2019). IMG/M v.5.0: An integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Research*, *47*(D1), D666–D677. https://doi.org/10.1093/nar/gky901

[44] Härtel, T., Eylert, E., Schulz, C., Petruschka, L., Gierok, P., Grubmüller, S., Lalk, M., Eisenreich, W., & Hammerschmidt, S. (2012). Characterization of central carbon metabolism of Streptococcus pneumoniae by isotopologue profiling. *Journal of Biological Chemistry*, *287*(6), 4260–4274. https://doi.org/10.1074/jbc.M111.304311

[45] Zhang, J. R., Idanpaan-Heikkila, I., Fischer, W., & Tuomanen, E. I. (1999). Pneumococcal licD2 gene is involved in phosphorylcholine metabolism. *Molecular Microbiology*, *31*(5), 1477–1488. https://doi.org/10.1046/j.1365-2958.1999.01291.x

[46] Clark, S. E., & Weiser, J. N. (2013). Microbial modulation of host immunity with the small molecule phosphorylcholine. *Infection and Immunity*, *81*(2), 392–401. https://doi.org/10.1128/IAI.01168-12