Identification of Genes That Contribute to the Pathogenesis of Invasive Pneumococcal Disease by In Vivo Transcriptomic Analysis

Abiodun D. Ogunniyi, Layla K. Mahdi, Claudia Trappetti, Nadine Verhoeven, Daphne Mermans, Mark B. Van der Hoek, Charles D. Plump tre, and James C. Paton

Research Centre for Infectious Diseases, School of Molecular and Biomedical Science, The University of Adelaide, Adelaide, South Australia, Australia; and Adelaide Microarray Centre, The University of Adelaide and SA Pathology, Adelaide, South Australia, Australia

Streptococcus pneumoniae (the pneumococcus) continues to be responsible for a high level of global morbidity and mortality resulting from pneumonia, bacteremia, meningitis, and otitis media. Here we have used a novel technique involving niche-specific, genome-wide in vivo transcriptomic analyses to identify genes upregulated in distinct niches during pathogenesis after intranasal infection of mice with serotype 4 or 6A pneumococci. The analyses yielded 28 common, significantly upregulated genes in the lungs relative to those in the nasopharynx and 25 significantly upregulated genes in the blood relative to those in the lungs in both strains, some of which were previously unrecognized. The role of five upregulated genes from either the lungs or the blood in pneumococcal pathogenesis and virulence was then evaluated by targeted mutagenesis. One of the mutants (ΔmalX) was significantly attenuated for virulence in the lungs, two (ΔaliA and ΔibH) were significantly attenuated for virulence in the blood relative to the wild type, and two others (ΔchiO and ΔpiuA) were completely avirulent in a mouse intranasal challenge model. We also show that the products of aliA, malX, and piuA are promising candidates for incorporation into multicomponent protein-based pneumococcal vaccines currently under development. Importantly, we suggest that this new approach is a viable complement to existing strategies for the discovery of genes critical to the distinct stages of invasive pneumococcal disease and potentially has broad application for novel protein antigen discovery in other pathogens such as S. pyogenes, Haemophilus influenzae type b, and Neisseria meningitidis.
strategies have also been employed to screen in vivo induced genes for the discovery and appraisal of novel candidate proteins for inclusion in a multicomponent pneumococcal protein vaccine under development. The strategies include, but are not limited to, in vivo expression technology (IVET) (27), signature-tagged mutagenesis (STM) (12, 18, 44), differential fluorescence induction (DFI) (4, 23), microarray transcriptomics (40), reverse vaccinology (3, 45), antigenome technology (10, 26), genomic array footprinting (GAF) (29), and protein expression library screening (28).

To gain novel insights into global expression of pneumococcal virulence genes during pathogenesis of invasive disease, we conducted a comprehensive microarray comparison of gene expression kinetics between pneumoccoci in the nasopharynx, lungs, and blood of mice infected with two virulent S. pneumoniae strains. We hypothesized that progression from carriage to invasive disease will require niche-specific alterations in virulence gene expression and that transcriptomic analysis will identify common virulence factors critical for disease progression. We also hypothesized that the virulence factors that are upregulated during progression from carriage to disease, if surface exposed, are likely to be protective immunogens. Accordingly, we tested and validated progression from carriage to disease, if surface exposed, are likely to be protective immunogens. Accordingly, we tested and validated.

### Materials and Methods

#### Bacterial strains and growth conditions.

The pneumococcal strains used in this study are serotype 2 (D39), serotype 4 (WCH43), serotype 6A (WCH16), and their respective isogenic mutant derivatives (Table 1). Serotype-specific capsule production was confirmed by Quellung reaction, as described previously (5). Opaque-phase variants of the three strains, selected on Todd-Hewitt broth supplemented with 1% yeast extract (THY)-catalase plates (50), were used in all animal experiments. Before infection, the bacteria were grown statically at 37°C in serum broth (10% heat-inactivated horse serum in nutrient broth) to an OD600 of approximately 1.6 (equivalent to approximately 5 × 10^7 CFU/ml).

#### Mice.

Outbred 5- to 6-week-old female CD1 (Swiss) mice, obtained from the Laboratory Animal Services breeding facility of The University of Adelaide, were used in all experiments. The Animal Ethics Committee of The University of Adelaide approved all animal experiments. The study was conducted in compliance with the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes (7th edition, 2004) and the South Australian Animal Welfare Act 1985.

#### Intranasal challenge of mice for gene expression analyses.

For each triplicate WCH16 and WCH43 challenge experiment, groups of 40 mice were used. The mice were anesthetized by intraperitoneal (i.p.) injection of pentobarbital sodium (Nembutal; Rhone-Merieux) at a dose of 66 mg per g of body weight and challenged intranasally (i.n.) with 50 μl of bacterial suspension containing approximately 1 × 10^7 to 2 × 10^7 CFU in serum broth. The challenge dose was confirmed retrospectively by serial dilution and plating of the inocula on blood agar. In each experiment, at 48, 72, and 96 h postchallenge, at least 12 mice were sacrificed by CO2 asphyxiation, and samples from the nasopharynx, lungs, and blood were processed as described previously (19, 33). Before processing, lung samples were further rinsed in 25 ml sterile, ice-cold phosphate-buffered saline (PBS) before homogenization to remove contaminating blood from the surface. A 40-μl aliquot of each sample was serially diluted in serum broth and plated on blood agar to enumerate pneumococci present in each niche. Blood plates were incubated at 37°C in 95% air, 5% CO2 overnight. Samples were then stored at −80°C until further processing was done. The experiment was performed three times for each strain.

#### Extraktion of total RNA from mouse tissues.

For microarray experiments, samples from each niche were pooled. Total RNA was extracted and purified as described previously (19, 22). Bacterial RNAs from a minimum of five mice that satisfied these criteria from each specific niche were pooled. The RNA was then purified further using a Qiagen RNeasy minikit. Purified RNA samples were checked for purity and integrity as described previously (19, 22). Bacterial RNAs from a minimum of five mice that satisfied these criteria from each specific niche were pooled. The DNA was then purified further using a Qiagen RNeasy minikit. RNA was further enriched for prokaryotic RNA using the MicroElutrix kit (Ambion). The amount of DNA recovered following purification and enrichment was determined by A_260/A_280 measurements.

#### Microarray analysis of bacterial RNA.

Microarray experiments were performed on whole-genome S. pneumoniae PCR arrays obtained from the Bacterial Microarray Group at St George’s Hospital Medical School, London, England (http://bugs.sgul.ac.uk/). The array was designed using TIGR4 base strain annotation (49) and extra target genes from strain R6 (15). The array design is available in B bugs@Sbase (accession no. A-BUGS-14; http://bugs.sgul.ac.uk/A-BUGS-14) and also ArrayExpress (accession no. A-BUGS-14).

#### Microarray probes were generated using the 3DNA Array 900 MPX labeling kit (Genisphere) following the manufacturer’s guidelines. Total amplified RNA of S. pneumoniae obtained from the nasopharynx, lungs, and blood was used, and pairwise comparisons were made between the nasopharynx and lung and between the lung and blood RNA samples from the 48-, 72-, and 96-h time points. RNA samples were

### Table 1. S. pneumoniae strains used in this study

<table>
<thead>
<tr>
<th>Strain</th>
<th>Description (sequence type)</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>D39</td>
<td>Capsular serotype 2 (395)</td>
<td>NCTC 7466</td>
</tr>
<tr>
<td>D39::ΔubiO</td>
<td>ubiO deletion mutant of D39 [Ery']</td>
<td>Present study</td>
</tr>
<tr>
<td>D39::ΔpiuA</td>
<td>piuA deletion mutant of D39 [Ery']</td>
<td>Present study</td>
</tr>
<tr>
<td>WCH16</td>
<td>Capsular serotype 6A clinical isolate (4966)</td>
<td>Women’s and Children’s Hospital, North Adelaide, Australia</td>
</tr>
<tr>
<td>WCH43</td>
<td>Capsular serotype 4 clinical isolate (205)</td>
<td>Women’s and Children’s Hospital, North Adelaide, Australia</td>
</tr>
<tr>
<td>WCH43::ΔadlB2</td>
<td>adlB2 deletion mutant of WCH43 [Spc']</td>
<td>Present study</td>
</tr>
<tr>
<td>WCH43::ΔaliA</td>
<td>aliA deletion mutant of WCH43 [Ery']</td>
<td>Present study</td>
</tr>
<tr>
<td>WCH43::ΔcmvO</td>
<td>cmvO deletion mutant of WCH43 [Ery']</td>
<td>Present study</td>
</tr>
<tr>
<td>WCH43::ΔcpeE</td>
<td>cpeE deletion mutant of WCH43 [Ery']</td>
<td>Present study</td>
</tr>
<tr>
<td>WCH43::ΔcgkB</td>
<td>cgkB deletion mutant of WCH43 [Spc']</td>
<td>Present study</td>
</tr>
<tr>
<td>WCH43::ΔdltB</td>
<td>dltB deletion mutant of WCH43 [Spc']</td>
<td>Present study</td>
</tr>
<tr>
<td>WCH43::ΔihVH</td>
<td>ihVH deletion mutant of WCH43 [Ery']</td>
<td>Present study</td>
</tr>
<tr>
<td>WCH43::ΔmalX</td>
<td>malX deletion mutant of WCH43 [Spc']</td>
<td>Present study</td>
</tr>
<tr>
<td>WCH43::ΔmrdD</td>
<td>mrdD deletion mutant of WCH43 [Spc']</td>
<td>Present study</td>
</tr>
<tr>
<td>WCH43::ΔpiuA</td>
<td>piuA deletion mutant of WCH43 [Ery']</td>
<td>Present study</td>
</tr>
</tbody>
</table>
reverse transcribed using Superscript III (Invitrogen) and then labeled with either Alexa Fluor 546 or Alexa Fluor 647 dye. The fluorescently labeled cDNAs for each pairwise comparison were then combined and hybridized to the surface of the microarray, essentially as described previously (25, 36).

Slides were scanned at 10-μm resolution using a Genepix 4000B scanner (Molecular Devices, USA). Detector photomultiplier tube (PMT) voltages were adjusted individually for each slide so that the total red (Alexa Fluor 647) and green (Alexa Fluor 546) fluorescence signals for each channel were approximately equal while minimizing the total number of features with signal above the maximum detectable. Foreground and background mean pixel intensity values were extracted from the scanned images for both channels (Alexa Fluor 546, Alexa Fluor 647) using the Spot plugin (CSIRO, Australia) within the R statistical software package (http://www.R-project.org). The Limma plugin for R (47) was used for data processing and statistical analysis. After background subtraction, the foreground intensities were log2 transformed and a single ratio (Alexa Fluor 647/Alexa Fluor 546) value was obtained for each probe. Ratio values were normalized using the print-tip Loess normalization routine (48). The replicate arrays were normalized to each other to give similar ranges of mRNA expression values. For each probe across the arrays a linear model was fitted to determine a final expression value for each mRNA probed and associated statistics (46). These statistics were used to rank the mRNAs from those most likely to be differentially expressed to the least likely using false-discovery rate values of P < 0.05. Microarray analysis examining RNA from infected nasopharynx versus lungs and lungs versus blood was performed on at least 9 independent hybridizations for each pairwise comparison from three separate assays, including at least one dye reversal per comparison for each strain.

Relative quantitation real-time reverse transcription-PCR (RT-PCR). For a subset of selected genes that showed significant levels of upregulation in the lungs or blood by microarray analysis, gene expression was validated using SuperScript III one-step RT-PCR kit (Invitrogen) in a LightCycler480 II (Roche) as described previously (22). The relative gene expression was analyzed using the 2−ΔΔCT method (21). The reference gene was 16S rRNA. The primer pairs used for gene expression analysis are listed in Table S1 in the supplemental material. All data were obtained from three biological replicates.

Construction of mutants and assessment of bacterial growth in vitro. Defined, nonpolar mutants of genes of interest were constructed in WCH43 (serotype 4) and, for comparison to the wild type, bacterial strains were grown in serum broth giving 50% of the highest absorbance reading above the background at 405 nm. After two weeks the third hybridization, mice were challenged i.p. with approximately 3 × 10^8 CFU (in 100 μl) of WCH43. Mice were closely monitored for survival over 21 days. Differences in median survival time of mice between groups were analyzed by the Mann-Whitney U test (one tailed).

Determination of membrane/surface localization. Wild-type WCH43 cells were grown in C+Y medium at an A600 of 0.3. After one wash in PBS, bacteria were resuspended in 50 μl of protein-specific polyclonal murine antisera or anti-alum serum (diluted 1:50, vol/vol), for 30 min at 37°C, followed by two more washes and then 50 μl of Alexa 488 donkey antiserum IgG, heavy and light chain (H+L) (Invitrogen; diluted 1:200, vol/vol) for 1 h on ice. Three replicates for each serum were analyzed using a BD FACSCanto flow cytometer (BD Biosciences, San Jose, CA) and the software package Weasel (Walter and Eliza Hall Institute of Medical Research). Data shown are representative of three independent experiments. Histograms are representative of fluorescence-activated cell sorter (FACS) data for each protein.

For cell fractionation studies, wild-type WCH43 and its isogenic ΔaliA, ΔmaI3X, and ΔpiuA mutants were grown as described above. Cell pellets were resuspended in PBS and lysed in a French pressure cell (SLM Aminco Inc.) at 12,000 lb/in^2. Lysates were then subjected to ultracentrifugation at 100,000 × g for 1 h to separate soluble (cytoplasmic) and insoluble (membrane and cell wall) fractions. After several washes in PBS, pellets containing the insoluble fractions were resuspended in PBS con-
Virulence Genes in Invasive Pneumococcal Disease

taining 2% SDS overnight at room temperature. All protein samples were quantitated by A_{280} measurements on a NanoDrop 1000 Spectrophotometer (Thermo Scientific). Approximately 50 μg of each sample was resuspended in sample loading buffer before electrophoresis on a 4 to 12% Bis Tris NuPAGE gel and then transferred on an iblot dry blotting (nitrocellulose) system (Invitrogen, Life Technologies). After transfer, each membrane was reacted with a dilution of 1/3,000 polyclonal mouse anti-AliA, MalX, and PiuA sera, using anti-PsaA as a loading control and membrane compartment marker for AliA and MalX blots and anti-AliA serum as a marker for the PiuA blot. Bound antibodies were detected by using a 1/3,000 dilution of alkaline phosphatase-conjugated goat anti-mouse IgG (H + L) (Bio-Rad) with disodium p-nitrophenylphosphate (Sigma) as the substrate.

Microarray data accession numbers. Fully annotated microarray data have been deposited in Bpe:G@Sbase (accession number E-BUGS-133; http://bugs.sgul.ac.uk/E-BUGS-133) and also ArrayExpress (accession number E-BUGS-133).

RESULTS
Transcriptomic analysis identifies novel, differentially expressed genes in discrete in vivo niches. Previous mouse intranasal challenge experiments in our laboratory had indicated that D39 and WCH43 are more virulent than WCH16 (A. D. Oggunniyi and J. C. Paton, unpublished data). Furthermore, after infection of mice, D39 causes fulminant bacteremia and WCH43 demonstrates the “classical” disease progression from the nasopharynx to the lungs, followed by dissemination to the blood and then to the brain, while WCH16 demonstrates minimal lung and blood involvement before translocation to the brain (22, 39).

The distinct virulence/pathogenic profiles exhibited by WCH16 and WCH43 reflect the spectrum of disease in humans, making these strains suitable for further analysis. Therefore, bacteria harvested from the nasopharynx, lungs, and blood of mice at 48, 72, and 96 h after infection with either WCH16 or WCH43, in three replicate experiments, were subjected to pairwise transcriptomic comparisons, as described in Materials and Methods. Our analysis shows dramatic differences in transcript abundance between the two strains. For WCH16, the comparisons yielded 253 significantly up- or downregulated genes in the lungs relative to the nasopharynx (Fig. 1A), and 171 genes differentially regulated in the blood relative to the lungs (Fig. 1B). For WCH43, 315 genes were differentially regulated in the lungs relative to the nasopharynx (Fig. 1A), while 370 genes were differentially regulated in the blood relative to the lungs (Fig. 1B).

Further, the genes that were differentially expressed in the discrete in vivo niches of mice infected with either WCH16 or WCH43 were cross-compared. In this manner, only genes that were consistently and reproducibly differentially expressed at all three time points, in all three replicate experiments, and in both strains were chosen for analysis. The analyses revealed 115 significantly differentially expressed genes in the lungs versus nasopharynx (Fig. 1A; see Fig. S1 and Table S3 in the supplemental material), and 43 differentially expressed genes in the blood versus lungs (Fig. 1B; see Fig. S2 and Table S3 in the supplemental material), common to both strains. We then used an arbitrary cutoff of ≥1.3-fold regulation for genes for further analysis, as any upregulation below this threshold was not considered to be physiologically relevant. In so doing, we found 28 genes upregulated in the lungs compared to the nasopharynx (Table 2) and 25 genes upregulated in the blood compared to the lungs (Table 3), common to both strains. For each pairwise comparison, 8 of the upregulated genes were randomly selected for validation by RT-PCR for both strains (Tables 2 and 3). Of the 28 genes upregulated in the lungs compared to the nasopharynx, 18 have been identified in other screens, and 14 of the 25 genes upregulated in the blood compared to the lungs have been identified in other screens (Tables 2 and 3).

Pneumococcal genes upregulated in the lungs and blood contribute to pathogenesis and virulence. To evaluate the role in pathogenesis and/or virulence of genes preferentially upregulated in either the lungs or blood, five genes specifically upregulated in either niche were subjected to targeted mutagenesis in which each deleted target gene is replaced by a nonpolar antibiotic resistance cassette using overlap PCR (see Materials and Methods). Selection criteria included, but were not limited to, in silico bioinformatics prediction of a role in metabolism, nutrient uptake for survival in the lungs and/or blood, and putative contribution to pathogenesis. The mutation was carried out in WCH43 because it is more virulent than WCH16. Mutation of the 10 genes chosen for further analysis did not adversely affect their growth in vitro in serum broth (data not shown).

The virulence characteristic of each mutant relative to that of the isogenic wild-type strain (WCH43) was assessed by intranasal (i.n.) challenge of mice. In this assay, two of the mutants (ΔcibO and ΔpiuA) were completely avirulent (Fig. 2). These mutants were also completely avirulent in this challenge model in a D39 background (not shown). Other mutants, such as the ΔmalX, ΔaliA, and ΔivhH strains, were also significantly attenuated (Fig. 2). To further examine the contribution of each differentially expressed gene to pathogenesis, we carried out competition experiments in which groups of mice were challenged i.n. with approximately equal numbers of mutant and wild-type pneumococci. This enabled evaluation of fine differences between wild type and mutant by calculating the competitive index for each mutant in the various niches at 48 h postinfection. The ΔcibO and ΔpiuA mutants were excluded from this assay because of their complete

![FIG 1 Venn diagram of microarray data comparisons of in vivo-expressed Streptococcus pneumoniae genes. Analysis of interaction of common, differentially expressed genes from microarray comparisons of lungs versus nasopharynx (L vs N) (A) and lungs versus blood (L vs B) (B) in WCH16 and WCH43.](attachment:image)
attenuation in the challenge model. Compared to the wild type, the ∆malX mutant was significantly attenuated in the lungs (Fig. 3A) and the ∆aliA mutant was significantly attenuated in the lungs and blood (Fig. 3B), while the ∆ilvH mutant was significantly attenuated in the blood (Fig. 3C). However, the numbers of the ∆ilvB mutant were not significantly different from those of the wild type in any of the niches (Fig. 3D). These results are complementary to the virulence data (Fig. 2), and they further validate the in vivo transcriptomic approach used to identify these genes.

Differentially expressed virulence factors are protective against pneumococcal sepsis. The promising pathogenesis and virulence data obtained for ΔcbiO, ΔpiuA, ΔmalX, ΔaliA, and ΔilvH mutants led us to carry out in silico bioinformatics analysis to predict cellular localization. Of the aforementioned factors, PiuA, MalX, and AliA had signal peptidase II recognition sequences and were predicted to be lipoproteins. Accordingly, the respective recombinant products were expressed and purified as His6-tagged fusion proteins in E. coli as detailed in Materials and Methods. Flow cytometry of whole cells with protein-specific murine polyclonal antiseras revealed a very low level of surface exposure of PiuA, MalX, and AliA compared to Pht proteins, which are known to be exposed on the pneumococcal surface (Fig. 4A and B). However, the difference in fluorescence for these antigens compared to negative-control anti-alum serum was similar to that of PsaA, a proven lipoprotein and vaccine candidate (25). We then used Western blotting of soluble and membrane compartments of fractionated cells (Fig. 4C to E) to show that like PsaA, PiuA, MalX, and AliA were almost exclusively present in the membrane fraction.

We then evaluated the potential of purified recombinant PiuA, MalX, and AliA to protect against fatal sepsis in a mouse intraperitoneal (i.p.) immunization challenge experiment, using the non-toxic pneumolysin derivative PdT as a control (see Materials and Methods). First, we found that strong, antigen-specific antibody responses were generated in mice to each of the immunogens (total IgG titer for AliA = 16,000 ± 1,000, for MalX = 48,000 ± 2,000, for PiuA = 16,000 ± 1,000, and for PdT = 12,000 ± 1,000). Second, we found that mice immunized with recombinant PiuA or AliA alone were significantly protected against challenge with WCH43 compared to mice in the placebo group (P = 0.004 and 0.003, respectively) (Fig. 5), while protection imparted by immunization with either MalX or PdT alone was not significant (P = 0.060 and P = 0.051, respectively).

Table 2 Differential gene expression profiles of S. pneumoniae WCH16 and WCH43 in the nasopharynx and lungs of mice by microarray analysis

<table>
<thead>
<tr>
<th>Gene ID (TIGR4)*</th>
<th>Gene annotation</th>
<th>Identified in previous screens†</th>
<th>Fold change (lungs/nasopharynx)‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sp-0043</td>
<td>Competence factor transport protein ComB</td>
<td>Yes (4, 18, 23, 40)</td>
<td>2.0</td>
</tr>
<tr>
<td>Sp-0049</td>
<td>VanZ protein, putative</td>
<td>Yes (12)</td>
<td>1.6 (7.7)</td>
</tr>
<tr>
<td>Sp-0202</td>
<td>Anaerobic ribonucleoside triphosphate reductase (NrdD)</td>
<td>Yes (40)</td>
<td>2.3 (8.2)</td>
</tr>
<tr>
<td>Sp-0463</td>
<td>Cell wall surface anchor family protein</td>
<td>Yes (10)</td>
<td>1.3</td>
</tr>
<tr>
<td>Sp-0545</td>
<td>Immunity protein (BlpY)</td>
<td>Yes (40)</td>
<td>1.6</td>
</tr>
<tr>
<td>Sp-0758</td>
<td>PTS system, IAABC components</td>
<td>No</td>
<td>1.7</td>
</tr>
<tr>
<td>Sp-0785</td>
<td>HlyD family secretion protein</td>
<td>Yes (10, 12)</td>
<td>1.7 (1.8)</td>
</tr>
<tr>
<td>Sp-0787</td>
<td>Antimicrobial peptide ABC transporter permease</td>
<td>Yes (40)</td>
<td>1.5</td>
</tr>
<tr>
<td>Sp-0803</td>
<td>Bacterial cell division membrane protein (RodA)</td>
<td>No</td>
<td>1.4</td>
</tr>
<tr>
<td>Sp-0804</td>
<td>4-Methyl-5(b-hydroxyethyl)-thiazole monophosphate biosynthesis protein</td>
<td>Yes (40)</td>
<td>3.6 (8.4)</td>
</tr>
<tr>
<td>Sp-1083</td>
<td>Putative transcriptional regulator</td>
<td>No</td>
<td>1.4</td>
</tr>
<tr>
<td>Sp-1123</td>
<td>Glycogen biosynthesis protein (GlgD)</td>
<td>Yes (10)</td>
<td>1.5</td>
</tr>
<tr>
<td>Sp-1368</td>
<td>Pat protein</td>
<td>Yes (23)</td>
<td>1.3</td>
</tr>
<tr>
<td>Sp-1382</td>
<td>Cytoplasmic alpha-amylase</td>
<td>No</td>
<td>1.6 (2.3)</td>
</tr>
<tr>
<td>Sp-1393</td>
<td>GfR family transcriptional regulator</td>
<td>Yes (40)</td>
<td>1.4 (1.3)</td>
</tr>
<tr>
<td>Sp-1395</td>
<td>Phosphate transport system regulatory protein (PhoU)</td>
<td>No</td>
<td>1.4 (1.5)</td>
</tr>
<tr>
<td>Sp-1590</td>
<td>Bacteriocin formation protein, putative (CylM protein)</td>
<td>No</td>
<td>1.5 (1.6)</td>
</tr>
<tr>
<td>Sp-2026</td>
<td>Bifunctional acetaldehyde-coenzyme A/alcohol dehydrogenase (Adh2)</td>
<td>Yes (40)</td>
<td>3.6 (8.4)</td>
</tr>
<tr>
<td>Sp-2052</td>
<td>Competence protein CglB</td>
<td>Yes (4, 12)</td>
<td>1.7 (18.1)</td>
</tr>
<tr>
<td>Sp-2073</td>
<td>Competence protein CglA</td>
<td>Yes (4, 12)</td>
<td>1.4 (1.9)</td>
</tr>
<tr>
<td>Sp-2106</td>
<td>Glycogen phosphorylase family protein</td>
<td>No</td>
<td>2.5 (3.3)</td>
</tr>
<tr>
<td>Sp-2107</td>
<td>4-Alpha-glucanotransferase</td>
<td>No</td>
<td>2.8 (2.8)</td>
</tr>
<tr>
<td>Sp-2108</td>
<td>Malto/maltodextrin ABC transporter, maltose/maltodextrin-binding protein (MalX)</td>
<td>Yes (10, 12, 28)</td>
<td>2.9 (6.8)</td>
</tr>
<tr>
<td>Sp-2109</td>
<td>Malto/maltodextrin transport system permease protein</td>
<td>No</td>
<td>1.7 (1.9)</td>
</tr>
<tr>
<td>Sp-2112</td>
<td>Maltope operon transcriptional repressor (MalR)</td>
<td>Yes (40)</td>
<td>1.3 (1.8)</td>
</tr>
<tr>
<td>Sp-2173</td>
<td>DltD protein</td>
<td>Yes (40)</td>
<td>1.4 (1.5)</td>
</tr>
<tr>
<td>Sp-2175</td>
<td>DltB protein</td>
<td>Yes (12)</td>
<td>1.4 (2.0)</td>
</tr>
<tr>
<td>Sp-2190</td>
<td>Choline binding protein A (CbpA)</td>
<td>Yes (12, 40)</td>
<td>1.4 (17.0)</td>
</tr>
</tbody>
</table>

* Gene IDs were obtained from the S. pneumoniae TIGR4 (serotype 4) genome as deposited in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database.
† Reference(s) for previous screens are in parentheses.
‡ Data in parentheses represent corresponding real-time RT-PCR values for the indicated genes.
DISCUSSION

The increase in the prevalence of antibiotic-resistant pneumococci and the existing and emerging shortcomings associated with current vaccine formulations have underscored the need for alternative approaches to management of pneumococcal disease. One of these strategies involves the development of vaccines based on pneumococcal proteins that contribute to pathogenesis and are common to all serotypes, to provide effective, affordable, and broad protection. Foremost among these antigens are choline binding protein A (CbpA, also known as PspC or SpsA), pneumococcal surface protein A (PspA), and pneumolysin (Ply) toxoid (2, 41, 43). These proteins have been shown to provide significant protective immunity alone; they work in synergy, and they serve as the gold standard for pneumococcal protein vaccine candidates (2, 34, 41–43). Nevertheless, protection using the current combination is by no means complete, and protein vaccine formulations are still being optimized.

An earlier study (40) used either in vitro surrogates for given in vivo niches (e.g., coculture with Detroit 562 cells in lieu of nasopharyngeal colonization; coculture with type II pneumocytes in lieu of lung infection), or cross-species comparisons of gene expression between distinct niches (e.g., blood of mice infected with D39 and adherence assay with Detroit 562 cells using a rough derivative of TIGR4), neither of which are appropriate models to investigate pneumococcal pathogenesis. Our mouse intranasal infection model mimics natural progression of invasive pneumococcal disease in humans, underscoring the superiority of our technique for identifying novel genes that might be critical to pneumococcal pathogenesis. The genome-wide transcriptomic comparisons of gene expression during transition of pneumococci from the nasopharynx to the lungs, and from there to the blood in the same animal, provide an alternative approach to characterizing niche-specific roles of virulence genes during pathogenesis. Such assessment gives insight into the optimization, design, and composition of surface protein combinations that might be more appropriate as vaccine antigens against carriage or invasive pneumococcal disease.

The analysis of gene expression profiles of pneumococci harvested from the nasopharynx compared with expression profiles of those harvested in the lungs revealed two important observations. First, the expression of many genes was higher in the nasopharynx compared with expression profiles of those harvested in the lungs, providing further molecular evidence in support of the hypothesis advanced by others that many pneumococcal virulence factors have dual roles in pathogenesis: establishment and maintenance of carriage, as well as involvement in certain stages of invasive pneumococcal disease.

Second, the relative expression profiles of many genes of WCH16 in the nasopharynx compared
Thus, we suggest that MalX is likely to play a role at mucosal surfaces: in nasopharynx versus lung comparisons. This strongly suggests limited regulation of pneumococcal gene expression during translocation from the lungs to the blood. Nevertheless, mutation in four of the five genes that were significantly upregulated in the blood relative to the lungs demonstrated either significant (for ΔaliA and ΔilvH mutants) or complete (for ΔbiO and ΔpiuA mutants) attenuation relative to the otherwise isogenic wild-type strain WCH43 in a mouse i.n. challenge model. Significant attenuation of the ΔaliA and ΔilvH mutants was further confirmed in a mouse i.n. competition assay, strongly suggesting a requirement for all these genes for optimal survival of pneumococci in the blood. The significant attenuation of the ΔaliA mutant in the blood is in contrast to that observed earlier in a D39 background (17), suggesting a strain-dependent contribution to virulence.

We then used an active i.p. immunization/challenge experiment to test the hypothesis that surface-accessible virulence factors that are upregulated in distinct host niches during pathogenesis are likely to be protective immunogens. For this purpose, we chose AliA, MalX, and PiuA for further analysis based on the presence of signal peptidase II cleavage sites at their N termini. We were particularly interested in protection against systemic disease, which is responsible for the significant mortality associated with S. pneumoniae. Significant protection was afforded by immunization with AliA and, particularly, with PiuA alone. PiuA is a lipoprotein first characterized as a highly conserved iron uptake ATP binding cassette (ABC) transporter required for bacterial growth and full virulence by Brown et al. (7) following identification by an STN screen (18). Intraperitoneal immunization challenge experiments showed that PiuA is protective against sepsis (8), and passive protection data obtained later confirmed this to occur via complement-dependent and -independent bacterial opsonophagocytosis, rather than by inhibition of iron transport (16). In this study, the level of protection observed for PiuA is better than what we obtained earlier using D39 as a challenge strain (8). The difference could be due to a higher challenge dose used for the D39 experiment but could also be strain dependent, as we proposed previously (34).

Western blot analysis confirmed that MalX, AliA, and PiuA, like PsaA, are located in the membrane compartment, as expected for lipoproteins. Thus, the likely explanation for the observed protection for PiuA and AliA is that exogenous antibody can penetrate the capsule/cell wall layer to a limited extent and bind the lipoproteins inhibiting their biological activity (nutrient transport), rather than promoting opsonophagocytosis. This is in agreement with the argument advanced by Whalan et al. for PiuA (51). The fact that labeling by FACS is very marginal could be due to limited penetration of both primary antibody as well as secondary antibody conjugate during the brief (30-min) incubation period employed. Nevertheless, labeling with anti-PiuA, anti-MalX, and anti-AliA was significantly greater than that with the negative-control serum when judged either by mean fluorescence intensity (Fig. 4B) or percent positive cells (result not shown).

Although the protection imparted by PdT in this study did not reach significance, immunization with PdT (or Pdb, a similar pneumolysin derivative) has been shown to confer protection against other virulent pneumococci (11, 34), and the protein is able to synergize with other protective immunogens in systemic challenge models (32, 34, 37). Given the fact that PiuA provided solid protection in this study, it would be of interest in a future study to investigate if a combination of PdT and PiuA would afford superior protection over each antigen alone in a high-dose challenge model.

FIG 2 Survival times for mice after i.n. challenge with WCH43 and isogenic mutant derivatives. Groups of 12 or 13 CD1 mice were challenged i.n. with approximately 1 × 10^7 CFU of the indicated strains. Each datum point represents one mouse. Upward arrows represent cohorts of pneumococcal genes upregulated in the lungs or blood. The horizontal broken lines denote the median survival time for each group. *, P < 0.05; ***, P < 0.001; Mann-Whitney U test; one and two tailed.
The current study provides further evidence that PiuA is a valid vaccine candidate against systemic pneumococcal disease, and it provides the first evidence that AliA is a promising protein vaccine candidate against systemic disease as well. We also provide proof-of-concept data that validate the *in vivo* microarray screen used to identify these promising vaccine candidates. The presence of elevated levels of antibody to PiuA in convalescent-phase sera of patients with pneumococcal septicemia, and the demonstration that the immune response elicited is serotype independent (51), supports this notion. This also strengthens our earlier suggestion that decisions on the best protein vaccine formulation that would proceed to clinical trials will require rigorous comparisons of individual antigens and all possible combinations thereof, using multiple mouse models and challenge strains (34). A careful examination of the literature shows that 15 of the 16 randomly selected genes (except the gene for CylM protein [Sp-1950]) listed in our analysis were identified by one or more, but not all, of the current screens for *in vivo*-expressed genes. This indicates that no single screening method is necessarily a stand-alone procedure but argues strongly in favor of our technique as valid, more robust, and more comprehensive, since it identified more candidate genes than other studies. Our technique also offers the advantage that it is less laborious and identifies key genes that are critical (either for survival or virulence) at each stage of the disease process. The screen also identified several other potentially interesting genes upregulated in discrete niches (at least 10 in the lungs and 8 in the blood) that are yet to be evaluated for their role in pathogenesis, and this presents avenues for future work. Such work would involve a thorough characterization of the critical regulatory pathways employed by the pneumococcus in discrete *in vivo* niches in accordance with host factors, nutrient availability, etc., and deployment of primary virulence determinants in those niches, leading to an improved understanding of pneumococcal pathogenesis, and potentially, identification of novel targets.
FIG 4 Membrane/surface localization studies. (A) Histograms showing marginal surface labeling of lipoproteins AliA, MalX, PiuA, and PsaA but substantial labeling of surface-localized PhtABDE with respective antisera. (B) Mean fluorescence data for AliA, MalX, PiuA, PsaA, and PhtABDE labeling with respective antisera, using alum-immunized serum as a negative control (**, P < 0.01; ***, P < 0.0001; unpaired t test; one tailed). (C, D, and E). Western blots of soluble (S) and membrane (M) fractions of S. pneumoniae WCH43 and its isogenic ΔaliA (C), ΔmalX (D), and ΔpiuA (E) mutants showing the reactivities of the antibodies raised against the various protein antigens. The samples were reacted with specific antisera generated from mice immunized with the various antigens; PsaA was used as a loading control and membrane compartment marker for AliA and MalX blots, and AliA was used as a marker for the PiuA blot. The expected mobilities and molecular mass of each test and marker protein are also indicated.
Influenzae type b, and recovery in other pathogens such as described here could also be easily applied to novel antigen discovery. The broken lines denote the median survival time for each group.

**ACKNOWLEDGMENTS**

This work was supported by Meningitis Research Foundation (United Kingdom) Research Grant 0802.0 to A.D.O., J.C.P., and L.K.M., as well as by National Health and Medical Research Council of Australia (NHMRC) Program Grant 565526 to J.C.P. and Project Grant 627142 to J.C.P. and A.D.O.

We acknowledge BµG@S (the Bacterial Microarray Group at St George’s, University of London) for supply of the microarray and advice and The Wellcome Trust for funding the multicollaborative microbial pathogen microarray facility under its Functional Genomics Resources Initiative. J.C.P. is an NHMRC Australia Fellow.


**REFERENCES**


