

# Comparative Genomic Analysis of Meningitis- and Bacteremia-Causing Pneumococci Identifies a Common Core Genome

Benard W. Kulohoma,<sup>a,b\*</sup> Jennifer E. Cornick,<sup>a,b</sup> Chrispin Chaguzo,<sup>a,b</sup> Feyruz Yalcin,<sup>c</sup> Simon R. Harris,<sup>c</sup> Katherine J. Gray,<sup>d</sup> Anmol M. Kiran,<sup>a,b</sup> Elizabeth Molyneux,<sup>d</sup> Neil French,<sup>b</sup>  Julian Parkhill,<sup>c</sup> Brian E. Faragher,<sup>e</sup> Dean B. Everett,<sup>a,b</sup> Stephen D. Bentley,<sup>b,c</sup>  Robert S. Heyderman<sup>a\*</sup>

Malawi-Liverpool-Wellcome Trust Clinical Research Programme, University of Malawi College of Medicine, Blantyre, Malawi<sup>a</sup>; Institute of Infection and Global Health, University of Liverpool, Liverpool, United Kingdom<sup>b</sup>; The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom<sup>c</sup>; University of Malawi College of Medicine, Blantyre, Malawi<sup>d</sup>; Liverpool School of Tropical Medicine, Liverpool, United Kingdom<sup>e</sup>

*Streptococcus pneumoniae* is a nasopharyngeal commensal that occasionally invades normally sterile sites to cause bloodstream infection and meningitis. Although the pneumococcal population structure and evolutionary genetics are well defined, it is not clear whether pneumococci that cause meningitis are genetically distinct from those that do not. Here, we used whole-genome sequencing of 140 isolates of *S. pneumoniae* recovered from bloodstream infection ( $n = 70$ ) and meningitis ( $n = 70$ ) to compare their genetic contents. By fitting a double-exponential decaying-function model, we show that these isolates share a core of 1,427 genes (95% confidence interval [CI], 1,425 to 1,435 genes) and that there is no difference in the core genome or accessory gene content from these disease manifestations. Gene presence/absence alone therefore does not explain the virulence behavior of pneumococci that reach the meninges. Our analysis, however, supports the requirement of a range of previously described virulence factors and vaccine candidates for both meningitis- and bacteremia-causing pneumococci. This high-resolution view suggests that, despite considerable competency for genetic exchange, all pneumococci are under considerable pressure to retain key components advantageous for colonization and transmission and that these components are essential for access to and survival in sterile sites.

*Streptococcus pneumoniae* is estimated to cause almost 15 million episodes of pneumonia, meningitis, and bloodstream infection in children under five annually, with the highest burden in sub-Saharan Africa (SSA) (1). An important HIV-related pathogen, the pneumococcus is also associated with a significant number of deaths in HIV-infected African adults (2). Nasopharyngeal colonization is a prerequisite for pneumococcal disease, which follows a complex multistep process of bloodstream and less frequently cerebrospinal fluid (CSF) invasion through a combination of engagement between bacteria and host epithelial and endothelial receptors, cytoskeletal rearrangement and transcytosis, and infection of migrating phagocytes (3–6). The avoidance of immune surveillance and the capacity of *S. pneumoniae* to grow and survive in the relatively hostile blood and meningeal compartments facilitate these processes (3, 7). Why some patients develop pneumococcal bacteremia but not meningitis while in others infections occur in both compartments is uncertain.

Our fundamental understanding of the bacterial mechanisms that determine these host-pathogen interactions is at best incomplete and is supported largely by data from tissue culture models and animal studies (4, 5, 8). In particular, specific pneumococcal determinants that explain the tropism of this capsulated bacterium for the brain have been elusive. Whether pneumococci that cause meningitis are genetically distinct from those that do not is uncertain. While some pneumococcal capsule serotypes are particularly associated with invasive pneumococcal diseases (IPD) in humans (9, 10), each with a different invasive potential, there is currently no evidence that either the serotypes or sequence types (STs) that cause bacteremia and meningitis are distinct (11). Likewise, numerous other bacterial molecules have been implicated in the pathogenesis of pneumococcal meningitis (e.g., phosphorylcholine, pneumolysin, pneumococcal surface adhesin A, neur-

amidase A, and pneumococcal adherence and virulence factor A), but whether these are absolutely required for meningeal invasion and disease causation remains to be determined. Adaptation of the pneumococcus to the host environment is facilitated by mutations and by frequent transfers of genetic material (horizontal gene transfer) between isolates and across bacterial species (12–14). This genetic variability provides considerable redundancy in the range of tools available to *S. pneumoniae* to target host receptors, overcome the mucosal barrier, and survive within the nasopharynx (7, 15–17).

Received 29 June 2015 Accepted 28 July 2015

Accepted manuscript posted online 10 August 2015

Citation Kulohoma BW, Cornick JE, Chaguzo C, Yalcin F, Harris SR, Gray KJ, Kiran AM, Molyneux E, French N, Parkhill J, Faragher BE, Everett DB, Bentley SD, Heyderman RS. 2015. Comparative genomic analysis of meningitis- and bacteremia-causing pneumococci identifies a common core genome. *Infect Immun* 83:4165–4173. doi:10.1128/IAI.00814-15.

Editor: A. Camilli

Address correspondence to Benard W. Kulohoma, bkulohoma@iscb.org, or Robert S. Heyderman, rheyderman@ucl.ac.uk.

\* Present address: Benard W. Kulohoma, Centre for Biotechnology and Bioinformatics, University of Nairobi, and International Center for Insect Physiology and Ecology, Nairobi, Kenya; Robert S. Heyderman, Division of Infection and Immunity, University College London, London, United Kingdom. B.W.K. and J.E.C. contributed equally to this work.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/IAI.00814-15>.

Copyright © 2015 Kulohoma et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](http://creativecommons.org/licenses/by/3.0/). doi:10.1128/IAI.00814-15

Here, we have used whole-genome sequencing of a large collection of disease-causing isolates to revisit pneumococcal meningitis in humans. We have specifically identified genes that are universally present in pneumococci associated with invasive disease (the pneumococcal core genome) and those that are present only in some invasive isolates (the pneumococcal accessory genome) and have investigated the hypothesis that meningitis-causing isolates have a genetic content distinct from isolates associated with bloodstream infection, which explains their ability to survive in these hostile host environments.

## MATERIALS AND METHODS

**Strain selection, DNA extraction, and genome sequencing.** *S. pneumoniae* isolates selected for genome sequencing (see Table S1 in the supplemental material) were obtained from the Malawi-Liverpool-Wellcome Trust Clinical Research Programme (MLW) strain archive. Isolates collected between 2002 and 2008 ( $n = 140$ ) were randomly selected from the archive isolate database using STATA v 9.2 (StataCorp, College Station, TX). MLW has conducted externally quality-controlled surveillance for bacterial meningitis and bloodstream infection at Queen Elizabeth Central Hospital, Blantyre, Malawi, for over 15 years (1–3). The inclusion criteria, culture rate, and laboratory methods have not changed appreciably over this period. Pneumococcal isolates were obtained from the CSF of children and adults where meningitis was suspected ( $n = 70$ ) and blood of children and adults with suspected bloodstream infection but not meningitis ( $n = 70$ ) (1, 3). Children were defined as patients <15 years of age. HIV status was not available for the isolates. Too few paired isolates for a robust analysis were available and therefore were not used. Likewise, too few noninvasive isolates from Malawi were available to allow a robust comparison of noninvasive to invasive (blood/CSF) isolates; therefore noninvasive isolates were not included in this analysis.

Isolates were grown overnight in Todd-Hewitt broth (Oxoid limited, Basingstoke, Hampshire, England), and genomic DNA was isolated from the bacteria using a Wizard genomic DNA purification kit (Promega, USA; product number A1125). Multiplex library construction, with a 200-bp insertion size and 54-nucleotide (nt) or 108-nt paired-end Illumina sequencing, followed by a third 7-nt tag read, was performed according to standard protocols (13, 18). Serotype and sequence type data were derived from genome sequences using previously described protocols (13). The isolates used in the study were anonymized. These data are published with the approval of the University of Malawi College of Medicine Research & Ethics Committee and conform to institutional guidelines.

**Assembly of draft genomes and annotation.** *De novo* assembly of draft genome sequences was achieved from multiplexed Illumina sequence data as previously described (13, 18). Briefly, assemblies were made using Velvet v 0.7.03 (19). Velvet was optimized to run with the longest  $k$ -mer that gave an expected coverage value above 20 (13). The draft genome assemblies were ordered against the *S. pneumoniae* ATCC 700669 genome (see Table S2 in the supplemental material) using ABACAS (20) and ACT (21). Coding sequences (CDS) were identified using Glimmer3 (22).

**Genome clustering and core genome analysis.** Gene annotations from 12 publicly available, complete, fully annotated reference genomes (see Table S2 in the supplemental material) were transferred onto the CDS in the draft genomes using RATT (23). Annotated genes were translated to protein sequences and assigned to orthologous gene “clusters” using tribeMCL (24) with default parameters (BLASTP E value cutoff of  $1e-5$  and inflation index of 4) (full details of the clustering pipeline can be found at <https://github.com/fy2/pneumoscript>). The OrthoMCL output was used to establish the core genome for the data sets.

The tribeMCL ortholog clusters were organized into a matrix of genome content using bespoke Perl scripts, with orthologs from the same genome arranged in columns and rows identifying annotated orthologs with similar functions. A blank was inserted where an ortholog was ab-

sent. Genomes (the matrix columns) were randomly sampled in an arithmetic progression fashion according to the equation  $S_N = N/2[2a + (N - 1)d]$ . The number of random sampling events,  $S_N$ , was established using the least number of genomes under consideration,  $a$  (i.e., 1 genome in this study), the total number of genomes under consideration (i.e., the data set size),  $N$ , and the common difference of successive genomes,  $d$  (i.e., 1) to be used during sampling. During random sampling,  $N$  was initialized as 1 genome for the first event and increased by one unit for each of the subsequent events, until it was equivalent to the total number of genomes in the data set in the final event. The random genomes were sampled only once during each event, and the total orthologs shared by all genomes were counted once for each cluster, thereby excluding paralogous counts. This was iterated 100 times, resulting in 100 input orders for each event; the arithmetic average core genome size was computed for each event to enable the average core genome size to be related to the number of genomes sampled. Comparisons of core genomes from different data sets were achieved using Perl scripts and STATA v 9.2 (Stata Corp, College Station, TX).

**Core genome extrapolation.** The absolute core genome size was extrapolated by fitting exponentially decaying functions as previously described by Tettelin et al. (29). For our data, two distinct decay phases were identified, so the following double-exponential model was fitted to the data using the Levenberg-Marquardt iterative estimation method:  $y = kc_1 \cdot \exp(-n/tc_1) + kc_2 \cdot \exp(-n/tc_2) + \text{gamma}$ . In this model,  $y$  is the size of the core genome,  $n$  is the number of sequenced strains,  $kc_1$  and  $tc_1$  are the exponent and free parameters, respectively, for the first decay phase,  $kc_2$  and  $tc_2$  are the exponent and free parameters, respectively, for the second decay phase, and gamma is the predicted value of the core genome size with an infinite number of sequenced strains.

## RESULTS

**The pneumococcal core genome.** We first set out to define the pneumococcal core genome, i.e., those genes that are common to all human invasive pneumococcal disease isolates. Hiller and colleagues have previously undertaken a similar analysis using a data set of 17 pneumococcal genomes (eight from nasopharyngeal isolates from pediatric patients and nine invasive and avirulent isolates that were publically available). They found that 1,454 gene clusters (46%) were conserved among all isolates (25). Here, we have subjected to Illumina sequencing 140 randomly selected invasive pneumococcal isolates, which originate from Malawian adults and children with meningitis ( $n = 70$ ) or bloodstream infection without clinical meningitis (bacteremia;  $n = 70$ ). The HIV status of the individuals from which the isolates were recovered was not available. These isolates were collected as part of routine invasive bacterial surveillance at a large teaching hospital in Malawi (2, 26, 27). The 140 genomes had a mean sequencing read depth of  $155\times$  (minimum depth of  $38\times$ ) (full details of assembly statistics are listed in Table S4 in the supplemental material). The isolates belong to different serotypes and clonal types as determined by multilocus sequence typing (MLST) (see Table S1 in the supplemental material) and consequently vary in their previously reported association with invasive disease (28). The genomes were first *de novo* assembled using Velvet v 0.7.03 (19); prediction of coding sequences and functional annotation were done using Glimmer3 and RATT (22, 23). A mapping approach, which involves the alignment of query genome sequence reads against a single reference genome, would not have been suitable for this analysis, as mapping highly recombinogenic clones from multiple diverse lineages against a single reference genome would considerably restrict the degree of variation detected.

We identified orthologous genes in the assembled genome by

the process of clustering based on homology. The clustering of orthologous genes provided a robust basis for comparing and combining data from multiple diverse genomes. Following clustering, genes from all isolates were divided into three categories: orthologous genes shared by all isolates (core genes), orthologous genes shared between two or more isolates (accessory genes), and genes unique to one isolate. We identified 3,164 orthologous gene clusters in all 140 isolates. Forty-five percent of the total clusters (1,428 core genes) were present in all of the study isolates (Fig. 1; see Table S3 in the supplemental material). Of the remaining clusters, 1,612 (51%) were present in two or more isolates, but not all isolates, while 124 genes (4%) did not group within a cluster and were unique to a single isolate. Of the core genes, 1,239/1,428 (87%) could be assigned a definitive function; the remaining 189 core genes (13%) were either of unknown or putative function.

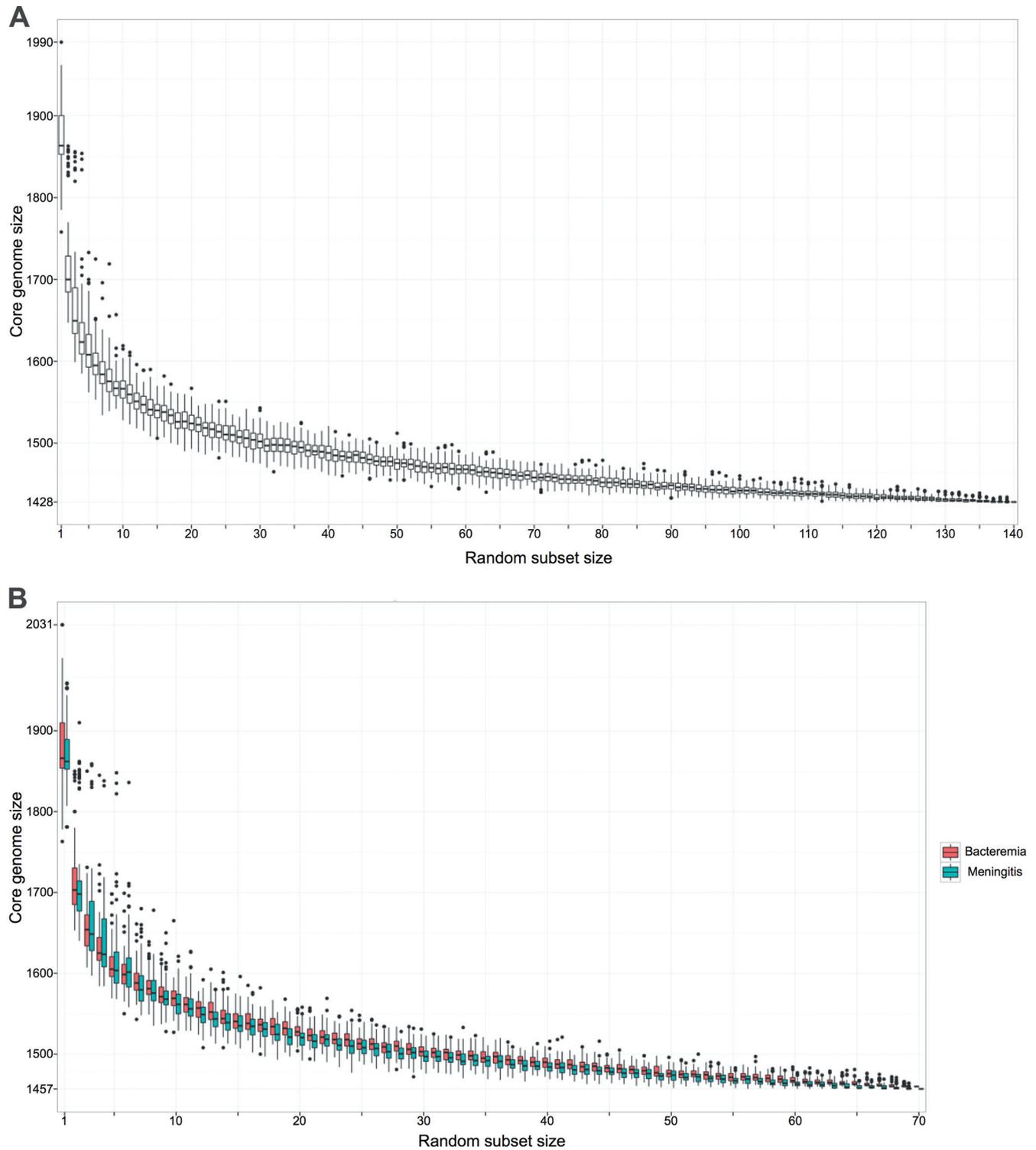
A considerable amount of care was taken to ensure that the genome assemblies were of good quality and therefore did not distort the outputs by applying a stringent quality control (QC) process. Nonetheless, poor assembly of small genomic regions may have resulted in legitimate core genes being wrongly defined as accessory genes in our analysis, thus causing us to marginally underestimate the core genome size. Relaxation of the stringency to allow for this possibility, whereby a core gene was defined as an orthologous gene shared by  $\geq 98\%$  or more isolates, marginally increased the core genome (the “extended core genome”) size to 1,469 genes. As reported previously, the core genome size was dependent on the number of isolates used during the analysis and gradually decreased with the addition of new genomes (Fig. 1A) but did not reach an absolute plateau, as has been reported in studies of group B *Streptococcus* that examined smaller data sets (29). To ascertain if we had reached a definitive core genome size for *S. pneumoniae*, we undertook a decaying-function analysis as previously described (29). However, instead of a single-exponential decaying function, we found that we could extrapolate the data until it plateaued with a higher degree of precision by fitting a double-exponential decaying function (see Fig. S1 in the supplemental material). Consistent with the core genome size observed in our analysis of 140 genomes, we found that, when extrapolated to infinity using the double-exponential decaying function, the core genome reached a minimum of 1,427 genes (95% confidence interval [CI] 1,425 to 1,435 genes;  $R^2 = 99.6$ ). It can therefore be concluded that the data set employed in this study allowed us to accurately assess the core genome content of pneumococci isolated in Malawi.

**Composition of the core genome.** The core genome common to all 140 human invasive pneumococcal disease isolates contains many genes that have been implicated in nasopharyngeal colonization by inhibiting mucus function, reducing niche competition by other bacterial species, subverting host cell function, and facilitating host cell attachment (Table 1; for a complete list of core genes see Table S3 in the supplemental material). In animal models, many of these virulence genes have also been shown to facilitate tissue invasion and promote disease. The virulence factors encoded by the overall core genome identified here include:  $\beta$ -galactosidase (BgaA) and  $\beta$ -*N*-acetylglucosaminidase (StrH), which deglycosylate mucus glycoconjugate to reduce mucus viscosity, thereby limiting mucus entrapment and exposing host cell surface adherence receptors that are exploited by the pneumococcus for attachment (3, 30), and pyruvate oxidase (SpxB), which catalyzes the production of hydrogen peroxide, which kills niche competi-

tors to facilitate pneumococcal colonization (8, 15). In view of the report that serotype 1 patient isolates from the ST217 clonal complex can have spontaneous mutations in *spxB* that seem to increase virulence (4), we investigated allelic variation in the *spxB* gene in the data set. We found the SpxB protein to be highly conserved, and no allelic variants of *spxB* were unique to either the bacteremia or the meningitis isolates.

Also identified in the overall core genome were genes encoding virulence proteins that enhance colonization either through binding host ligands or subverting immune defense, including pneumolysin (Ply) (3, 8, 17), autolysin (LytA) (17), choline-binding protein E (CbpE) (5, 16), pneumococcal adhesion and virulence A (PavA) (3, 16), and immunoglobulin A1 protease (IgA1 protease). The overall core genome contains genes that enable niche adaptation and regulation of metabolism in response to alterations in available carbon sources, cations and amino acids, and anaerobic conditions (31). These include the formate acetyltransferase gene (*pfl*), which is essential for fermentative metabolism. Pfl mutants have reduced virulence, evidenced by prolonged survival times in mouse models of bacteremia (31).  $\text{Fe}^{2+}$  ATP-binding cassette (ABC) transporters, which are members of the ABC transporter superfamily, translocating a variety of substrates such as cations, lipids, and amino acids across membranes (32), have been specifically implicated in pneumococcal virulence (33). Previous PCR-based studies have shown that several of the associated genes, while present in all of our invasive isolates, are not always present in carriage isolates (34, 35). For example, genes encoding NanB and CbpA have been found to be present in only 81% and 54% of carriage pneumococci screened, respectively (35). This implies that carriage isolates encode an alternative set of proteins that aid colonization, suggesting that there is considerable redundancy in the range of tools available to the pneumococcus for survival in the nasopharyngeal niche.

Once the pneumococcus has reached the bloodstream, invasion of the meninges is thought to require a series of genes that mediate host cell attachment and transcytosis through the blood-brain barrier (BBB) and then survival within the CSF (3, 4, 8). Since bacteremia and meningitis are a dead end for the pneumococcus, it is unlikely that these core genes have evolved solely to facilitate invasion of the brain. We speculate that these core genes are important for niche survival at the mucosal surface but also facilitate invasion into and survival in the blood and CSF. Ply disrupts BBB integrity, PavA facilitates endothelial fibronectin engagement, and Hyl degrades the endothelial extracellular matrix; in animal models all three of these functions have been strongly implicated in the pathogenesis of meningitis. We identified a number of other genes that have been specifically implicated in pneumococcal meningitis in the overall core genome. These include genes encoding choline binding protein A (CbpA, SpsA, or PspC), a member of a family of surface proteins implicated in colonization (17), which binds to the polymeric immunoglobulin receptor (pIgR) in order to cross from the mucosal surface into the blood (3, 36). In the absence of pIgR on the surface of brain endothelium, CbpA binds laminin to initiate contact and translocation across the BBB (3, 4). The core genome also encodes pneumococcal competence proteins (ComA, ComD, ComE, and ComX), which regulate biofilm formation and are associated with virulence in animal models of meningitis (30, 37). The prominence of DNA repair and regulation genes in the overall core genome identified by our analysis suggests the critical nature of gene



**FIG 1** Meningitis- and bacteremia-associated pneumococcal core genomes based on 140 invasive isolates. (A) Box and whisker plot of the number of core gene orthologous clusters observed as the subset of isolates included in the analysis increases, where the subset ranged from 1 to the total number of isolates ( $n = 140$ ). To generate a subset of isolates,  $n$  isolates are randomly selected from the data set. The core genome size is then calculated for  $n$  isolates. Each random subset of a given size is generated 100 times. (B) Box and whisker plot of the number of meningitis and bacteremia core gene orthologous clusters observed as the number of isolates included in the analysis increases, where the subset ranged from 1 to the total number of isolates ( $n = 70$ ). To generate a subset of isolates,  $n$  isolates are randomly selected from the data set. The core genome size is then calculated for  $n$  isolates. Each random subset of a given size is generated 100 times. (C) Map of the proportion of meningitis-specific core genes in the bacteremia-associated pneumococcal data set (genes present and absent are represented in gray and white, respectively). A list of meningitis-specific proteins is on the  $x$  axis, and individual bacteremia-specific isolates with a phylogenetic tree are on the  $y$  axis. (D) Map of the proportion of bacteremia-specific genes in the meningitis-associated pneumococcal data set (genes present and absent are represented in gray and white, respectively). A list of bacteremia-specific proteins is on the  $x$  axis, and individual meningitis isolates with a phylogenetic tree are on the  $y$  axis.

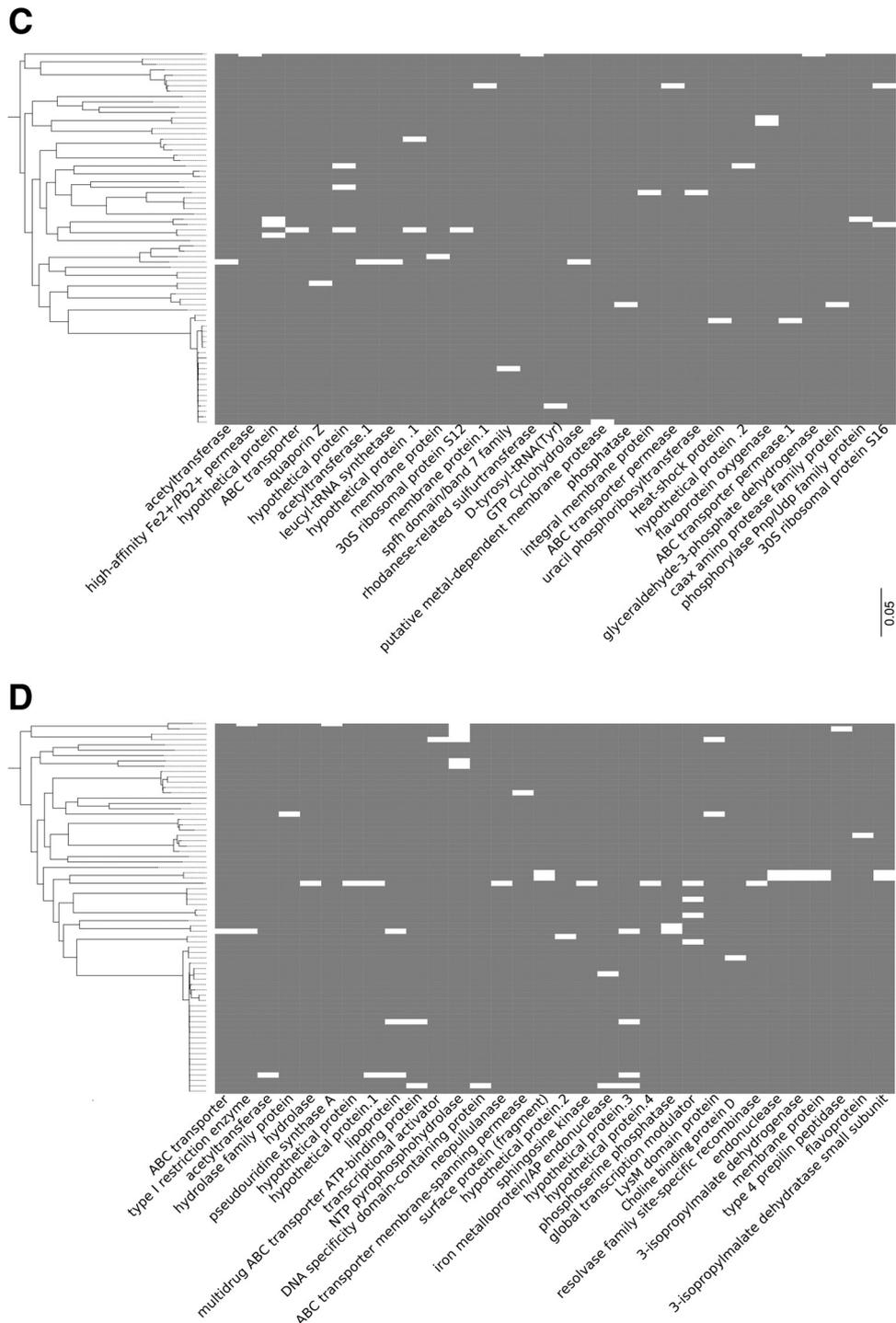


FIG 1 continued

variation for pneumococcal survival in a variety of environments (12, 38).

Some of the genes identified in the common core genome also encode putative pneumococcal protein antigen candidates currently under evaluation (Table 1), with the notable exception of pilin and NanA, whose genes were not present in the core genome (3, 8, 17, 39). Other key virulence factors whose genes were not present in the core genome include pneumococcal

histidine triad protein B (PhtB), pneumococcal histidine triad protein E (PhtE), pneumococcal surface protein A (PspA), neuraminidase A (NanA), neuraminidase C (NanC), pilus subunit A (RrgA), pilus subunit B (RrgB), pilus subunit C (RrgC), and bacteriocin (Blp). Mahdi and colleagues recently conducted a genome-wide *in vivo* transcriptomic analysis of pneumococci reaching the brain after murine intranasal infection (40). They identified a previously uncharacterized protein, alpha-glycero-

**TABLE 1** Summary of key proteins associated with the pneumococcal core genome

Category	Core gene products
Antimicrobial resistance	ABC transporter antimicrobial extrusion family protein, MATE Multidrug resistance protein, MdtG
Transporters	Manganese efflux system, SP1552 ABC transporter, BlpC Cobalt ABC transporter, CbiO
Virulence factors	$\beta$ -Galactosidase, BgaA BlpC (quorum-sensing pheromone), which controls Blp Choline binding protein A, CbpA Choline binding protein D, CbpD Choline binding protein E, CbpE Choline binding protein G, CbpG Choline binding protein H, CbpH
Competence regulation	Competence protein A, ComA Competence protein D, ComD Competence protein E, ComE Competence protein X, ComX Autolysin, LytA Cell wall hydrolase, LytB Cell wall hydrolase, LytC Neuraminidase, NanB $\beta$ -N-Acetylglucosaminidase, StrH Serine/threonine protein kinase
DNA synthesis, repair, and regulation	Superoxide dismutase, SodA DNA repair protein, RecN DNA mismatch protein, RadC DNA inducible damage protein D

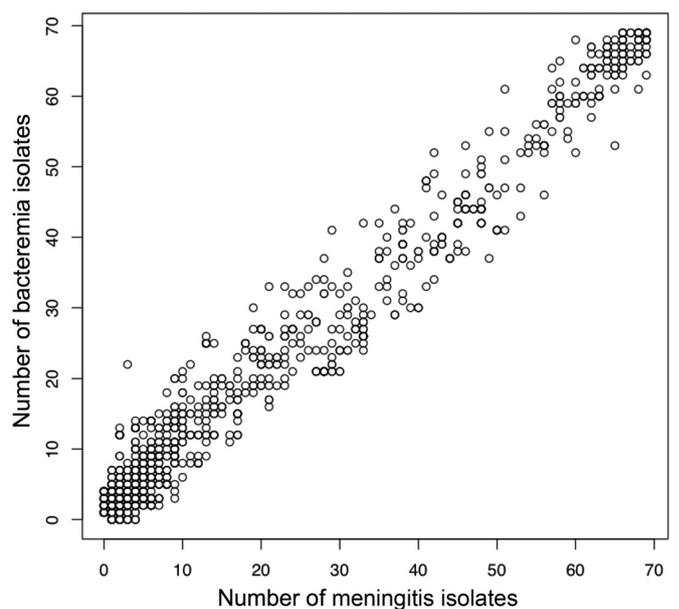
phosphate oxidase (GlpO), which was substantially upregulated in brain tissue. GlpO was cytotoxic for human brain microvascular endothelial cells (HBMECs), and immunization of mice with GlpO protected against invasive pneumococcal disease. The gene encoding GlpO was present in the overall core genome (Table 1). The finding of these putative vaccine candidates in the overall core genome is especially attractive as it suggests that a vaccine based on these candidate molecules is likely to be protective against a wide range of invasive pneumococcal strains.

**The core genomes of bloodstream infection and meningitis pneumococcal isolates.** We then conducted separate analyses of the core genomes of the pneumococcal bloodstream and meningitis subsets to test the hypothesis that there are distinct gene sets that are absolutely required for invasion, growth, and persistence in these different niches. The core genomes of the bloodstream and meningitis subsets were very similar in size (1,457 versus 1,460 genes, respectively) (Fig. 1B). Our analysis identified a subset of 32 genes that were present in the bloodstream core genome but absent in the meningitis core genome (meningitis-specific core genes) and a subset of 29 genes that were present in the meningitis core genome but absent in the bloodstream core genome (bloodstream-specific core genes). However, the genetic differences identified by this comparative analysis (Fig. 1C and D) were due to the absence of key genes in just 3/70 (4%) of the bloodstream isolates (Fig. 1C) and 4/70 (6%) of the meningitis isolates

(Fig. 1D). We found no impact of imbalanced sampling leading to the dominance of a particular clone or an overrepresentation of prophage and integrative conjugative elements (ICE) on the results of the comparative analysis. Therefore, given the known biological functions of the genes identified and the relatively small proportion of isolates responsible for the differences, we think it unlikely that the core genome of meningitis-causing pneumococci is truly distinct from the core genome of bloodstream isolates.

**The accessory genomes of bloodstream infection and meningitis pneumococcal isolates.** We next compared the accessory genomes of the bloodstream and meningitis isolates to interrogate the possibility that a large proportion of the bloodstream isolates possess a subset of accessory genes that were not present in the meningitis accessory genome (Fig. 2). No accessory gene clusters were composed of only bloodstream or meningitis isolates. Among the accessory gene clusters there was a strong positive correlation between the number of bloodstream isolates and the number of meningitis isolates represented in a cluster (Spearman's correlation = 0.79;  $P < 0.0001$ ). No accessory genes were more strongly associated with the bacteremia or meningitis data set.

**Factors influencing the pneumococcal core genome.** The pneumococcal isolates analyzed were randomly selected to avoid a selection bias, and given the size of the data set, there should be a clonal balance representative of the invasive strains in circulation. However, in many SSA countries, serotype 1 (ST217) accounts for over 30% of IPD (9, 10), and in our randomly selected Malawian meningitis data set it accounted for 24% (33/140) of isolates (14 ST217 isolates recovered from blood, 19 isolates recovered from CSF). It has been suggested that highly virulent serotype 1 pneumococcal isolates have genomic regions associated with pathogenicity that are absent in others that are intermediately virulent or avirulent (41). Therefore, to address the possibility that "genetic noise" from the other sequence types (STs) present in the data set may have masked legitimate differences in the core genome com-

**FIG 2** Plot of the number of meningitis and bacteremia isolates represented in each accessory gene cluster. Each circle represents a single gene cluster.

position between ST217 causing bacteremia and ST217 causing CSF invasion, we repeated the above analyses using the ST217 clones alone. The core genomes of the ST217 bloodstream and meningitis subsets differed in size (1,822 versus 1,795 genes, respectively), and we identified a subset of 42 genes that were present in the ST217 bloodstream core genome but absent in the ST217 meningitis core genome (ST217 meningitis-specific core genes) and a subset of 15 genes that were present in the ST217 meningitis core genome but absent in the ST217 bloodstream core genome (ST217 bloodstream-specific core genes). However, as with the observed core genome differences between the bacteremia and meningitis causing isolates in the overall core genome analysis, the difference in the ST217 core genome sizes was due to the absence of key genes in only a very small number of isolates. The 42 core genes that were unique to the ST217 bacteremia isolates were at most absent in 4/19 ( $\leq 21\%$ ) of the ST217 meningitis isolates; likewise the 15 core genes that were unique to the ST217 bacteremia isolates were absent in at most 2/14 ( $\leq 14\%$ ) of the ST217 bacteremia isolates. Again, as with the whole data set, we next compared the accessory genomes of the ST217 meningitis and bloodstream isolates to assess if a large proportion of the ST217 meningitis isolates possess a subset of accessory genes that were not present in the ST217 bloodstream accessory genome. We found that no ST217 accessory gene clusters consisted only of bloodstream or meningitis isolates. We observed a positive correlation between the number of ST217 meningitis isolates and the number of ST217 bloodstream isolates represented in an accessory gene cluster (Spearman's rank coefficient = 0.84;  $P < 0.0001$ ); this strongly suggests that no ST217 accessory genes were more strongly associated with a specific invasive disease phenotype.

We next carried out the same analysis with all ST217 isolates excluded and found that there was no difference between the genome compositions of the bacteremia and meningitis isolates when ST217 was removed from the data set. The core genomes of the bacteremia and meningitis isolates, with ST217 excluded, comprised 1,467 and 1,463 genes, respectively. The 29 bacteremia "unique" isolates were absent in at most 3/53 ( $\leq 6\%$ ) of the meningitis isolates; the meningitis "unique" isolates were absent in at most 4/45 ( $\leq 8\%$ ) of the bacteremia isolates. We observed a positive correlation between the number of meningitis isolates and the number of bloodstream isolates represented in an accessory gene cluster when the ST217 isolates were excluded from the analysis (Spearman's rank coefficient = 0.92;  $P < 0.0001$ ). This additional analysis based on subsets of the data set strongly suggests that our initial finding, which reported that there is no difference in the core genome content between bacteremia and meningitis *S. pneumoniae* isolates, is not an artifact of sample bias.

## DISCUSSION

Numerous human host factors that determine pneumococcal disease susceptibility and severity have been identified, including extremes of age, HIV, splenic dysfunction, sickle cell disease, malignancy, chronic liver disease, malnutrition, and diabetes (42, 43). In the past our understanding of the bacterial components contributing to disease was largely supported by tissue culture models and animal studies. Here, we have used an innovative bioinformatics approach and a stringent analysis pipeline to interrogate the pneumococcal genome associated with invasive disease. We have found, by fitting a double-exponential decaying function,

that disease isolates share a core of 1,427 genes (95% CI, 1,425 to 1,435 genes). The core genome size we report is marginally lower than the core genome size reported for *Streptococcus pneumoniae* by Obert et al. ( $n = 1,553$ ) (44) and Hiller et al. ( $n = 1,454$ ) (25). However, the reduced core genome size can be explained by the higher number of isolates and serotypes included in our analyses, which have likely resulted in greater saturation and a more precise estimate.

We also show that pneumococci that infect the meninges do not differ in their core or accessory genome content from those isolated from the blood, and therefore gene presence/absence alone does not explain the virulence behavior of some pneumococci that reach the brain. Although the development of meningitis may simply be a function of bacterial load and time (26, 45–48), the potential effect of more-subtle differences at the genomic level, specifically allelic variation and gene expression, requires further investigation. For example, a comparison of the genomic sequences of four nasopharyngeal PMEN1 isolates and their capacities to cause disease in a chinchilla otitis media model revealed allelic differences in Blp regulatory proteins and pneumococcal adherence virulence factor B (PavB) in those isolates that spread to the brain (49). Gerlini et al. have shown that bacteremia is a result of a single pneumococcal clone invading the blood; this clone then disseminates, and the resulting invasive pneumococcal population acquires genomic mutations over time as it adapts to the selective pressures of the bloodstream. It is feasible that the same behavior is true for meningitis, whereby a single clone crosses the BBB, proliferates, and rapidly acquires genomic mutations to permit survival in the CSF (50). Microarray analysis of pneumococcal gene expression from blood and CSF infected *in vitro* has previously identified site-specific patterns of gene expression (51). Furthermore, host differences identified by numerous genetic association studies have shown that single nucleotide polymorphisms (SNPs) in innate immunity genes are linked to an individual's susceptibility to pneumococcal meningitis. This suggests that specific pneumococcal genotypes may be able to cause meningitis in some individuals but not others, irrespective of the core genome composition and SNP profile of that genotype (52–54).

A limitation of the study was that it was not possible from the clinical data available to determine if any of the bacteremia isolates included in the data set went on to invade the meninges. It is therefore feasible that some meningitis isolates were incorrectly assigned as bacteremia isolates. This may have masked subtle differences in the accessory genome composition between the bacteremia and meningitis isolates, whereby accessory genes that are more strongly associated with meningitis isolates were found to be evenly distributed between the meningitis and bacteremia isolates in our analysis. This, however, will not have affected the overall core genome composition reported here. It is feasible that the ability of pneumococci to cause meningitis is dictated by the presence of a subset of meningitis-specific genes linked through function and that there may be more than one subset combination that confers the ability to cause meningitis. This hypothesis was not addressed in the current analysis.

The finding of a repertoire of over 1,400 core genes that are common to all invasive-disease-causing pneumococci despite the highly recombinogenic nature of the multiple diverse lineages investigated provides an important biological insight. The presence of these core genes and their associated proteins in all disease-

causing isolates suggests that they are presumably advantageous for colonization and transmission, as well as being essential for the pathogen to invade and survive within the bloodstream and meninges. They include a range of previously identified vaccine candidate molecules and a number of the core gene clusters (13%) that could not be assigned a definitive function. The further characterization of these genes together with an assessment of allelic variation and gene expression among pneumococci from different sites may also reveal important new bacterial housekeeping functions and identify novel candidates that can be exploited as vaccine and therapeutic targets.

## ACKNOWLEDGMENTS

We gratefully acknowledge the clinicians, microbiologists, and study participants of the Malawi-Liverpool Wellcome Trust Clinical Research Program. We thank W. P. Hanage, M. Maiden and E. R. Moxon for their useful comments.

This work was supported by a Wellcome Trust Major Overseas program award (award no. 084679/Z/08/Z) and a Wellcome Trust Grant (award no. 098051).

We declare no conflicts of interest.

## REFERENCES

- O'Brien KL, Wolfson LJ, Watt JP, Henkle E, Deloria-Knoll M, McCall N, Lee E, Mulholland K, Levine OS, Cherian T. 2009. Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. *Lancet* 374:893–902. [http://dx.doi.org/10.1016/S0140-6736\(09\)61204-6](http://dx.doi.org/10.1016/S0140-6736(09)61204-6).
- French N, Gordon SB, Mwalukomo T, White SA, Mwfulirwa G, Longwe H, Mwaiponya M, Zijlstra EE, Molyneux ME, Gilks CF. 2010. A trial of a 7-valent pneumococcal conjugate vaccine in HIV-infected adults. *N Engl J Med* 362:812–822. <http://dx.doi.org/10.1056/NEJMoa0903029>.
- Mook-Kanamori BB, Geldhoff M, van der Poll T, van de Beek D. 2011. Pathogenesis and pathophysiology of pneumococcal meningitis. *Clin Microbiol Rev* 24:557–591. <http://dx.doi.org/10.1128/CMR.00008-11>.
- Orihuela CJ, Mahdavi J, Thornton J, Mann B, Wooldridge KG, Abouseada N, Oldfield NJ, Self T, Ala'Aldeen DA, Tuomanen EI. 2009. Laminin receptor initiates bacterial contact with the blood brain barrier in experimental meningitis models. *J Clin Invest* 119:1638–1646. <http://dx.doi.org/10.1172/JCI36759>.
- Ring A, Weiser JN, Tuomanen EI. 1998. Pneumococcal trafficking across the blood-brain barrier. Molecular analysis of a novel bidirectional pathway. *J Clin Invest* 102:347–360.
- Tuomanen E. 1996. Entry of pathogens into the central nervous system. *FEMS Microbiol Rev* 18:289–299. <http://dx.doi.org/10.1111/j.1574-6976.1996.tb00245.x>.
- Weiser JN, Bae D, Fasching C, Scamurra RW, Ratner AJ, Janoff EN. 2003. Antibody-enhanced pneumococcal adherence requires IgA1 protease. *Proc Natl Acad Sci U S A* 100:4215–4220. <http://dx.doi.org/10.1073/pnas.0637469100>.
- Orihuela CJ, Gao G, Francis KP, Yu J, Tuomanen EI. 2004. Tissue-specific contributions of pneumococcal virulence factors to pathogenesis. *J Infect Dis* 190:1661–1669. <http://dx.doi.org/10.1086/424596>.
- Hausdorff WP, Bryant J, Paradiso PR, Siber GR. 2000. Which pneumococcal serogroups cause the most invasive disease: implications for conjugate vaccine formulation and use, part I. *Clin Infect Dis* 30:100–121. <http://dx.doi.org/10.1086/313608>.
- Hausdorff WP, Bryant J, Kloek C, Paradiso PR, Siber GR. 2000. The contribution of specific pneumococcal serogroups to different disease manifestations: implications for conjugate vaccine formulation and use, part II. *Clin Infect Dis* 30:122–140. <http://dx.doi.org/10.1086/313609>.
- Greenwood B. 2006. Pneumococcal meningitis epidemics in Africa. *Clin Infect Dis* 43:701–703. <http://dx.doi.org/10.1086/506943>.
- Hanage WP, Fraser C, Tang J, Connor TR, Corander J. 2009. Hyper-recombination, diversity, and antibiotic resistance in pneumococcus. *Science* 324:1454–1457. <http://dx.doi.org/10.1126/science.1171908>.
- Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, McGee L, von Gottberg A, Song JH, Ko KS, Pichon B, Baker S, Pary CM, Lambertsen LM, Shahinas D, Pillai DR, Mitchell TJ, Dougan G, Tomasz A, Klugman KP, Parkhill J, Hanage WP, Bentley SD. 2011. Rapid pneumococcal evolution in response to clinical interventions. *Science* 331:430–434. <http://dx.doi.org/10.1126/science.1198545>.
- Golubchik T, Brueggemann AB, Street T, Gertz RE, Jr, Spencer CC, Ho T, Giannoulitou E, Link-Gelles R, Harding RM, Beall B, Peto TE, Moore MR, Donnelly P, Crook DW, Bowden R. 2012. Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event. *Nat Genet* 44:352–355. <http://dx.doi.org/10.1038/ng.1072>.
- Selva L, Viana D, Regev-Yochay G, Trzcinski K, Corpa JM, Lasa I, Novick RP, Penades JR. 2009. Killing niche competitors by remote-control bacteriophage induction. *Proc Natl Acad Sci U S A* 106:1234–1238. <http://dx.doi.org/10.1073/pnas.0809600106>.
- Bergmann S, Hammerschmidt S. 2006. Versatility of pneumococcal surface proteins. *Microbiology* 152:295–303. <http://dx.doi.org/10.1099/mic.0.28610-0>.
- Jedrzejak MJ. 2001. Pneumococcal virulence factors: structure and function. *Microbiol Mol Biol Rev* 65:187–207. <http://dx.doi.org/10.1128/MMBR.65.2.187-207.2001>.
- Harris SR, Feil EJ, Holden MT, Quail MA, Nickerson EK, Chantratita N, Gardete S, Tavares A, Day N, Lindsay JA, Edgeworth JD, de Lencastre H, Parkhill J, Peacock SJ, Bentley SD. 2010. Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 327:469–474. <http://dx.doi.org/10.1126/science.1182395>.
- Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18:821–829. <http://dx.doi.org/10.1101/gr.074492.107>.
- Assefa S, Keane TM, Otto TD, Newbold C, Berriman M. 2009. ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics* 25:1968–1969. <http://dx.doi.org/10.1093/bioinformatics/btp347>.
- Carver T, Berriman M, Tivey A, Patel C, Bohme U, Barrell BG, Parkhill J, Rajandream MA. 2008. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* 24:2672–2676. <http://dx.doi.org/10.1093/bioinformatics/btn529>.
- Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679. <http://dx.doi.org/10.1093/bioinformatics/btm009>.
- Otto TD, Dillon GP, Degraeve WS, Berriman M. 2011. RATT: Rapid Annotation Transfer Tool. *Nucleic Acids Res* 39:e57. <http://dx.doi.org/10.1093/nar/gkq1268>.
- Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30:1575–1584. <http://dx.doi.org/10.1093/nar/30.7.1575>.
- Hiller NL, Janto B, Hogg JS, Boissy R, Yu S, Powell E, Keefe R, Ehrlich NE, Shen K, Hayes J, Barbadora K, Klimke W, Dernovoy D, Tatusova T, Parkhill J, Bentley SD, Post JC, Ehrlich GD, Hu FZ. 2007. Comparative genomic analyses of seventeen *Streptococcus pneumoniae* strains: insights into the pneumococcal supragenome. *J Bacteriol* 189:8186–8195. <http://dx.doi.org/10.1128/JB.00690-07>.
- Carroll ED, Guiver M, Nkhoma S, Mankambo LA, Marsh J, Balmer P, Banda DL, Jeffers G, White SA, Molyneux EM, Molyneux ME, Smyth RL, Hart CA. 2007. High pneumococcal DNA loads are associated with mortality in Malawian children with invasive pneumococcal disease. *Pediatr Infect Dis J* 26:416–422. <http://dx.doi.org/10.1097/01.inf.0000260253.22994.61>.
- Molyneux EM, Walsh AL, Forsyth H, Tembo M, Mwenechanya J, Kayira K, Bwanaisa L, Njobvu A, Rogerson S, Malenga G. 2002. Dexamethasone treatment in childhood bacterial meningitis in Malawi: a randomised controlled trial. *Lancet* 360:211–218. [http://dx.doi.org/10.1016/S0140-6736\(02\)09458-8](http://dx.doi.org/10.1016/S0140-6736(02)09458-8).
- Brueggemann AB, Griffiths DT, Meats E, Peto T, Crook DW, Spratt BG. 2003. Clonal relationships between invasive and carriage *Streptococcus pneumoniae* and serotype- and clone-specific differences in invasive disease potential. *J Infect Dis* 187:1424–1432. <http://dx.doi.org/10.1086/374624>.
- Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, Deboy RT, Davidsen TM, Mora M, Scarselli M, Margarit y Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou L, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K,

- O'Connor KJ, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM. 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome." *Proc Natl Acad Sci U S A* 102:13950–13955. <http://dx.doi.org/10.1073/pnas.0506758102>.
30. Weiser JN. 2010. The pneumococcus: why a commensal misbehaves. *J Mol Med* 88:97–102. <http://dx.doi.org/10.1007/s00109-009-0557-x>.
31. Yesilkaya H, Spissu F, Carvalho SM, Terra VS, Homer KA, Benisty R, Porat N, Neves AR, Andrew PW. 2009. Pyruvate formate lyase is required for pneumococcal fermentative metabolism and virulence. *Infect Immun* 77:5418–5427. <http://dx.doi.org/10.1128/IAI.00178-09>.
32. Basavanna S, Khandavilli S, Yuste J, Cohen JM, Hosie AH, Webb AJ, Thomas GH, Brown JS. 2009. Screening of *Streptococcus pneumoniae* ABC transporter mutants demonstrates that LivJHMGF, a branched-chain amino acid ABC transporter, is necessary for disease pathogenesis. *Infect Immun* 77:3412–3423. <http://dx.doi.org/10.1128/IAI.01543-08>.
33. Brown JS, Gilliland SM, Ruiz-Albert J, Holden DW. 2002. Characterization of pit, a *Streptococcus pneumoniae* iron uptake ABC transporter. *Infect Immun* 70:4389–4398. <http://dx.doi.org/10.1128/IAI.70.8.4389-4398.2002>.
34. Pettigrew MM, Fennie KP, York MP, Daniels J, Ghaffar F. 2006. Variation in the presence of neuraminidase genes among *Streptococcus pneumoniae* isolates with identical sequence types. *Infect Immun* 74:3360–3365. <http://dx.doi.org/10.1128/IAI.01442-05>.
35. Sakai F, Talekar SJ, Klugman KP, Vidal JE, for the Investigators G. 2013. Expression of virulence-related genes in the nasopharynx of healthy children. *PLoS One* 8:e67147. <http://dx.doi.org/10.1371/journal.pone.0067147>.
36. Zhang JR, Mostov KE, Lamm ME, Nanno M, Shimida S, Ohwaki M, Tuomanen E. 2000. The polymeric immunoglobulin receptor translocates pneumococci across human nasopharyngeal epithelial cells. *Cell* 102:827–837. [http://dx.doi.org/10.1016/S0092-8674\(00\)00071-4](http://dx.doi.org/10.1016/S0092-8674(00)00071-4).
37. Claverys JP, Prudhomme M, Martin B. 2006. Induction of competence regulons as a general response to stress in gram-positive bacteria. *Annu Rev Microbiol* 60:451–475. <http://dx.doi.org/10.1146/annurev.micro.60.080805.142139>.
38. Prudhomme M, Libante V, Claverys JP. 2002. Homologous recombination at the border: insertion-deletions and the trapping of foreign DNA in *Streptococcus pneumoniae*. *Proc Natl Acad Sci U S A* 99:2100–2105. <http://dx.doi.org/10.1073/pnas.032262999>.
39. Bogaert D, De Groot R, Hermans PW. 2004. *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. *Lancet Infect Dis* 4:144–154. [http://dx.doi.org/10.1016/S1473-3099\(04\)00938-7](http://dx.doi.org/10.1016/S1473-3099(04)00938-7).
40. Mahdi LK, Wang H, Van der Hoek MB, Paton JC, Ogunniyi AD. 2012. Identification of a novel pneumococcal vaccine antigen preferentially expressed during meningitis in mice. *J Clin Invest* 122:2208–2220. <http://dx.doi.org/10.1172/JCI45850>.
41. Harvey RM, Stroehrer UH, Ogunniyi AD, Smith-Vaughan HC, Leach AJ, Paton JC. 2011. A variable region within the genome of *Streptococcus pneumoniae* contributes to strain-strain variation in virulence. *PLoS One* 6:e19650. <http://dx.doi.org/10.1371/journal.pone.0019650>.
42. Williams TN, Uyoga S, Macharia A, Ndila C, McAuley CF, Opi DH, Mwarumba S, Makani J, Komba A, Ndiritu MN, Sharif SK, Marsh K, Berkley JA, Scott JA. 2009. Bacteraemia in Kenyan children with sickle-cell anaemia: a retrospective cohort and case-control study. *Lancet* 374:1364–1370. [http://dx.doi.org/10.1016/S0140-6736\(09\)61374-X](http://dx.doi.org/10.1016/S0140-6736(09)61374-X).
43. Berkley JA, Lowe BS, Mwangi I, Williams T, Bauni E, Mwarumba S, Ngeta C, Slack MP, Njenga S, Hart CA, Maitland K, English M, Marsh K, Scott JA. 2005. Bacteraemia among children admitted to a rural hospital in Kenya. *N Engl J Med* 352:39–47. <http://dx.doi.org/10.1056/NEJMoa040275>.
44. Obert C, Sublett J, Kaushal D, Hinojosa E, Barton T, Tuomanen EI, Orihuela CJ. 2006. Identification of a candidate *Streptococcus pneumoniae* core genome and regions of diversity correlated with invasive pneumococcal disease. *Infect Immun* 74:4766–4777. <http://dx.doi.org/10.1128/IAI.00316-06>.
45. Moxon ER, Smith AL, Averill DR, Smith DH. 1974. Haemophilus influenzae meningitis in infant rats after intranasal inoculation. *J Infect Dis* 129:154–162. <http://dx.doi.org/10.1093/infdis/129.2.154>.
46. Moxon ER, Ostrow PT. 1977. Haemophilus influenzae meningitis in infant rats: role of bacteremia in pathogenesis of age-dependent inflammatory responses in cerebrospinal fluid. *J Infect Dis* 135:303–307. <http://dx.doi.org/10.1093/infdis/135.2.303>.
47. Moxon ER, Murphy PA. 1978. Haemophilus influenzae bacteremia and meningitis resulting from survival of a single organism. *Proc Natl Acad Sci U S A* 75:1534–1536. <http://dx.doi.org/10.1073/pnas.75.3.1534>.
48. Huang SH, Stins MF, Kim KS. 2000. Bacterial penetration across the blood-brain barrier during the development of neonatal meningitis. *Microbes Infect* 2:1237–1244. [http://dx.doi.org/10.1016/S1286-4579\(00\)01277-6](http://dx.doi.org/10.1016/S1286-4579(00)01277-6).
49. Hiller NL, Eutsey RA, Powell E, Earl JP, Janto B, Martin DP, Dawid S, Ahmed A, Longwell MJ, Dahlgren ME, Ezzo S, Tettelin H, Daugherty SC, Mitchell TJ, Hillman TA, Buchinsky FJ, Tomasz A, de Lencastre H, Sa-Leao R, Post JC, Hu FZ, Ehrlich GD. 2011. Differences in genotype and virulence among four multidrug-resistant *Streptococcus pneumoniae* isolates belonging to the PMEN1 clone. *PLoS One* 6:e28850. <http://dx.doi.org/10.1371/journal.pone.0028850>.
50. Gerlini A, Colomba L, Furi L, Braccini T, Manso AS, Pammolli A, Wang B, Vivi A, Tassini M, van Rooijen N, Pozzi G, Ricci S, Andrew PW, Koedel U, Moxon ER, Oggioni MR. 2014. The role of host and microbial factors in the pathogenesis of pneumococcal bacteraemia arising from a single bacterial cell bottleneck. *PLoS Pathog* 10:e1004026. <http://dx.doi.org/10.1371/journal.ppat.1004026>.
51. Orihuela CJ, Radin JN, Sublett JE, Gao G, Kaushal D, Tuomanen EI. 2004. Microarray analysis of pneumococcal gene expression during invasive disease. *Infect Immun* 72:5582–5596. <http://dx.doi.org/10.1128/IAI.72.10.5582-5596.2004>.
52. Brouwer MC, de Gans J, Heckenberg SG, Zwinderman AH, van der Poll T, van de Beek D. 2009. Host genetic susceptibility to pneumococcal and meningococcal disease: a systematic review and meta-analysis. *Lancet Infect Dis* 9:31–44. [http://dx.doi.org/10.1016/S1473-3099\(08\)70261-5](http://dx.doi.org/10.1016/S1473-3099(08)70261-5).
53. Sanders MS, van Well GT, Ouburg S, Morre SA, van Furth AM. 2011. Genetic variation of innate immune response genes in invasive pneumococcal and meningococcal disease applied to the pathogenesis of meningitis. *Genes Immun* 12:321–334. <http://dx.doi.org/10.1038/gene.2011.20>.
54. Woehrl B, Brouwer MC, Murr C, Heckenberg SG, Baas F, Pfister HW, Zwinderman AH, Morgan BP, Barnum SR, van der Ende A, Koedel U, van de Beek D. 2011. Complement component 5 contributes to poor disease outcome in humans and mice with pneumococcal meningitis. *J Clin Invest* 121:3943–3953. <http://dx.doi.org/10.1172/JCI57522>.

## ERRATUM

# Erratum for Kulohoma et al., Comparative Genomic Analysis of Meningitis- and Bacteremia-Causing Pneumococci Identifies a Common Core Genome

Benard W. Kulohoma,<sup>a,b</sup> Jennifer E. Cornick,<sup>a,b</sup> Chrispin Chaguza,<sup>a,b</sup> Feyruz Yalcin,<sup>c</sup> Simon R. Harris,<sup>c</sup> Katherine J. Gray,<sup>d</sup> Anmol M. Kiran,<sup>a,b</sup> Elizabeth Molyneux,<sup>d</sup> Neil French,<sup>b</sup> Julian Parkhill,<sup>c</sup> Brian E. Faragher,<sup>e</sup> Dean B. Everett,<sup>a,b</sup> Stephen D. Bentley,<sup>b,c</sup> Robert S. Heyderman<sup>a</sup>

Malawi-Liverpool-Wellcome Trust Clinical Research Programme, University of Malawi College of Medicine, Blantyre, Malawi<sup>a</sup>; Institute of Infection and Global Health, University of Liverpool, Liverpool, United Kingdom<sup>b</sup>; The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom<sup>c</sup>; University of Malawi College of Medicine, Blantyre, Malawi<sup>d</sup>; Liverpool School of Tropical Medicine, Liverpool, United Kingdom<sup>e</sup>

Volume 83, no. 10, p. 4165–4173, 2015. Page 4165, abstract, line 5, “1,427” should read “1,428.”

Page 4167, column 1, line 40: “1,427” should read “1,428.”

Page 4171, column 2, line 1: “1,427” should read “1,428.”

**Citation** Kulohoma BW, Cornick JE, Chaguza C, Yalcin F, Harris SR, Gray KJ, Kiran AM, Molyneux E, French N, Parkhill J, Faragher BE, Everett DB, Bentley SD, Heyderman RS. 2015. Erratum for Kulohoma et al., Comparative genomic analysis of meningitis- and bacteremia-causing pneumococci identifies a common core genome. *Infect Immun* 83:4896. doi:10.1128/IAI.01156-15.

Copyright © 2015, American Society for Microbiology. All Rights Reserved.