Pneumococcal protein antigen serology varies with age and may predict antigenic profile of colonizing isolates

Azarian T[1], Grant LR[2], Georgieva M[1], Hammitt LL[2], Reid R[2], Bentley SD[3], Goldblatt D[4], Santosham M[2], Weatherholtz R[2], Burbidge P[4], Goklish N[2], Thompson CM[1], Hanage WP[1], O'Brien KL[2], Lipsitch M[1]

1 Center for Communicable Disease Dynamics, Department of Epidemiology, T.H. Chan School of Public Health, Harvard University; 2 Center for American Indian Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland; 3 Wellcome Trust Sanger Institute, Cambridge, UK; 4 Immunobiology Section, Institute of Child Health, University College London, London, UK.

Word Count:
Abstract: 200
Main test: 3497
References: 37

Lindsay R. Grant lgrant10@jhu.edu
Maria Georgieva georgiev@hsph.harvard.edu
Laura Hammitt lhammitt@jhu.edu
Ray Reid rreid2@jhu.edu
Stephen D Bentley sdb@sanger.ac.uk
David Goldblatt d.goldblatt@ucl.ac.uk
Mathuram Santosham msantosham@jhu.edu
Robert Weatherholtz rweathe1@jhu.edu
Polly Burbidge p.burbidge@ucl.ac.uk
Novalene Goklish ngoklish@jhsph.edu
Claudette M Thompson cthompso@hsph.harvard.edu
Bill Hanage whanage@hsph.harvard.edu
Kate O'Brien klobrien@jhsph.edu
Marc Lipsitch mlipsitc@hsph.harvard.edu

Corresponding Author:
Taj Azarian, PhD MPH
352.494.2011
Center for Communicable Disease Dynamics,
Harvard T.H. Chan School of Public Health,
677 Huntington Avenue, Suite 506, Boston, MA 02115
tazarian@hsph.harvard.edu

Abstract

Background: Several *Streptococcus pneumoniae* proteins play a role in pathogenesis and are being investigated as vaccine targets. It is largely unknown whether naturally-acquired antibodies reduce the risk of colonization with strains expressing a particular antigenic variant.

Methods: Serum IgG titers to 28 pneumococcal protein antigens were measured among 242 individuals, aged < 6 months - 78 years in Native American communities between 2007-2009. Nasopharyngeal swabs were collected at least 30 days after serum collection, and the protein antigen variant in each pneumococcal isolate was determined using genomic data. We assessed the association between preexisting variant-specific antibody titers and subsequent carriage of pneumococcus expressing a particular antigen variant.

Results: Antibody titers often increased across pediatric groups before decreasing among adults. PspA and StkP IgG titers decreased from <6 months to 6-12 months ($p<0.01$). Individuals with low titers against Group 3 PspC variants were more likely to be colonized with pneumococci expressing those variants. For other antigens, variant-specific IgG titers do not predict colonization with pneumococci expressing particular variants.

Conclusion: We observed an inverse association between variant-specific antibody concentration and homologous pneumococcal colonization for only one protein. Further assessment of antibody repertoires may elucidate the nature of anti-pneumococcal antibody-mediated mucosal immunity while informing future vaccine development.

Key words: Streptococcus pneumoniae, pneumococci, protein antigens, sera, immunology, PspC, PspA, vaccine, pilus, antibody

Introduction

Current pneumococcal conjugate vaccines have significantly reduced invasive disease caused by the included *Streptococcus pneumoniae* (pneumococcal) serotypes. However, the currently licensed vaccines, PCV-10 and PCV-13, target only 10 or 13 of the ~90 recognized pneumococcal capsular serotypes. In addition to incomplete coverage of disease-causing types, significant disadvantages of capsular vaccines include their production cost, production complexity, and serotype replacement. While PCV formulations are still an attractive vaccine approach, these limitations have motivated pursuit of pneumococcal protein antigens as vaccine candidates. Protein-based vaccines would, in theory, generate robust antibody responses, be efficacious in young children, and may decrease carriage (1).

Pneumococcal surface protein A (PspA), PspC, pilus (RrgA/B/C), pneumolysin (Ply) and neuraminidase (NanA) are among the pneumococcal proteins being investigated for use in vaccine formulations (1). Studies suggest that in some cases combinations of these proteins may elicit better protection than any of the proteins themselves (1,2). In humans, antibodies to pneumococcal proteins can be detected during colonization and natural infection, providing protection from subsequent colonization and invasive disease (3–8). Virolainen *et al.* showed that among children with invasive pneumococcal infections, those with lowest antibody titers to PspA were infected most frequently with pneumococci (9). However, animal data show that while antiprotein antibodies are correlated with protection against subsequent challenge, the mechanism of protection is not necessarily antibody mediated, suggesting antibody levels may correlate with degree of immune response but not necessarily exclusively mediate protection. Evidence of variant-specific protection, in which antibodies to a particular protein antigenic variant correlate with protection against colonization by homologous pneumococci (i.e. those with that protein variant), would be more strongly

indicative of antiprotein antibodies' causal role in protection, as has been observed for serotype-specific anticapsular antibodies (10,11).

At the same time, such evidence would provide a mechanism to explain the high level of sequence variation and signs of diversifying selection at these loci. While these protein antigens are present in almost all pneumococci, they are also very diverse, and pneumococcal strains differ considerably in the particular antigenic variants they express (12). Two clear examples are the surface associated choline-binding proteins PspA and PspC. Both are encoded by polymorphic genes with clear structural variability, which becomes the basis for their division into 3 PspA families and 11 PspC groups (12,13). Studies suggest structural differences in these proteins impact the nature and specificity of the antibody response generated toward them. For example, family-specific antibody responses among children exposed to pneumococci possessing family 1 and 2 PspA variants have been observed (14).

Important uncertainties remain about the biologic function of protein antibodies and the extent to which their binding and activity are specific to particular variants of polymorphic antigens. Here, we investigate whether naturally acquired antibodies to protein antigens reduce the risk of nasopharyngeal acquisition (i.e. colonization) with strains containing particular variants of diverse proteins such as PspA and PspC. To address this question, we used pneumococcal genomic data to identify variants of 21 pneumococcal protein antigens present in *S. pneumoniae* carriage isolates. We then assessed the association between antibody titers and subsequent colonization with *S. pneumoniae* strains expressing an antigen recognized by preexisting antibodies. We posited that individuals who had low antibody titers to a specific protein antigen variant would be more likely to be colonized with *S. pneumoniae* expressing that variant.

Methods

Study Population, Serum Collection, and Nasopharyngeal Colonization

Individuals included in this study were a subset of participants in a larger prospective, longitudinal, observational cohort study of pneumococcal carriage among Navajo and White Mountain Apache families described elsewhere (15). Briefly, participants living on reservations in the southwest USA were enrolled from March 2006 to March 2008. Demographic and epidemiological data are provided in Supplementary Table 1. Serum and nasopharyngeal (NP) specimens were obtained on the initial visit after recruitment, and NP samples were collected at each of six follow-up visits at one-month intervals to determine pneumococcal carriage status (16). We selected individuals who were negative for carriage at the initial visit, had an available enrollment serum sample, and subsequently had a pneumococcal isolate detected $\geq$30 days after serum collection. The Navajo Nation, White Mountain Apache tribe and the IRBs of the Johns Hopkins Bloomberg School of Public Health, the Navajo Nation and the Phoenix Area IHS approved this study. Written informed consent was obtained from adult participants and from caregivers of child participants. Assent was obtained from children 7-17 years.

Protein Antigen Serology

Serum IgG to 28 pneumococcal protein antigens were measured using direct binding electrochemiluminescence-based multiplex assay (MSD, Rockville, MD) (7). Antibody levels among participants are expressed as a titer relative to the amount in a reference serum. The 28 antigens represent 21 pneumococcal proteins and alleles or structural variants of polymorphic proteins (Table 1). In particular, antibody titers were

148 measured for variants of PcpC, PspA, pilus subunit RrgB, pneumolysin (ply), and
149 pneumococcal histidine triad D (PhtD). Four PspC variants, representing four of the 11
150 recognized major groups (12), were selected based on their prevalence in a
151 pneumococcal carriage study in Massachusetts, USA (17). Two variants, var-I and var-
152 II, contain choline-binding domains (CBD), while var-III (Group 7) and var-IV (Group 8)
153 contain the LPXTG (sortase) motif. Truncated PspC constructs were designed to
154 uniquely represent each PspC variant. Truncation removed the proline-rich and cell wall
155 anchor motifs, which are highly homologous to those found in PspA (Supplementary
156 Figure 1).
157
158 *Genome Sequencing and Protein Antigen Identification*
159
160 Genomic DNA from *S. pneumoniae* isolates were sequenced on the Illumina
161 HiSeq to produce paired-end 100 bp reads at ≥30-fold coverage. Serotypes were
162 determined by mapping reads to concatenated CPS locus sequences of 90
163 pneumococcal serotypes using SRST2 (18,19). *De novo* genome assemblies were
164 generated with Velvet (20) and annotated using Prokka (21). Pangenome analysis was
165 conducted with Roary (22) to cluster and abstract protein antigen genes. The coding
166 sequence for each protein antigen in the MSD assay was downloaded from KEGG
167 (http://www.kegg.jp/) and orthologs from *S. pneumoniae* reference strains were identified
168 (Table 1). The MSD index variant, *S. pneumoniae* reference orthologs, and *de novo*
169 assembled protein antigen genes were aligned with PRANK using a codon-aware
170 algorithm (23). The diversity of each protein antigen among sequenced isolates was
171 assessed to classify each antigen as polymorphic or conserved. Maximum likelihood
172 phylogenies were inferred from the alignments using RAxML v8.0.0. For polymorphic
173 antigens, one reference variant was selected from each monophyletic clade and used to

construct a protein database for SRST2 (19). To determine the antigenic profile of each carried pneumococci, reads were mapped to each variant and the highest scoring match was reported.

Statistical Analysis

To investigate correlation among participants' anti-protein titers and association with age, we performed hierarchical clustering of 28 titers using the mclust v5.2 package in R. We then assessed the association between titers and subsequent carriage of a pneumococcus possessing alleles against which the participant had antibodies. For each polymorphic protein antigen, log antibody titers were compared among individuals carrying a pneumococcus with the respective variant. Statistical significance was assessed using analysis of variance and Tukey's HSD. For the pilus (RrgA/B), which is present in only a fraction of pneumococci, we assessed anti-pilus titers between individuals colonized with piliated and non-piliated strains.

Results

This analysis included 242 participants who had new pneumococcal colonization $\geq$30 days after serum collection (range 30-225 days, median=69). Individuals ranged in age from <1 month to 78 years of age, and had 34 pneumococcal serotypes identified in addition to 4 non-typable isolates (Supplementary Table 2). For two participants, two serotypes were identified, and a single serotype was selected for WGS. Among participants, 14 carried pneumococci that were PCV-7 vaccine types, and the distribution of PCV-7 types did not significantly differ by age group (Fisher's Exact, p=0.28).

Classification of pneumococcal protein antigens

Phylogenetic analysis identified 11 polymorphic protein antigens and 10 that were largely conserved. Table 2 lists the variant frequencies for the 21 protein antigens. In some cases, the protein antigen variant was unable to be assigned; therefore, counts for some antigens do not total 242. We measured antibody titers to multiple variants of three proteins with polymorphisms (PspC, PspA, and RrgB). For the remaining 10 proteins, we measured antibody titers for one variant and compared titers among individuals with carriage isolates possessing polymorphic variants.

Protein Antigen Serology

Participants' responses to different PspC variants were positively correlated, ($r=0.29$-0.90, $p<0.05$ for all correlations), such that those with high PspC var-I titers also had high titers to the other PspC variants (Figure 1 and Supplementary Figure 2). Titers against two of the three RrgB variants were highly correlated with each other (0.96, $p<0.01$), and less correlated with those against Var-III (Figure 1 and Supplemental Figure 3). Among other protein antigens, we found significant correlation between variants of the same antigen (Figure 1). There was also high correlation among antigens including PiuA, PiaA, PcsB, Spr2021, PcpA, CbpA, PhtE, and PhtD.

Antibody titers for most proteins increased with age including PspC variants (Figure 2A), PspA variants (Figure 2B), Ply variants (Figure 2D) and others (Supplementary Figure 4). Anti-pilus (RrgA/B) titers did not vary across pediatric ages; however, adults (18+ years) had significantly higher anti-RrgA/B titers compared to pediatric participants (<18 years) ($p<0.001$) (Figure 2C). For PspA ($p<0.01$), StkP ($p<0.01$) and PhtD ($p<0.01$), participants 6-12 months and 12-24 months had lower titers than <6-month-old participants. Among most protein antigens, the magnitude of

titers were comparable, with the exception of PspC var-IV titers that were an order of magnitude lower across all ages, compared to other PspC variants.

Hierarchical clustering of individual sera by the patterns of scaled antibody titers identified two clusters. The larger cluster (A) included substantial numbers of participants from all age groups, while cluster B was comprised largely of participants <5 years of age (86.8%) (Supplemental Figure 5). This cluster of pediatric participants had antibody titers below the population mean for a large proportion of protein antigens, compared to pediatric participants in cluster A, suggesting this population was either comparatively unexposed to pneumococcal protein antigens, or unresponsive to them.

Analysis of individual protein sequence variation and association with antibody titers

PspC

Among PspC variants, var-I was the most prevalent (84.1% of carriage isolates) followed by var-II (12%). Var-I corresponds to Groups 2, 3 and 6 of PspC proteins based on sequence identity and structural organization, and var-II corresponds to Group 4 (Table 1) (12). The low prevalence of other PspC variants largely limited statistical comparison of anti-PspC titers to var-I and var-II. As age may mediate pneumococcal carriage, we first investigated carriage of the PspC var-I by age group and found that it did not differ significantly between pediatric and adult participants (X^2=0.03, p-value=0.87). Similarly, PspC variants II-IV did not significantly differ by age group. Anti-PspC titers did not differ significantly by pneumococcal carriage isolate variant for var-II (p=0.18), var-III (p=0.23), or var-IV (p=0.53) (Figure 3A); however, differences in anti-var-I titers among participants carrying PspC variants approached statistical significance [F(3)=2.41, p=0.07]. We collapsed non-var-I PspC variants into a single category and

compared titers between var-I and non-var-I. We found anti-var-I titers were significantly lower among individuals who went on to acquire strains possessing var-I [median log antibody titer 3.20 (var-I) vs. 3.31 (non-var-I) (p=0.019)] (Figure 3B). This correlation was found in all age groups but only significant for participants <1 year of age (Supplementary Figure 6). In a logistic regression model controlling for age, low var-I titers were significantly associated with subsequent carriage of a strain carrying the var-I variant (OR=0.36, 95% CI 0.14-0.81, p= 0.02). This suggests that low anti-var-I titers may be positively associated with acquisition of a strain possessing the PspC var-I.

PspA

Among pneumococci in this sample, we identified all three families of PspA variants described by Hollingshead and colleagues (13). Most isolates (70%) were family 1 PspA variant (Table 1). Additionally, 1% of strains (n=3) were a PspA variant that formed an out-group to Families 1-3. Anti-PspA antibody titers were assessed among individuals carrying four PspA variants (Family 1-3 and Unknown). Among participants carrying pneumococcal strains with polymorphic PspA variants, anti-PspA titers for Family 1 and 2 did not significantly differ (p=0.53 and 0.62) (Figure 3C). Additionally, carriage of PspA variants was not found to vary by age.

Type 1 Pilus (RrgA-C)

Using the presence of RrgB as a marker, we found 11% of carriage strains were piliated (Table 1). Among piliated strains (n=24), RrgB var-II was most common, followed closely by the var-III. Piliated strains were most often serotype 35B (42.9%), 7B (21.4%), or 19A (14.3%). Carriage of a piliated pneumococcus did not significantly

differ between pediatric and adult participants (X^2=0.49, p=0.49). Anti-RrgB antibody titers of var-II and var-I did not significantly differ by carriage variant (p=0.38 and 0.37) (Figure 3C). However, we found that anti-RrgB var-III titers were significantly higher (p<0.001) among participants with subsequent carriage strains possessing var-III RrgB. This is contrary to what would be expected if higher variant specific titers were protective against colonization with respective strains. Anti-RrgA titers also did not significantly differ among carriage variants (p= 0.148). Last, we assessed anti-RrgB titers among carriage variants, comparing participants carrying piliated and non-piliated strains. Anti-RrgB titers did not significantly differ among participants subsequently colonized with piliated or non-piliated strains (Supplemental Figure 7).

Other polymorphic protein antigens

For the polymorphic protein antigens NanA, SP0609, SP2194, PhtD, StkP, and StrH, where antibody titers were measured for only one variant, we compared titers among participants carrying pneumococci with heterologous antigen variants (Supplementary Figure 8). Distribution of NanA variants I-III did not differ between pediatric and adult participants (p=0.32). Anti-NanA titers varied among participants (p=0.08), with var-I titers marginally lower among participants carrying pneumococci with var-I. Interestingly, var-I was the least common, present in only 3.0% of carriage isolates. No statistically significant differences in anti-protein antibody titers were observed among the remaining polymorphic proteins. Additionally, with the exception of SP0609, for which the var-I was more prevalent among pediatric participants (p=0.001), protein antigen variant frequencies did not differ by age group.

Discussion

303 Protein-based vaccines aim to further reduce the morbidity of pneumococcal
304 disease; however, a clearer understanding of the dynamics of protein antigen immunity
305 and pneumococcal carriage is required. As with previous studies, we found that adult
306 sera possessed antibodies to multiple antigens and to multiple variants of variable
307 antigens, and antibody titers usually increased across participant age groups, with adults
308 having higher overall mean anti-protein titers (7,14). For PspA, StkP, and PhtD, a
309 significant reduction in protein antibodies was observed between infants <6-months-old
310 and children 6-24 months, likely resulting from the loss of maternally acquired antibodies
311 and slower acquisition in infancy and toddlerhood than for other antigens. While
312 increasing natural immunity with age has been shown in previous studies of
313 pneumococcal protein antigens (5,7,24), this is the first study to assess variant-specific
314 antibodies and their effect on protection against homologous pneumococcal
315 colonization. We found increased carriage of pneumococci expressing var-I PspC
316 among participants with lower anti-PspC var-I titers. However, among all other protein
317 antigens, including PspA, we found no difference in anti-protein antibody titers by
318 subsequent colonization status with those variants. Because our study was designed to
319 focus on variant-specific protection by serum antibody against colonization with a
320 particular antigenic variant, we did not assess antibody levels in individuals who were
321 subsequently not colonized with pneumococcus, so we were unable to assess the
322 absolute level of protection against colonization associated with particular levels of
323 serum antibodies. Our results suggest that whatever protection these antibodies offer, it
324 is modestly variant-specific for PspC and not measurably variant-specific for other
325 antigens.
326 PspC and PspA are structurally similar surface proteins (25) known for their
327 interaction with the host immune system (26,27), involvement in host epithelial
328 adherence (28), and high recombination rates (17,29). Host immune evasion is often

invoked to explain the apparent increased recombination rates, sequence diversity and evidence of diversifying selection in *pspA* and psp*C*; however, there is no direct experimental evidence for this (30,31). Studies so far have implicated anti-PspC in protection against colonization, while anti-PspA seems important for protection against invasive disease (14). The growing interest in both PspC and PspA as vaccine candidates underscores the need for better understanding and characterization of the nature of anti-PspA and anti-PspC immunity.

In this study, we investigated four variants of PspC, each representing one of the 11 recognized PspC groups (12). Our data suggest there may exist specificity against colonization based on exposure history, where colonization with a rare PspC variant leaves individuals susceptible to colonization with the most prevalent circulating strain. This observation is consistent with the concept of balancing selection in which the selective pressure of host immunity is sufficient to impact the frequency of specific PspC variants in the overall pneumococcal population. In general, this hypothesis is supported by the observation of high rates of recombination in the PspC locus, which may provide a way for pneumococci to obtain a PspC variant with low population level host immunity. However, these underlying population dynamics require further exploration.

In prior studies, anti-PspA titers in children reflected the PspA family to which they had been exposed (14), while adults often possessed high titers to both major families. Among participants of this study, family 2 PspA variants were found in only 27% of carriage isolates and mean anti-PspA family 2 variant titers were consistently lower among each age group. However, we found that anti-PspA titers were not predictive of the subsequently carried strain. In the context of previous studies demonstrating the protective role of anti-PspA antibodies, it appears that the cross-reactivity of antibodies may provide a broad level of protection from all PspA families.

354 The pneumococcal pilus, comprised of proteins RrgA, RrgB and RrgC, facilitates
355 binding to lung epithelial cells and colonization (32) and has been explored as a potential
356 vaccine candidate (33). The pilus operon has been shown to recombine and be acted
357 on by positive selection (34) Immunization with pilus subunits has been shown to
358 provide protection against systemic challenge with piliated pneumococcal strains (35).
359 Previously estimated to be present in ~30% of pneumococci (36), pilus was present in
360 only 11% of our studied strains. We found that while anti-pilus antibody titers were
361 significantly higher among adults compared with children, there was no difference in the
362 proportion of carried strains that were piliated between these age groups. Furthermore,
363 anti-pilus titers were not predictive of either carriage of piliated strains or RrgB variant
364 among carried piliated strains. This was unexpected considering the high anti-RrgB
365 titers in adults and the elevated RrgB var-III titers compared to variants I and II, which
366 were less prevalent among piliated strains (Figure 2C). The low prevalence of piliated
367 strains in the pneumococcal isolates from this study or variations in pilus expression (37)
368 may have reduced our ability to detect differential pneumococcal carriage according to
369 antibody titers.
370 While for the majority of protein antigen variants we failed to find any variant-
371 specific protective effect against carriage, it should be noted that immunity generated by
372 naturally acquired antibodies is likely more complex than a simple variant-specific
373 protective effect. Several prior studies have suggested some degree of cross-reactivity
374 between antibodies to one variant of a protein and pneumococci carrying another. This
375 could be consistent with subtle selective pressures imposed by modestly greater
376 protection against homologous than heterologous variants, which may be hard to
377 measure in a host population. Furthermore immune responses against different proteins
378 may have cumulative effects, such that the variant-specificity of antibodies to a particular
379 protein would be obscured when considering the ability of these antibodies to protect

against a strain to which the individual may have many other effective antiprotein antibody responses. Indeed, challenge studies of protein-based vaccines in animal models have clearly shown that combination vaccines including 2-3 pneumococcal proteins are more efficacious for protection against invasive challenge (1,2). This likely reflects the *in vivo* interaction of host immunity with multiple pneumococcal antigens; however, this interaction is difficult to investigate. We observed a correlation of anti-protein titers among several antigen variants, suggesting that individuals likely possess a repertoire of antibodies to a specific set of protein antigen variants.

Certain limitations should be considered when interpreting our results. We selected genetic variants of protein antigens based on extant literature and phylogenetic analysis, identifying one strain from each monophyletic clade to represent a putatively antigenically distinct variant. However, each of these protein variants possesses several distinct epitopes, each likely generating a different antibody response. It is conceivable that antibodies against one protein antigen are in fact a mixture of antibodies of varying specificity, complicating the characterization of the overall response against one antigen. This is certainly a confounding factor in our analysis and future studies should assess pneumococcal anti-protein antibody repertoires and functionally characterize distinct protein antigen variants (e.g. PspC) to determine the nature of the antibody response they elicit and further our understanding of protection while informing vaccine development. Secondly, the exposure histories of participants are unknown, and participants were seemingly less likely to be exposed to rare protein variants in the pneumococcal population. Overall, using a large sample pneumococcal strains and sera from those who carried these strains we found only modest evidence for variant-specificity of protection against pneumococcal carriage.

Conflict of Interest

The authors disclose no conflicts of interest.

Funding Sources

This study was supported by R01 R01AI048935, the Grand Challenges in Global Health initiative through the Bill & Melinda Gates Foundation, the Native American Research Centers for Health (U26IHS300013/03), the Centers for Disease Control and Prevention National Vaccine Program Office, and the Thrasher Research Fund (02820-9).

Correspondence and requests for reprints should be addressed to Taj Azarian

tazarian@hsph.harvard.edu

References

1. Darrieux M, Goulart C, Briles D, Leite LC de C. Current status and perspectives on protein-based pneumococcal vaccines. Crit Rev Microbiol. Informa Healthcare; **2015**2; 41(2):190–200.
2. Cao J, Chen D, Xu W, et al. Enhanced protection against pneumococcal infection elicited by immunization with the combination of PspA, PspC, and ClpP. Vaccine. Elsevier; **2007**; 25(27):4996–5005.
3. Lebon A, Verkaik NJ, Labout JAM, et al. Natural antibodies against several pneumococcal virulence proteins in children during the pre-pneumococcal-vaccine era: the generation R study. Infect Immun. **2011**; 79(4):1680–7.
4. McCool TL, Cate TR, Tuomanen EI, Adrian P, Mitchell TJ, Weiser JN. Serum Immunoglobulin G Response to Candidate Vaccine Antigens during Experimental Human Pneumococcal Colonization. Infect Immun. **2003**; 71(10):5724–5732.
5. Rapola S, Jäntti V, Haikala R, et al. Natural development of antibodies to pneumococcal surface protein A, pneumococcal surface adhesin A, and pneumolysin in relation to pneumococcal carriage and acute otitis media. J Infect Dis. **2000**; 182(4):1146–52.
6. Holmlund E, Quiambao B, Ollgren J, et al. Antibodies to pneumococcal proteins PhtD, CbpA, and LytC in Filipino pregnant women and their infants in relation to pneumococcal carriage. Clin Vaccine Immunol. Am Soc Microbiol; **2009**; 16(6):916–923.
7. Turner P, Turner C, Green N, et al. Serum antibody responses to pneumococcal colonization in the first 2 years of life: results from an SE Asian longitudinal cohort study. Clin Microbiol Infect. **2013**; 19(12):E551–8.
8. Simell B, Melin M, Lahdenkari M, et al. Antibodies to pneumococcal surface protein A families 1 and 2 in serum and saliva of children and the risk of

pneumococcal acute otitis media. J Infect Dis. **2007**; 196(10):1528–36.

9. Virolainen A, Russell W, Crain MJ, Rapola S, Käyhty H, Briles DE. Human antibodies to pneumococcal surface protein A in health and disease. Pediatr Infect Dis J. LWW; **2000**; 19(2):134.

10. Weinberger DM, Dagan R, Givon-Lavi N, Regev-Yochay G, Malley R, Lipsitch M. Epidemiologic evidence for serotype-specific acquired immunity to pneumococcal carriage. J Infect Dis. **2008**; 197(11):1511–8.

11. Trzcinski K, Thompson C, Malley R, Lipsitch M. Antibodies to Conserved Pneumococcal Antigens Correlate with, but Are Not Required for, Protection against Pneumococcal Colonization Induced by Prior Exposure in a Mouse Model. Infect Immun. American Society for Microbiology; **2005**; 73(10):7043–7046.

12. Iannelli F, Oggioni MR, Pozzi G. Allelic variation in the highly polymorphic locus pspC of *Streptococcus pneumoniae*. Gene. **2002**; 284(1-2):63–71.

13. Hollingshead SK, Becker R, Briles DE. Diversity of PspA: Mosaic Genes and Evidence for Past Recombination in *Streptococcus pneumoniae*. Infect Immun. **2000**; 68(10):5889–5900.

14. Melin MM, Hollingshead SK, Briles DE, Lahdenkari MI, Kilpi TM, Käyhty HM. Development of antibodies to PspA families 1 and 2 in children after exposure to *Streptococcus pneumoniae*. Clin Vaccine Immunol. **2008**; 15(10):1529–35.

15. Scott JR, Millar E V, Lipsitch M, et al. Impact of more than a decade of pneumococcal conjugate vaccine use on carriage and invasive potential in Native American communities. J Infect Dis. **2012**; 205(2):280–8.

16. Millar E V, O'Brien KL, Zell ER, Bronsdon MA, Reid R, Santosham M. Nasopharyngeal carriage of *Streptococcus pneumoniae* in Navajo and White Mountain Apache children before the introduction of pneumococcal conjugate vaccine. Pediatr Infect Dis J. **2009**; 28(8):711–6.

17. Croucher NJ, Finkelstein JA, Pelton SI, et al. Population genomics of post-vaccine changes in pneumococcal epidemiology. Nat Genet. **2013**; 45(6):656–63.
18. Bentley SD, Aanensen DM, Mavroidi A, et al. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. PLoS Genet. Public Library of Science; **2006**; 2(3):e31.
19. Inouye M, Dashnow H, Raven L-A, et al. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. Genome Med. BioMed Central; **2014**; 6(11):90.
20. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **2008**; 18(5):821–9.
21. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. **2014**; 30(14):2068–9.
22. Page AJ, Cummins CA, Hunt M, et al. Roary: Rapid large-scale prokaryote pan genome analysis. Bioinformatics. **2015** [cited 2015 Jul 22]; 31(22):btv421.
23. Löytynoja A. Phylogeny-aware alignment with PRANK. Methods Mol Biol. **2014**; 1079:155–70.
24. Simell B, Lahdenkari M, Reunanen A, Käyhty H, Väkeväinen M. Effects of ageing and gender on naturally acquired antibodies to pneumococcal capsular polysaccharides and virulence-associated proteins. Clin Vaccine Immunol. **2008**; 15(9):1391–7.
25. Brooks-Walter A, Briles DE, Hollingshead SK. The pspC Gene of *Streptococcus pneumoniae* Encodes a Polymorphic Protein, PspC, Which Elicits Cross-Reactive Antibodies to PspA and Provides Immunity to Pneumococcal Bacteremia. Infect Immun. **1999**; 67(12):6533–6542.
26. Mukerji R, Mirza S, Roche AM, et al. Pneumococcal surface protein A inhibits complement deposition on the pneumococcal surface by competing with the

binding of C-reactive protein to cell-surface phosphocholine. J Immunol. Am Assoc Immnol; **2012**; 189(11):5327–5335.

27. Ren B, McCrory MA, Pass C, et al. The Virulence Function of *Streptococcus pneumoniae* Surface Protein A Involves Inhibition of Complement Activation and Impairment of Complement Receptor-Mediated Protection. J Immunol. **2004**; 173(12):7506–7512.

28. Vadesilho CFM, Ferreira DM, Gordon SB, et al. Mapping of epitopes recognized by antibodies induced by immunization of mice with PspA and PspC. Clin Vaccine Immunol. **2014**; 21(7):940–8.

29. Chewapreecha C, Harris SR, Croucher NJ, et al. Dense genomic sampling identifies highways of pneumococcal recombination. Nat Genet. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; **2014**; 46(3):305–9.

30. Chaguza C, Cornick JE, Everett DB. Mechanisms and impact of genetic recombination in the evolution of *Streptococcus pneumoniae*. Comput Struct Biotechnol J. **2015**; 13:241–7.

31. Li Y, Gierahn T, Thompson CM, et al. Distinct effects on diversifying selection by two mechanisms of immunity against *Streptococcus pneumoniae*. PLoS Pathog. Public Library of Science; **2012**; 8(11):e1002989.

32. Barocchi MA, Ries J, Zogaj X, et al. A pneumococcal pilus influences virulence and host inflammatory responses. Proc Natl Acad Sci U S A. National Acad Sciences; **2006**; 103(8):2857–2862.

33. Spraggon G, Koesema E, Scarselli M, et al. Supramolecular organization of the repetitive backbone unit of the *Streptococcus pneumoniae* pilus. PLoS One. Public Library of Science; **2010**; 5(6):e10919.

34. Muzzi A, Moschioni M, Covacci A, Rappuoli R, Donati C. Pilus operon evolution in

Streptococcus pneumoniae is driven by positive selection and recombination. PLoS One. **2008**; 3(11):e3660.

35. Harfouche C, Filippini S, Gianfaldoni C, et al. RrgB321, a fusion protein of the three variants of the pneumococcal pilus backbone RrgB, is protective in vivo and elicits opsonic antibodies. Infect Immun. Am Soc Microbiol; **2012**; 80(1):451–460.

36. Paterson GK, Mitchell TJ. The role of *Streptococcus pneumoniae* sortase A in colonisation and pathogenesis. Microbes Infect. Elsevier; **2006**; 8(1):145–153.

37. Basset A, Turner KH, Boush E, Sayeed S, Dove SL, Malley R. Expression of the type 1 pneumococcal pilus is bistable and negatively regulated by the structural component RrgA. Infect Immun. Am Soc Microbiol; **2011**; 79(8):2974–2983.

Figures and Tables

Table 1. Pneumococcal serotypes by age group and ordered by prevalence.

	Age Group									
Serotype	**<1**		**1-4 y**		**5-17y**		**18y+**		**Total**	**% of Total**
19A	5	2.07%	7	2.89%	0	0.00%	10	4.13%	22	9.07%
22A	4	1.65%	10	4.13%	4	1.65%	3	1.24%	21	8.67%
35B	6	2.48%	6	2.48%	3	1.24%	6	2.48%	21	8.67%
23B	2	0.83%	8	3.31%	3	1.24%	4	1.65%	17	7.02%
15A	1	0.41%	7	2.89%	2	0.83%	3	1.24%	13	5.37%
34	0	0.00%	6	2.48%	1	0.41%	6	2.48%	13	5.36%
6B	1	0.41%	4	1.65%	2	0.83%	5	2.07%	12	4.95%
11D	2	0.83%	3	1.24%	4	1.65%	2	0.83%	11	4.54%
23A	1	0.41%	4	1.65%	1	0.41%	4	1.65%	10	4.13%
31	2	0.83%	3	1.24%	3	1.24%	0	0.00%	8	3.31%
16F	0	0.00%	2	0.83%	4	1.65%	2	0.83%	8	3.30%
17F	2	0.83%	0	0.00%	2	0.83%	4	1.65%	8	3.30%
37	2	0.83%	0	0.00%	1	0.41%	4	1.65%	7	2.89%
6A	0	0.00%	3	1.24%	3	1.24%	0	0.00%	6	2.48%
21	0	0.00%	3	1.24%	2	0.83%	1	0.41%	6	2.48%
35F	2	0.83%	2	0.83%	0	0.00%	2	0.83%	6	2.48%
7B	1	0.41%	1	0.41%	0	0.00%	4	1.65%	6	2.47%
10A	0	0.00%	3	1.24%	1	0.41%	1	0.41%	5	2.06%
15C	1	0.41%	1	0.41%	0	0.00%	3	1.24%	5	2.06%
3	0	0.00%	1	0.41%	2	0.83%	1	0.41%	4	1.65%
35A	0	0.00%	3	1.24%	0	0.00%	1	0.41%	4	1.65%
NT	2	0.83%	1	0.41%	0	0.00%	1	0.41%	4	1.65%
10B	0	0.00%	1	0.41%	2	0.83%	0	0.00%	3	1.24%
7C	0	0.00%	1	0.41%	1	0.41%	1	0.41%	3	1.24%
1	0	0.00%	1	0.41%	0	0.00%	2	0.83%	3	1.24%
19F	1	0.41%	0	0.00%	1	0.41%	0	0.00%	2	0.83%
22F	0	0.00%	1	0.41%	1	0.41%	0	0.00%	2	0.83%
38	0	0.00%	1	0.41%	1	0.41%	0	0.00%	2	0.83%
12A	0	0.00%	1	0.41%	0	0.00%	1	0.41%	2	0.82%
12B	0	0.00%	1	0.41%	0	0.00%	1	0.41%	2	0.82%
15B	0	0.00%	0	0.00%	1	0.41%	0	0.00%	1	0.41%
5	0	0.00%	0	0.00%	1	0.41%	0	0.00%	1	0.41%
7A	1	0.41%	0	0.00%	0	0.00%	0	0.00%	1	0.41%
8	0	0.00%	1	0.41%	0	0.00%	0	0.00%	1	0.41%
9A	0	0.00%	1	0.41%	0	0.00%	0	0.00%	1	0.41%
10F	0	0.00%	0	0.00%	0	0.00%	1	0.41%	1	0.41%
Total	36	14.9%	87	36.0%	46.0	19%	73	30.2%	242	100.00%

Table 2. Protein antigens and variant frequencies identified through genomic analysis of pneumococcal carriage isolates. The antigen name and function are listed with the number of variants tested. Variants with measured titers are specified with an asterisk.

Polymorphic								
Antigen			**Variant**	**Strain**	**Accesion/Gene**	**Count**	**Freq**	**P-Distance**
I.	PspC	Pneumococcal surface protein C	Var-I*	ND6053	ERR129207	203	83.9%	0.025 (0.025-0.025)
			Var-II*	CH2016	ERR129074	31	12.8%	0.007 (0.006-0.008)
			Var-III*	BR1086	ERR129054	3	1.2%	0.005 (0.001-0.009)
			Var-IV*	MD5090	ERR129180	1	0.4%	-
II.	PspA	Pneumococcal surface protein A	Family 1*	D39	SPD_0126	166	68.6%	0.200 (0.199-0.201)
			Family 2*	TIGR4	SP_0117	69	28.5%	
			Family 3	BG6380	AF071823	4	1.7%	
III.	RrgA	RrgA pilus subunit, adhesin	Var-I*	TIGR4	SP_0462	24	88.9%	0.032 (0.027-0.037)
			Var-II	670-6B	SP670_0540	3	11.1%	
IV.	RrgB	RrgB pilus subunit, backbone	Var-I*	670-6B	SP670_0541	4	13.8%	0.216 (0.199-0.233)
			Var-II*	TIGR4	SP_0463	11	37.9%	
			Var-III*	23F_Taiwan_15	EF560629: 5159-7123	14	48.3%	
V.	NanA	Neuraminidase	Var-I*	D39	SPD_1504	7	3.0%	0.042 (0.041-0.042)
			Var-II	INV200	SPNINV200_15140	75	32.2%	
			Var-III	ATCC 700669	SPN23F16920	151	64.8%	
VI.	SP0609	Amino acid ABC transporter	Var-I*	TIGR4	SPD_0530	161	66.8%	0.015 (0.015-0.015)
			Var-II	ATCC 700669	SPN23F05490	80	33.2%	
VII.	SP2194	ATP-dependent Clp protease	Var-I	Taiwan19F-14	SPT_2213	222	91.7%	0.010 (0.010-0.010)
			Var-II*	TIGR4	SP_2194	20	8.3%	
VIII.	PhtD	Pneumococcal histidine triad D	Var-I	ATCC 700669	SPN23F09290	168	70.9%	0.062 (0.060-0.064)
			Var-II*	D39	SPD_0889	69	29.1%	
IX.	StkP	Serine threonine kinase protein	Var-I	D39	SPD_1542	181	74.8%	0.007 (0.007-0.007)

			Var-II*	TIGR4	SP_1732	61	25.2%	
X.	StrH	Beta-N-acetylhexosaminidase	Var-I*	D39	SPD_0063	101	42.1%	0.007 (0.007-0.007)
			Var-II	TIGR4	SP_0057	139	57.9%	
XI.	Ply	Pneumolysin	Var-I*	D39	SPD_1726	29	12.0%	0.003 (0.003-0.003)
			Var-II*	TIGR4	SP_1923	213	88.0%	
Conserved								
Antigen			**Variant**	**Strain**	**Accession/Gene**	**-**	**-**	
XII.	LysM	LysM domain-containing protein	Var-I*	TIGR4	SP_0107	-	-	0.004 (0.003-0.004)
XIII.	LytB	Endo-beta-N-acetylglucosaminidase	Var-I*	D39	SPD_0853	-	-	0.008 (0.008-0.008)
XIV.	LytC	Lysozyme (C-ter)	Var-I*	D39	SPD_1403	-	-	0.005 (0.005-0.005)
XV.	PcpA	Choline binding protein	Var-I*	D39	SPD_1965	-	-	0.003 (0.003-0.003)
XVI.	PcsB	Secreted 45 kDa protein	Var-I*	TIGR4	SP_2216	-	-	0.006 (0.006-0.007)
XVII.	PhtE	Truncated histidine triad protein	Var-I*	D39	SPD_0890	-	-	0.008 (0.008-0.008)
XVIII.	PiaA	Part of iron uptake ABC transporter	Var-I*	D39	SPD_0915	-	-	0.002 (0.002-0.002
XIX.	PiuA	Part of iron uptake ABC transporter	Var-I*	D39	SPD_1652	-	-	0.005 (0.005-0.005)
XX.	PsaA	Pneumococcal surface adhesin A	Var-I*	TIGR4	SP_1650	-	-	0.003 (0.002-0.003)
XXI.	SP2027	Conserved hypothetical protein	Var-I*	TIGR4	SP_2027	-	-	0.005 (0.005-0.006)

Figure 1. Pearson correlations of log10 antibody titers for 28 protein antigens clustered heuristically by correlation value. Correlations between normalized antibody titers of protein antigens were clustered using heuristic methods. Protein antigens including multiple variants of the same antigen are labeled on the x- and y- axes, and the heatmap displays the correlation values between antigens. The dendrogram on the left represents the results of the heuristic clustering of correlated antibody titers. Significant correlation between variants of the same antigen was observed as well as high correlation among several antigens, which likely exist on the same genomic background.

Figure 2. Protein antibodies titers by age for variants of PspC, PspA, pilus, and ply. Antibody levels were measured using direct binding electrochemiluminescence-based multiplex assay are expressed as a titer relative to the amount in a reference serum. Structural variants of polymorphic proteins were measured individually and were compared among participant age groups. A.) Variant-specific anti-PspC antibodies to Var-I (ND6053), Var-II (CH2016), Var-III (BR1086), and Var-IV (MD5090). B.) Variant-specific anti-PspA antibodies to Family 1 and Family 2. C.) Anti-pilus antibodies to RrgA-I (TIGR4) and RrgB pilus variants RrgB-I (670-6B), RrgB-II (Taiwan 23F), and RrgB-III (TIGR4). D.) Variant-specific anti-pneumolysin (ply) antibodies to variants I and II.

Figure 3. Protein antibody titers among participants carrying pneumococcal strains with specific polymorphic protein-antigen variants. Serum was collected from participants at enrollment, and nasopharyngeal swabs for pneumococcal carriage detection were collected at least 30 days after serum collection. The protein antigen variant in each pneumococcal isolate was determined using genomic data. Antibody levels were

measured using direct binding electrochemiluminescence-based multiplex assay are expressed as a titer relative to the amount in a reference serum. Structural variants of polymorphic proteins were measured individually. The y-axis represents the variant-specific antibody titers, and the carriage isolate protein antigen variant is specified on the x-axis with labels colored to match the corresponding titers. If susceptibility to a strain possessing a specific variant were observed, the respective antibody titer would be the lowest among all other titers. A) Anti-PspC titers vs. carriage isolate PspC variants I-IV and non-typable. B) Anti-PspC titers vs. carriage isolate PspC var-I and combined var-II-IV. C) Anti-PspA Family 1 and Family 2 titers vs. carriage isolate PspA variants Families 1, 2, 3, and unknown. D) Anti-pilus titers vs. carriage isolate RrgB variants I-III and not present.