Harvey, Richard Manuel; Stroeher, Uwe Horst; Ogunniyi, Abiodun David; Smith-Vaughan, Heidi C.; Leach, Amanda J.; Paton, James Cleland

A variable region within the genome of Streptococcus pneumoniae contributes to strain-strain variation in virulence, PLoS One, 2011; 6(5):e19650

© 2011 Harvey et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

PERMISSIONS

http://www.plosone.org/static/license

Open-Access License

No Permission Required

PLOS applies the Creative Commons Attribution License (CCAL) to all works we publish (read the human-readable summary or the full license legal code). Under the CCAL, authors retain ownership of the copyright for their article, but authors allow anyone to download, reuse, reprint, modify, distribute, and/or copy articles in PLOS journals, so long as the original authors and source are cited. No permission is required from the authors or the publishers.

In most cases, appropriate attribution can be provided by simply citing the original article (e.g., Kaltenbach LS et al. (2007) Huntingtin Interacting Proteins Are Genetic Modifiers of Neurodegeneration. PLOS Genet 3(5): e82. doi:10.1371/journal.pgen.0030082). If the item you plan to reuse is not part of a published article (e.g., a featured issue image), then please indicate the originator of the work, and the volume, issue, and date of the journal in which the item appeared. For any reuse or redistribution of a work, you must also make clear the license terms under which the work was published.

This broad license was developed to facilitate open access to, and free use of, original works of all types. Applying this standard license to your own work will ensure your right to make your work freely and openly available. Learn more about open access. For queries about the license, please contact us.

4th July 2013

http://hdl.handle.net/2440/66433
A Variable Region within the Genome of *Streptococcus pneumoniae* Contributes to Strain-Strain Variation in Virulence

Richard M. Harvey¹, Uwe H. Stroeher¹*, Abiodun D. Ogunniyi¹, Heidi C. Smith-Vaughan², Amanda J. Leach², James C. Paton¹*

¹ Research Centre for Infectious Diseases, School of Molecular and Biomedical Science, University of Adelaide, Adelaide, South Australia, Australia, ² Menzies School of Health Research, Charles Darwin University, Darwin, Northern Territory, Australia

Abstract

The bacterial factors responsible for the variation in invasive potential between different clones and serotypes of *Streptococcus pneumoniae* are largely unknown. Therefore, the isolation of rare serotype 1 carriage strains in Indigenous Australian communities provided a unique opportunity to compare the genomes of non-invasive and invasive isolates of the same serotype in order to identify such factors. The human virulence status of non-invasive, intermittently virulent and highly virulent serotype 1 isolates was reflected in mice and showed that whilst both human non-invasive and highly virulent isolates were able to colonize the murine nasopharynx equally, only the human highly virulent isolates were able to invade and survive in the murine lungs and blood. Genomic sequencing comparisons between these isolates identified 8 regions >1 kb in size that were specific to only the highly virulent isolates, and included a version of the pneumococcal pathogenicity island 1 variable region (PPI-1v), phage-associated adherence factors, transporters and metabolic enzymes. In particular, a phage-associated endolysin, a putative iron/lead permease and an operon within PPI-1v exhibited niche-specific changes in expression that suggest important roles for these genes in the lungs and blood. Moreover, in vivo competition between pneumococci carrying PPI-1v derivatives representing the two identified versions of the region showed that the version of PPI-1v in the highly virulent isolates was more competitive than the version from the less virulent isolates in the nasopharyngeal tissue, blood and lungs. This study is the first to perform genomic comparisons between serotype 1 isolates with distinct virulence profiles that correlate between mice and humans, and has highlighted the important role that hypervariable genomic loci, such as PPI-1v, play in pneumococcal disease. The findings of this study have important implications for understanding the processes that drive progression from colonization to invasive disease and will help direct the development of novel therapeutic strategies.

Introduction

*Streptococcus pneumoniae* (the pneumococcus) is a leading cause of bacterial pneumonia, invasive disease (bacteremia and meningitis [IPD]) and otitis media, and is responsible for bacterial pneumonia, invasive disease in both humans and mice when compared to certain other serotypes and clones to cause IPD varies and has led to the grouping of serotypes and clonal clusters according to invasive potential [2–5]. In particular, serotype 1 pneumococci have repeatedly been reported to have a high invasive potential due to the rarity of asymptomatic carriage [2,5]. Furthermore, serotype 1 isolates frequently cause disease in patients without an underlying illness and as such behave as a primary pathogen [6]. In spite of this, whilst serotype 1 isolates have a high invasive potential, commonly studied serotype 1 clones tend to cause less severe disease in both humans and mice when compared to certain other serotypes and clones that behave as opportunistic pathogens [4,6]. However, a number of clones from the ST217 clonal cluster (CC217) have been responsible for African epidemics of IPD with unusually high mortality rates, and are considered to be hypervirulent. In contrast to these hypervirulent clones, relatively high rates of serotype 1 asymptomatic carriage have been reported in a number of communities following vaccination with the 7-valent conjugate vaccine [7,8]. In particular, serotype 1 carriage by ST304 and ST227 clones was detected in a number of remote Indigenous communities in Australia, without an associated increase in serotype 1 IPD in the same communities [8]. Therefore, it is clear that considerable variation in virulence exists between strains of the same serotype, which in turn highlights the contribution that serotype-independent factors play in virulence. Of particular interest is the existence of many ‘accessory regions’ (AR) within the pneumococcal genome that may contribute to such differences in virulence [9–14]. However, whilst a number of potential virulence determinants such as the pilus encoded within...
the nblA islet and the pneumococcal serine rich repeat protein (PsrP) have received particular attention, little consistency between the presence of specific virulence determinants and invasive potential has been found in large-scale comparisons [10,14–21]. Nevertheless, it is possible that such large-scale comparisons fail to take into account the significant differences in virulence that may exist within groups of isolates with apparently equivalent invasive potential, such as hypervirulent and moderately virulent serotype 1 clones [14]. In addition, comparisons across serotypes risk underestimating the impact of the serotype itself on virulence, due to serotype-specific structural differences in the capsule that can affect complement deposition and resistance against phagocytosis [22–25]. Therefore, by comparing the genomes of non-invasive and invasive serotype 1 isolates, it is possible to identify serotype-independent factors that alter the outcome of infection.

Initially, the human virulence status of non-invasive and invasive serotype 1 isolates was confirmed in mouse models of infection, enabling these isolates to be grouped as non-invasive, intermediately virulent and highly virulent. Using these virulence profiles as a basis, preliminary comparisons between the three virulence phenotypes were performed using comparative genomic hybridization (CGH), and next generation genome sequencing technology. These comparisons identified a number of previously described ARs as well as new ARs that were present only in the highly virulent isolates. Of particular significance was that the highly virulent isolates harbored a version of the pneumococcal pathogenicity island 1 (PPI-1) that conferred greater competitiveness in vivo than the versions in less virulent isolates.

**Results and Discussion**

The virulence of serotype 1 isolates in mouse models of infection mimics human disease

An intraperitoneal (i.p.) mouse challenge model identified three distinct virulence phenotypes by comparing the virulence of a collection of serotype 1 isolates (Fig. 1A). Menzies1-1 (strain 1) and Menzies1-2 (strain 2) were avirulent in mice, which mimicked their non-invasiveness in humans. Menzies1-1861 (strain 1861) and WCH4496 (strain 4496) were highly virulent in mice, as mice infected with these strains had significantly reduced survival times than those challenged with the other two invasive isolates. Menzies1-3415 (strain 3415) and Menzies1-5482 (strain 5482) were considered to be intermediately virulent, as they were not carried asymptomatically in humans, but they were less virulent than strains 1861 and 4496 in mice. The virulence of strains 3415 and 5482 appears to be similar to the relatively low virulence of most serotype 1 isolates in mice that has been reported previously [4,6]. An intranasal (i.n.) mouse challenge model also confirmed the differences in virulence between the non-invasive strain 1 and the highly virulent strains 1861 and 4496 (Fig. 1B). A feature of the highly virulent isolates was their ability to either cause rapidly fulminant infection within 60 h of challenge, or not cause detectable disease at all. Interestingly, daily analysis of bacteremia in all mice revealed that in surviving mice pneumococci did not reach a detectable level within the blood at any time (data not shown). Further characterization of the pathogenicity of strains 1, 1861 and 4496 was performed by comparing the recovery of pneumococci in the nasopharynx, blood and lungs at both 48 h and 96 h post-challenge using an i.n. mouse challenge model (Fig. 2). At 48 h a small but significant difference was observed between the number of pneumococci in the nasopharynx between strains 1861 and 4496 (Fig. 2A[i]). However, there was no significant difference in the number of bacteria recovered from the nasopharynx of strain 1-infected mice when compared to either of the highly virulent strains at either timepoint (Fig. 2A[ii]). No pneumococci were recovered from either the blood or lungs of strain 1-infected mice, which was in stark contrast to the numbers recovered from both strain 1861- and 4496-infected mice at 48 h (Fig. 2B[i] and C[i]). Whilst there was a small, but significant difference in the level of bacteremia between strains 1861 and 4496, this difference was minimal compared to strain 1. As shown in Figure 2, mice challenged with either of the highly virulent strains either develop fulminant infection within 60 h or survive the challenge completely. Therefore, the strain 1861- and 4496-infected mice analyzed at 96 h represent surviving mice, and unsurprisingly lack detectable numbers of pneumococci in either

---

**Figure 1. Confirmation of virulence in mice of non-invasive and invasive serotype 1 isolates.** Survival times were recorded following i.p. (A) or i.n. (B) challenge with approximately $10^6$ CFU and $10^7$ CFU of the relevant strain, respectively. Experiments were undertaken for 216 h and 252 h following i.p. and i.n. challenge, respectively. Horizontal broken lines indicate the median survival time. Statistical significance was calculated using the two-tailed Mann-Whitney U test. *P*, P < 0.05). ** indicates comparison with non-invasive isolates; *** indicates comparison with intermediately virulent isolates.

doi:10.1371/journal.pone.0019650.g001
the blood or lungs (Fig. 2B[iii] and 2C[iii]). The fact that a proportion of mice completely survive i.n. challenge is most likely a consequence of mouse-mouse variation in the actual number of pneumococci aspirated into the lungs immediately following i.n. challenge. However, most importantly the key difference between the non-invasive and highly virulent phenotypes is the ability of the latter to invade and survive in the blood and lungs. We used MLST to examine whether the strains with differing virulence profiles had any clonal relationship. Strains 1 and 2 (ST304), and strains 3413 and 5482 (ST227), belong to the lineage A of serotype 1 clones [26], whereas the highly invasive strains 1861 (ST3079) and 4496 (ST3018) belong to lineages B and C, respectively. In particular, strain 1861 was found to be a single-locus variant of ST217 and a double-locus variant of ST618 and ST303, which were the dominant clones responsible for epidemics of severe pneumococcal pneumonia in humans in recent years [27–30]. Therefore, the heightened virulence of strain 1861 in mice is consistent with the severity of disease in humans caused by clonally related strains. A summary of the virulence and MLST data of the serotype 1 isolates used in this study is shown in Table 1.

Genetic differences between the non-invasive, intermediately virulent and highly virulent strains were identified by genomic sequencing

Genome comparisons were performed between the serotype 1 isolates in order to identify ARs present in the highly virulent strains that might be responsible for their heightened virulence. Genomic comparisons were initially performed by sequencing the genomes of strains 1 and 1861. The presence of regions >1 kb in size that were present only in strain 1861 was subsequently tested in all six serotype 1 isolates by PCR to identify those regions associated with heightened virulence. The primers used in these PCRs are listed in Table S1. In this study, we focused attention on regions >1 kb, as these were likely to encode an intact gene product, and probably represent horizontally acquired genetic material. Of course, the importance of smaller regions (<1 kb), which might encode bacteriocins, signaling peptides, or regulatory RNAs, as well as single nucleotide polymorphisms (SNPs) in coding and non-coding regions should not be discounted. Nevertheless, the expansive pool of ARs available to the pneumococcal genome warranted paying particular attention to these larger regions. The presence of key regions that have previously been associated with virulence in other studies but not associated with virulence in this study was confirmed by comparative genomic hybridization (CGH). The P1031 (ST303; lineage B) and INV104B (ST227; lineage A) genome sequences were most closely aligned with strains 1861 and 1, respectively.

Regions associated with heightened virulence

Eight regions >1 kb in size were identified in both of the highly virulent isolates but absent in all four less invasive isolates (Table 2). A detailed list of each gene present in each AR is included in Table S2. Of these regions, those designated 1, 2, 5, 6 and 7 have not previously been identified as ARs [14].

Figure 2. Number of \textit{S. pneumoniae} CFU recovered from the nasopharynx, lungs and blood of infected mice. The number of CFU recovered from the nasopharynx (A), lungs (B) and blood (C) of CD1 mice was determined at 48 h (i) and 96 h (ii) post-challenge. The horizontal bars indicate the geometric mean number of CFU that were recovered. Statistical differences were analyzed by two-tailed unpaired t-test on log-transformed values (*, P<0.05; **, P<0.01; ***, P<0.001). ‘’ indicates comparison with 4496 and ‘++’ indicates comparison with 1861. The single horizontal spotted line indicates the limit of detection (LD) which equates to 102 CFU/nasopharynx, 2 x 102 CFU/lung and 102 CFU/ml blood.

doi:10.1371/journal.pone.0019650.g002

| Table 1. Serotype 1 isolates used in this study. |
|---|---|---|
| Isolate | ST (Lineage*) | Virulence status* |
| Menzies* | 304 (A) | Non-invasive |
| Menzies | 304 (A) | Non-invasive |
| Menzies* | 227 (A) | Intermediately virulent |
| Menzies | 227 (A) | Intermediately virulent |
| Menzies* | 3079 (B) | Highly virulent |
| WCH4496 | 3018 (C) | Highly virulent |

*Lineage is as described in Brueggemann et al. [26].
*Non-Indigenous isolate. Other isolates were all of Indigenous origin.

Non-invasive isolates were both non-invasive in humans and mice; intermediately virulent strains were virulent in humans, but significantly less virulent in mice than strains 1861 and 4496. Highly virulent strains were significantly more virulent than the other four strains.

doi:10.1371/journal.pone.0019650.t001

A Pathogenicity Island and Pneumococcal Disease

Regions associated with heightened virulence

Eight regions >1 kb in size were identified in both of the highly virulent isolates but absent in all four less invasive isolates (Table 2). A detailed list of each gene present in each AR is included in Table S2. Of these regions, those designated 1, 2, 5, 6 and 7 have not previously been identified as ARs [14].

Region 1 consists of a temperate bacteriophage genome with greatest homology to \textit{SPP.0028–0084} in P1031 and to a lesser extent the \textit{Streptococcus oralis} PH10 phage [30]. The prophage is inserted into a position between genes homologous to \textit{SP.0019} (adenylsuccinate synthetase) and \textit{0020} (cytidine/deoxycytidylate deaminase) in TIGR4 and \textit{SPN104.00170} and \textit{00100} in INV104. In addition, examination of the sequence of the integrase gene (\textit{SPP.0028}) suggests that the phage in strains 1861 and 4496 belongs to the group 1 pneumogres [31,32]. This prophage encodes genes homologous to platelet-binding protein B (PblB) and an endolysin, both of which are required for the virulence of \textit{Streptococcus mitis} in an animal model of infective endocarditis [33]. PblB and the endolysin have been shown to be required by \textit{S. mitis} for adherence to human platelets [34–36]. However, a similar role for the products of these genes in \textit{S. pneumoniae} has yet to be demonstrated. Interestingly, the endolysin gene shares 80% nucleotide sequence identity with the major autolysin (N-acetylglucosamyl-L-alanine amidase; \textit{lytA}), which is an important pneumococcal virulence factor [37–40]. Of the publicly available genome sequences, only strain P1031 contains sequences homologous to the full length of region 1 represented by the prophage. However, Hungary\textsuperscript{34,35}, JJA, 70358, OXC141, SPN994039, SPN034183, SPN994038, and SPN034156 possess pblB- and endolysin-(in addition to \textit{lytA}) like genes at a similar genomic location.

Region 2 encodes a putative MerR family transcriptional regulator and MuT/Nudix family protein homologous to \textit{SPP.0750} and \textit{0731}, respectively. In addition, the gene encoding a putative sodium-dependent permease was truncated in the non-invasive and intermediately virulent strains due to the absence of region 2. However, it is not clear whether the gene is functional in the highly virulent strains due to a frameshift mutation approximately 220 bp downstream of the start codon. Region 2 is inserted among genes homologous to \textit{SPN104.06130} and \textit{06140} in INV104. The genes in region 2 are also present in numerous other genomes including TIGR4 (\textit{SP.0737–0740}), D39, ATCC 700669, JJA, 70358, Taiwan\textsuperscript{10–14}, G54, CGSP14, INV200, SPN034156, OXC141, SPN994039, SPN034183, SPN994038 and TCH0431/19A. A MerR family transcriptional
regulator in *S. pneumoniae* (*SP_1856*) has been shown to be involved in nitric oxide stress and is required for full systemic virulence [41].

Regions 3 and 4 consist of the previously described AR 22 [14], also known as the PPI-1 variable region (PPI-1v). PPI-1v appears to be a hotspot for recombination and in some strains contains the PezAT toxin-antitoxin (TA) system, which has been implicated in virulence [42–44]. An alignment between PPI-1v from the lineage A isolates and the highly virulent isolates shows the relative position of regions 3 and 4 and is consistent with the region in ATCC 700669 and G54 (Fig. 3). In addition, this alignment shows that PPI-1v consists of two variable components: the pezAT region (region 3) and an accessory region (region 4) (Fig. 3). In this study, *pezAT* was present only in the highly virulent isolates and not in either the intermediately virulent or non-virulent isolates (Table 2). In addition to *pezAT* itself, the non-invasive and intermediately virulent isolates lack most of the neopullulanase locus (*SP_1046*–*1058*, *nptI*). However, since *nptI* is fragmented in strains 1861 and 4496, it is unlikely that a functional protein is produced by these strains. The *pezAT* region is present in numerous other *S. pneumoniae* genomes; A, ATCC 700669; B, JJA; C, 70585; D, Taiwan19F-14; E, Hungary19A-6; F, G54; G, CGSP14.

### Table 2. Genomic regions present only in the genomes of the highly virulent isolates.

<table>
<thead>
<tr>
<th>Region (AR)*</th>
<th>Size</th>
<th>Putative annotation/function of key gene(s) in region</th>
<th>Homologous ORFs</th>
<th>Other genomesβ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>P1031</td>
<td>TIGR4</td>
</tr>
<tr>
<td>1</td>
<td>35 kb</td>
<td>Group 1 pneumophagae (pbp8-like, endolysin)</td>
<td>0008–9535</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1 kb</td>
<td>Sodium-dependent permease, transcription regulator (MerR)</td>
<td>0750–0751</td>
<td>0739–0740</td>
</tr>
<tr>
<td>3 (22)</td>
<td>6.3 kb</td>
<td>PPI-1v (PezAT, NpIT)</td>
<td>1049–1058</td>
<td>1046–1056</td>
</tr>
<tr>
<td>4 (22)</td>
<td>8 kb</td>
<td>PPI-1v (Metabolic/hypothetical)</td>
<td>1059–1069</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>58 kb</td>
<td>ZmpD, Tn5253 (TA system, UmuCD)</td>
<td>1141–1198</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1.7 kb</td>
<td>High-affinity iron/lead permease</td>
<td>1340–1343</td>
<td>1155–1157</td>
</tr>
<tr>
<td>7</td>
<td>2.9 kb</td>
<td>ABC-type transporter, transcription regulator (ArsR)</td>
<td>1637–1643</td>
<td></td>
</tr>
<tr>
<td>8 (28)</td>
<td>18 kb</td>
<td>ABC transporter, sialic acid degradation enzymes</td>
<td>1350–1371</td>
<td>1346–1349</td>
</tr>
</tbody>
</table>

*AR designated by Bloomberg et al. [14].

βPresence in other *S. pneumoniae* genomes: A, ATCC 700669; B, JJA; C, 70585; D, Taiwan19F-14; E, Hungary19A-6; F, G54; G, CGSP14.

A detailed list of genes present in each AR is included in Table S2. doi:10.1371/journal.pone.0019650.t002

However, unlike strains 1861 and 4496, the genomes of G54, 11-B570 and MLV-016 lack *pezAT*. The PPI-1v accessory region in the non-invasive and intermediately virulent isolates contains a fragmented lantibiotic modification and export gene, and a putative lantibiotic immunity system ABC transporter (Table 3). However, the region lacks the structural gene for mersacidin lantibiotic and lantibiotic modifying enzymes. This version of PPI-1v has previously been described in the *pezAT*-positive strain ATCC 700669 [43] and is homologous to ORFs *SPINV104_09130–09190* in the *pezAT*-negative strain INV104. This version of the region is also present in the *pezAT*-positive strains JJA, SPN030308 and SPN032672.

Region 5 includes the putative zinc metalloprotease D (ZmpD) and Tn5253, which are homologous to *SPP_1141–1198* in P1031. However, this region is not present in TIGR4 or INV104 and is inserted between homologues of *SP_1154* and 1155, and *SPINV104_09960* and 09970, respectively. Other pneumococcal Zmps include the IgA protease and ZmpC, which function to cleave host IgA and activate human matrix metalloprotease 9, respectively, and thus contribute to virulence [46–52]. However, as the substrate of ZmpD is unknown, it is not clear whether this enzyme plays a role in virulence. Other strains harboring *zmpD* include ATCC 700669, JJA, G54, CGSP14 and INV200, and the gene was detected in 49% of 218 mostly invasive disease isolates in one study [53]. The large conjugative transposon Tn5253 is a composite element consisting of regions from Tn916 and Tn5252 and contains genes that confer resistance to tetracycline and chloramphenicol [34,55]. In the highly virulent serotype 1 isolates, the region encodes a putative TA system homologous to PezAT in PPI-1 and an *umuCD*-like operon. *UmuCD* is involved in a branch of the SOS response in *E. coli* and in *S. pneumoniae* has been shown to confer greater UV tolerance [56,57]. Whilst a number of other strains contain distinct Tn5252 and Tn916 elements with homology to region 5, ATCC 700669 also harbors the composite Tn5253 element [43].

Region 6 encodes a putative high-affinity iron/lead permease and a fragmented DyP-type peroxidase that are homologous to *SPP_1340* and 1341, respectively in P1031. This region is inserted between homologues of *SPINV104_11210* and 11240 in INV104 and *SP_1299* and 1306 in TIGR4. In both TIGR4 and INV104 the corresponding region consists of a number of small hypothetical genes. Whilst the substrate of the permease in region
6 requires experimental confirmation, other pneumococcal transporters of metal ions such as Psa (manganese) and Pia (iron) are important in pneumococcal pathogenesis [58–63]. The putative metal permease is also present in the genomes of D39, Hungary3–6, CGSP14, INV200 and CDC3059-06.

Region 7 encodes a putative ArsR family transcriptional regulator, and an ABC-2 type transporter including the transmembrane and ATP-binding components homologous to SPP_1779 and 1780 in P1031, respectively. The region is absent in both TIGR4 and INV104, and if present would be inserted between SP_1779 and 1780, and SPINV104_15230 and 15250, respectively. ArsR family transcriptional regulators are often responsive to metal ions, and like MerR regulators, have been implicated in resistance against environmental stresses [64]. Examination of other pneumococcal genome sequences revealed that strains MLV-016, 11-BS70, 9-BS68, 14-BS69 and 18-BS74 also harbor this region.

Region B is homologous to SPP_1350–1367 in P1301, which corresponds to the previously described AR 28 [14]. In particular, the AR 28 subregion RD8b1 [10,65] encodes a putative ABC transporter for the transport of glutathione and putative N-acetylmannosamine-6-phosphate 2-epimerase, kelch-like protein, glycoside hydrolase family protein and N-acetylneuraminate lyase genes. Much of this region is also present in D39 (SPD_1164–1174). A putative ABC transporter is also encoded within the same, but divergent region in the non-invasive and intermediately virulent isolates (Table 3). The transporter appears to be similar to the LplABC polysaccharide transporter in Bacillus subtilis and Agrobacterium radiobacter [66]. Glycoside hydrolase and N-acetylmannosamine-6-phosphate epimerase genes present within the region in both groups of strains share 50% nucleotide sequence identity. Putative diadenosine tetraphosphate hydrolase and dihydrolipoamide dehydrogenase genes also exist in this region in the intermediately virulent and non-invasive isolates. Furthermore, the AR 28 subregion RD8b2 that is homologous to SP_1345–1349 in TIGR4 and includes a second ABC transporter [10,65] is present in the non-invasive and intermediately virulent isolates, but absent in the highly virulent isolates. The RD8b3

Figure 3. Alignment of PPI1v between ATCC 700669, 1, 1861 and G54. The nucleotide sequences of PPI-1v were aligned using the artemis comparison tool between ATCC 700669, 1, 1861 and G54. Strain 1 represents the non-invasive and intermediately virulent isolates, and strain 1861 represents the highly virulent isolates. Blue shading between strains indicates nucleotide sequence identity exceeding 90%. Green arrows indicate the npIT ORFs, red arrows indicate the pezAT operon, orange arrows indicate ORFs within divergent sequence that are absent in strain 1, yellow arrows indicate ORFs that are absent in strain 1861 and purple indicates ftsW. Tn5252 ORFs are not shown.

doi:10.1371/journal.pone.0019650.g003

Figure 4. General component structure of PPI-1v. Summary of PPI-1v derived from alignments of the region between ATCC 700669, 1, 1861 and G54. Blue indicates regions of shared sequence between the aligned strains, red indicates the pezAT region and yellow indicates a region of divergent sequence (accessory region).

doi:10.1371/journal.pone.0019650.g004
subregion that is homologous to SP_1343–1349 and was reported to be required for wild-type virulence in TIGR4 [65], was present in all six serotype 1 isolates in this study. The overall region in the non-invasive and intermediately virulent isolates is homologous to INV104_11270–11410 in INV104.

Key regions found not to be associated with a virulence phenotype in this study

Previous genomic comparisons have identified a number of ARs that were associated with clonal clusters with high invasive potential. While it is likely that a number of these regions are involved in virulence, it is unlikely that they are responsible for the differences in virulence that we observed between the strains in this study. ARs 5, 6, 9, 15, 20, 21, 29 and 39 were reported to be present in the majority of tested isolates from highly invasive clonal clusters [14]. In particular, ARs 5, 6, 15 and 29 have previously been shown to impact virulence using signature tagged mutagenesis (STM) [67,68]. However, AR 5 is absent in all six isolates in this study (Table 4), and AR 6 is present only in the lineage A isolates and not in the highly virulent isolates. ARs 9, 15, 20, 21, 29 and 39 were present in all six strains, which is not surprising since Blomberg et al. [14] found that these regions are present in the majority of serotype 1 isolates that were tested in their study [14]. However, this finding implies that these regions are not responsible for the virulence differences between the isolates in this study. In addition, AR 6 is also not likely to be required for the heightened virulence of strains 1861 and 4496. Furthermore, ARs 10, 16, 19 and 27 were shown to be present in some serotype 1 isolates, and not others [14]. In this study ARs 16 and 19 were present in all isolates and AR 10 was present in the lineage A isolates, but not the highly virulent isolates (Table 4). Similar to Blomberg et al., [14] the presence of AR 27 did not appear to correlate with the invasive potential of the strains in this study (Table 4) [10,14]. Whilst the deletion of AR 31 has previously been reported to significantly reduce the virulence of TIGR4 [65], the region was present in strain 1861, but not in either strain 4496 or the lineage A strains. Therefore, the region is unlikely to be responsible for the differences in virulence between serotype 1 isolates in this study. In addition, ARs 11, 30 and 34, which encode the pilus, pneumococcal collagen-like protein (PclA) and PsnP, respectively, have been implicated in virulence [10,16,19,69]. However, AR 11 was absent in all six isolates, and AR 34 was present in the intermediate virulent isolates and strain 1861, but not in the non-invasive isolates or strain 4496. AR 30 was present only in the intermediately virulent isolates. Therefore, these regions are unlikely to be important for the differences in virulence between the isolates in this study. In addition to the pilus encoded within the n/a islet, there is a second pilus (PI-2), which has been reported to be present in emerging serotypes such as 1, 2, 7F and 19A and is homologous to SPT_1056–1064 in Taiwan19F-14 [70]. However, genome sequencing and PCR (data not shown) showed that PI-2 is present in the non-invasive and intermediately virulent isolates, but absent in the highly virulent isolates. Therefore, possession of PI-2 is not required for the heightened virulence of strains 1861 and 4496.

Differential in vivo expression of key virulence-associated genes

Genes associated with heightened virulence (Tables 2 and 3) were selected for in vivo gene expression comparisons to identify genes that exhibit niche-specific changes in expression. Expression analysis was performed on nasal lavage, blood and lung samples of strain 1861- and 4496-infected mice (Figure 2), using qRT-PCR. Expression of the endolysin, nplT, 3HIBDH and iron/lead permease was elevated in the blood and lungs of both strain 1861- and 4496-infected mice, when compared to the nasopharyngeal lavage fluid (Table 5), which suggests that the products of these genes are more important in the blood and lungs than the nasopharyngeal surface. Interestingly, since endolysin activity has been shown to be required for Phib surface expression in S. mitis [33], surface expression of this potential adherence factor could be indirectly elevated in the lungs and blood due to endolysin activity, despite little difference in phib expression between niches. In PPI-1v, 3HIBDH is part of an operon that also encodes PDt and galE (data not shown), which implies that the expression of all three of these enzymes is elevated in the lungs and blood compared to the nasopharyngeal surface. While the activities of these genes remain to be confirmed experimentally, the products of these genes could provide a survival advantage and facilitate disease. Interestingly, expression of the iron/lead permease was greatest in the blood followed by the lungs and lowest on the nasopharyngeal surface, which suggests that this transporter is more important in the blood and lungs than on the nasopharyngeal mucosa. Expression of the major facilitator transporter in PPI-1v and the glycoside hydrolase in region 8 was significantly greater in the blood than either the lungs or nasopharyngeal surface. However, since the substrate of the major facilitator is unknown it is not clear whether the expression of this gene provides a survival advantage in this niche. As the glycoside hydrolase is encoded immediately downstream of the putative ABC transporter in region 8, and no expression-attenuating secondary structures were predicted in the intervening sequence, it is likely that the expression of this enzyme reflects the expression of the ABC transporter. Expression of the sodium-dependent permease, pezAT and biotin carboxylase expression in PPI-1v did not appear to be niche-specific since changes in the expression of these genes were not consistent between the two strains.
Table 4. Other notable regions not consistently associated with a virulence profile in this study.

<table>
<thead>
<tr>
<th>AR*</th>
<th>Putative annotation/function of key gene(s) in region</th>
<th>Homologous ORFs</th>
<th>Serotype 1 isolates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>TIGR4</td>
<td>D39</td>
</tr>
<tr>
<td>5</td>
<td>Hypothetical proteins</td>
<td>0296–0298</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>β-glucosidase, PTS – metabolic</td>
<td>0300–0310</td>
<td>0276–0283</td>
</tr>
<tr>
<td>9</td>
<td>Mevalonate pathway - metabolism</td>
<td>0382–0387</td>
<td>0347–0390</td>
</tr>
<tr>
<td>10</td>
<td>Mannitol PTS</td>
<td>0394–0399</td>
<td>0360–0364</td>
</tr>
<tr>
<td>11</td>
<td>RIIA islet - adherence</td>
<td>0461–0468</td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td>β-galactosidase – metabolism</td>
<td>0643–0648</td>
<td>0559–0562</td>
</tr>
<tr>
<td>16</td>
<td>Zinc metalloproteinase B</td>
<td>0664–0666</td>
<td>-</td>
</tr>
<tr>
<td>19</td>
<td>Type I restriction modification system</td>
<td>0887–0890</td>
<td>-</td>
</tr>
<tr>
<td>20</td>
<td>Amino acid metabolism</td>
<td>0918–0923</td>
<td>0811–0816</td>
</tr>
<tr>
<td>21</td>
<td>CelA – competence</td>
<td>0949–0954</td>
<td>0839–0843</td>
</tr>
<tr>
<td>27</td>
<td>V-type sodium ATP synthase</td>
<td>1315–1331</td>
<td>-</td>
</tr>
<tr>
<td>29</td>
<td>ABC transporter</td>
<td>1432–1442</td>
<td>1261–1272</td>
</tr>
<tr>
<td>30</td>
<td>Collagen-like protein, PclA – adherence</td>
<td>1612–1620</td>
<td>-</td>
</tr>
<tr>
<td>31</td>
<td>Ribulose PTS – uptake of ribulose</td>
<td>1930–1936</td>
<td>1731–1736</td>
</tr>
<tr>
<td>39</td>
<td>Type II restriction modification system</td>
<td>220.30*c</td>
<td>1.27</td>
</tr>
</tbody>
</table>

*AR designated by Bloomberg et al. [14].

+, Region present; –, Region absent; Determined from genome sequence of strains 1 and 1861, and by CGH for strains 2, 3415, 5482 and 4496. Presence of PI-2 was determined by PCR.

doi:10.1371/journal.pone.0019650.t004

Table 5. Relative expression of selected virulence associated genes between different niches of the mouse.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Region*</th>
<th>Expression*</th>
<th>Blood vs. nose</th>
<th>Lungs vs. nose</th>
<th>Blood vs. lungs</th>
</tr>
</thead>
<tbody>
<tr>
<td>pblB</td>
<td>1</td>
<td>+1.22**</td>
<td>+1.79**</td>
<td>+1.17**</td>
<td>-2.43**</td>
</tr>
<tr>
<td>Endolysin</td>
<td>1</td>
<td>+1311.20**</td>
<td>+221.32**</td>
<td>+1028.74**</td>
<td>+220.30**</td>
</tr>
<tr>
<td>Na+ dep. transporter</td>
<td>2</td>
<td>-1.28**</td>
<td>-1.28**</td>
<td>-3.15**</td>
<td>-1.54</td>
</tr>
<tr>
<td>npIT</td>
<td>3</td>
<td>+55.72**</td>
<td>+42.62**</td>
<td>+28.38**</td>
<td>+69.07**</td>
</tr>
<tr>
<td>pezAT</td>
<td>3</td>
<td>+1.06**</td>
<td>+11.71*</td>
<td>-1.70**</td>
<td>+1.14**</td>
</tr>
<tr>
<td>3HIBDH</td>
<td>4</td>
<td>+64.59**</td>
<td>+32.53**</td>
<td>+115.89**</td>
<td>+97.46**</td>
</tr>
<tr>
<td>Biotin carboxylase</td>
<td>4</td>
<td>-1.63**</td>
<td>+97.01*</td>
<td>-2.61**</td>
<td>+9.78**</td>
</tr>
<tr>
<td>Major facilitator</td>
<td>4</td>
<td>+3.27*</td>
<td>+10.95*</td>
<td>+1.08**</td>
<td>+2.04*</td>
</tr>
<tr>
<td>Fe**/Pb** permease</td>
<td>6</td>
<td>+5.92*</td>
<td>+24.36*</td>
<td>+2.17**</td>
<td>+12.10*</td>
</tr>
<tr>
<td>Gly. Hydrolase</td>
<td>8</td>
<td>+29.18**</td>
<td>369.65**</td>
<td>-624.55**</td>
<td>+25.22**</td>
</tr>
</tbody>
</table>

*Indicates where the target mRNA was below the limit of detection in nasal wash-derived RNA.

**Indicates where the target mRNA was below the limit of detection in both samples.

The values represent the relative amounts of mRNA in the first niche compared to the second. Results of statistical analysis using t-test: ns, not significant (includes values <2); *P<0.05; **P<0.01; ***P<0.001.

As numbered in Table 2.

doi:10.1371/journal.pone.0019650.t005
In vivo competition between D39 PPI-1 mutants replacing the endogenous version of the region with the regions of strain 1 and strain 1861

Since the content of PPI-1v varies between the highly virulent isolates and the four less virulent isolates, and the expression of a number of PPI-1v genes was favored in niches associated with disease, it was decided to confirm the role of this region in virulence determination by mutagenesis. However, a significant roadblock to such experiments was the inability to genetically transform the serotype 1 isolates in this study, despite numerous attempts using various known transformation protocols. Therefore, it was decided to construct PPI-1v derivatives in the easily transformable laboratory strain D39 [71], which has a different version of PPI-1v compared to strain 1 and 1861. Derivatives of PPI-1v were constructed and used to replace the endogenous region with the version of the region in the highly virulent isolates (designated D391861) and the version in the non-invasive isolates (designated D391). PPI-1v in D391 was also representative of that present in the intermediately virulent isolates. In addition, a PPI-1v deletion mutant (D39APPI-1) was constructed to examine the contribution of the wild-type D39 version of PPI-1v to virulence.

In order to obtain a D39APPI-1 knockout, an intermediate mutant lacking pezT (D39APezT) was constructed by replacing SPD_0931–0951 with ermK, as simultaneous deletion of pezA and pezT has been reported to be lethal [44]. The subsequent D39APPI-1 mutant was constructed from D39APezT by replacing SPD_0927–0930 with ermK. D39v was also constructed from D39APezT by replacing the remaining non-homologous D39 PPI-1v sequence (SPD_0927–0930) with strain 1 sequence (SPINV104_09310–09190) (Fig. 5A). D391861 was constructed by replacing the non-homologous D39 PPI-1v sequence (SPD_0936–0951) with strain 1861 PPI-1v sequence (SPP_1056–1072) (Fig. 5B). Expression of 1861-derived PPI-1v genes in D391861 was confirmed in vitro by qRT-PCR (data not shown). In addition, there was no detectable difference in in vitro growth rate between the mutants and D39 (data not shown). The ability of the mutants and wild-type to cause disease was subsequently compared in mice using mixed infections using the combinations D39 vs. D39APPI-1, D39 vs. D391, D39 vs. D391861 and D391861 vs. D39. Data were obtained from the nasal lavage fluid, nasal tissue, lungs and blood at both 24 h and 48 h post-challenge. At 24 h both D39APPI-1 and D391 were less competitive than the wild type in both the lungs and blood (Fig. 6A and 6B). However, D39v was more competitive than the wild-type at the nasopharyngeal surface. In contrast, there was no significant difference between D391861 and the wild type in the blood, lungs or nasal tissue (Fig. 6C). However, D391861 was more competitive than the wild type on the nasopharyngeal mucosa. Interestingly, D391861 was more competitive than D39v in all four niches (Fig. 6D).

Furthermore, at 48 h the wild type was more competitive than D39APPI-1 in the nasal tissue, lungs and blood (Fig. 6E). In contrast, D39v was less competitive than the wild type in the nasal tissue and the lungs, but was equally competitive in the blood at 48 h (Fig. 6F). Similar to 24 h, D391861 was as competitive as the wild-type in the nasal tissue, blood and lungs, but more competitive on the nasopharyngeal surface at 48 h (Fig. 6G). D391861 was also more competitive than D39v in the nasal tissue and blood, but equally competitive in the lungs and at the nasopharyngeal surface (Fig. 6H). In summary, it is clear that PPI-1v plays a role in virulence in a D39 background, and that it is possible that PezAT is important as has been previously reported [42]. However, the importance of other components of PPI-1v in virulence cannot be ruled out. In addition, while the endogenous version of PPI-1v is required for wild-type virulence in D39, it is also true that in a D39 background, PPI-1v from the highly virulent serotype 1 strains was also more competitive than PPI-1v from the non-invasive and intermediately virulent serotype 1 isolates in a number of niches, which correlates with differences in virulence between the isolates in both mice and humans. Therefore, while it is possible that the relative impact of PPI-1v on virulence could be different in a D39 background than in the serotype 1 isolates themselves, the correlation between the virulence of the wild-type strains, the niche-specific changes in expression of PPI-1v genes and in vivo competition between PPI-1v D39 mutants is compelling.

This study aimed to identify serotype-independent virulence determinants within the genomes of a selection of serotype 1 isolates with wide-ranging virulence in both mice and humans. In particular, isolates closely related to the hypervirulent clones responsible for epidemic IPD were found to readily invade and survive in the blood, whereas non-invasive clones isolated from episodes of asymptomatic serotype 1 carriage in remote Indigenous Australian communities were only able to colonize the nasopharynx in mice. A number of regions >1 kb in size were present in the genomes of only the highly virulent isolates, which included a phage genome with putative adherence factors, a number of putative metabolic enzymes, an ABC transporter, an ion transporter as well as a number of potential stress-responsive transcriptional regulators. In particular, the expression of some metabolic enzymes, transporters and adherence factors appeared to exhibit preferential expression in niches associated with IPD. In addition, the various versions of PPI-1v were shown to impact on virulence differently and suggests that this region is at least partly responsible for the greater virulence of highly virulent isolates compared to less virulent isolates. PPI-1v appears to be a highly variable AR within the pneumococcal genome that alters the survival of the bacterium within the host in a content-dependent fashion.

Materials and Methods

Ethics statements

This 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. All animal experiments were approved by the Animal Ethics Committee of the University of Adelaide (Project Number: S-86-2006). Written consent was obtained for studies of human specimens and ethics approval was obtained for further molecular analyses from the Human Research Ethics Committee of the Menzies School of Health Research and Department of Health and Families.

Strains and media

The \textit{S. pneumoniae} serotype 1 clinical strains used in this study that were of Indigenous Australian origin included non-invasive isolates (strains Menzies1-1 [ST304] and Menzies1-2 [ST304]), and invasive isolates (strains Menzies1-3415 [ST227], Menzies1-5482 [ST227] and Menzies1-1861) and were provided by the Royal Darwin Hospital Pathology Services. An invasive serotype 1 isolate (strain WCH4496) was of non-Indigenous origin and was obtained from the Women’s and Children’s Hospital, North Adelaide, Australia. The virulent serotype 2 strain, D39 (NCTC 7466) was also used in this study. The sequence type (ST) of serotype 1 strains was determined by MLST as described in Enright & Spratt (1998) and in accordance with the instructions at http://spneumoniae.mlst.net. Opaque-phase variants of all strains selected on Todd–Hewitt broth supplemented with 1% yeast.
extract (THY)-catalase plates [72] were used in all animal experiments. Before infection, the bacteria were grown in serum broth (SB) (nutrient broth [10 g/l peptone (Oxoid), 10 g/l Lab Lemco powder (Oxoid) and 5 g/l NaCl] and 10% [v/v] donor horse serum) to an optical density at 600 nm (OD600) of 0.16, which approximates $1 \times 10^8$ CFU/ml.

PCR

Chromosomal DNA for PCR was extracted and purified using the Wizard genomic DNA purification kit (Promega Corporation, Madison, WI), with the exception of cell lysis, which was performed by incubating cells at 37°C for 10 min with 0.1% (w/v) sodium deoxycholate. PCR reactions were performed using a G-STORM GS482 thermal cycler (Gene Technologies, UK). Standard reactions were performed using Taq DNA polymerase (Roche Diagnostics, Basel Switzerland) according to the manufacturer’s instructions. The Expand™ Long template or High fidelity PCR systems were used when high fidelity amplification was required. Overlap-extension PCR was carried out essentially as previously described [73,74], using the Expand™ Long Template PCR system. DNA sequencing reactions were carried out using the BigDye® Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, CA, USA).

Animal studies

Inbred 5- to 6-week old female Balb/c mice were used in i.p. challenge experiments and outbred 5- to 6-week old female CD1 (Swiss) mice were used in i.n. challenge experiments. For i.p. challenge experiments groups of 5 mice were used. Mice were challenged i.p with 100 μl of bacterial suspension containing approximately $1 \times 10^4$ CFU in SB. The challenge dose was confirmed retrospectively by serial dilution and plating on blood agar. Mice were monitored for signs of illness over 9 days and were euthanized when moribund. Blood was taken from euthanized mice and plated on blood agar to confirm the presence of *S. pneumoniae* in the blood. For i.n. challenge, groups of 10 mice were anaesthetized by i.p. injection of pentobarbital sodium (Nembutal; Rhone-Merieux) at a dose of 66 mg per g of body weight and challenged with 50 μl of bacterial suspension containing approximately $1 \times 10^7$ CFU in SB. The challenge dose was confirmed retrospectively as described above for i.p. challenge. For quantification of pneumococci and gene expression analysis in infected mouse tissues, groups of 30 mice for strains Menzies1-1861 and WCH4496 and a group of 20 mice for strain Menzies1-1 were challenged via the i.n. route with anesthesia as described above. At 48 h and 96 h, the mice were euthanized by CO2.
For harvest, nasal lavage, blood, lung and nasal samples were collected as previously described [75,76]. A 40-μl aliquot of each sample was serially diluted in phosphate-buffered saline and plated on blood agar to enumerate pneumococci present in each niche and to determine the presence, if any, of contaminating microflora. Blood plates were incubated at 37°C in 95% air, 5% CO2 overnight. Samples were then stored at −80°C until further processing was performed. In addition, a 400-μl aliquot of blood and homogenized nasal and lung tissue was also harvested from each mouse for extraction of prokaryotic RNA.

To assess for pneumococcal presence in the nasal cavity, competition experiments were performed on groups of ten mice per competition group were challenged with a mixed culture of approximately 8 × 10^9 CFU of both strains (as described above) with a mixed culture of approximately 5 × 10^8 CFU per strain. The competitive index (CI) within nasal wash, nasal tissue, blood and lung samples was determined at 24 h and 48 h post-challenge on selective media by calculating the ratio of wild-type to mutant or mutant to mutant as required, relative to the input ratio. As the CI values were log transformed, a value close to 0 is expected if strains compete equally.

**Extraction of total RNA from infected host tissues**

RNA was extracted from host tissues, purified and enriched for bacterial RNA essentially as described previously [75,77]. In this experiment bacterial RNA was pooled from the same 4 mice per niche.

**Linear amplification of total RNA**

Bacterial-RNA samples were amplified using a RNA linear amplification kit SenseAMP (Genisphere), as described previously [75,77].

**Real-time relative qRT-PCR**

The abundance of mRNAs of the genes listed in Table 5 present in amplified RNA recovered from pneumococci harvested from all niches was measured by real-time quantitative RT-PCR (qRT-PCR). Gene specific primers were designed using OligoPerfectTM software (Invitrogen), and primers specific for 16S rRNA were used as internal controls. Data normalization (Table S1). qRT-PCR was performed using a LightCycler® 480 II (Roche) using the Superscript III One-step RT-PCR kit (Invitrogen) according to the manufacturer’s instructions. Expression data were calculated using the 2^(-ΔΔCT) method [78]. Expression data are expressed as a relative increase/decrease between niches.

**Comparative Genomic Hybridization**

Comparative genomic hybridization (CGH) experiments were performed on whole genome S. pneumoniae PCR microarrays as described previously [75,76]. The microarray design is available in BJG@Shace (Accession No. A-BUGS-14; http://bugs.sgd.ac.uk/A-BUGS-14) and ArrayExpress (Accession No. A-BUGS-14). S. pneumoniae DNA for CGH was extracted and purified using a phenol extraction method as described previously [79]. DNA (10.5 μg in 100 μl) was digested with SmalI (New England Biolabs [NEB], MA, USA), and purified using the Qiagen MinElute® PCR Purification kit. Thereafter, 20 μl of purified digest was labeled using the Genisphere Array 900 DNASmDNA labeling kit for Microarrays (Genisphere, PA, USA) for each of the dyes used (Alexa Fluor 555 and Alex Fluor 647). Slides were incubated overnight in a dark humidified chamber at 65°C, washed in a 3-step process (15 min with 2× SSC, 0.03% (v/v) SDS, at 65°C; 15 min with 1× SSC, RT; 15 min with 0.2× SSC, at RT) and dried. Slides were scanned using GenePix® Pro 6.0 software (Axon). CGH was performed in pairs of strains with the same virulence phenotype per slide (1 & 2, 3415 & 5482, 1861 & 4496). Approximately equal fluorescence per spot on the array between channels on one slide indicated presence in both strains of the same phenotype, unequal fluorescence indicated presence only in one strain of the same virulence phenotype, and fluorescence not significantly different from the background indicated absence in both strains of the same virulence phenotype.

**Genomic sequencing**

Sequencing and genome assembly were performed by Gene-works (Thebarton, Adelaide, Australia) using chromosomal DNA prepared as described for CGH, using an Illumina Genome Analyzer II (California, USA) and Lasergene® 8 software (DNASTAR Inc, WI, USA). The sequenced 35-bp reads were assembled against the P1301 (serotype 1, ST303) genome (Genbank accession no. CP000920). The assembly of the strain 1 reads generated a 1,947,650-bp consensus sequence (92.22% of P1031) with an average read depth of 35.72 reads. The assembly of the strain 1861 reads generated a 2,105,218-bp consensus sequence (99.68% of P1031) with an average read depth of 43.66 reads. Unassembled sequences were subsequently assembled de novo into 103 and 62 contigs >300-bp in size from strain 1 and 1861, respectively. While many smaller contigs of boneyard sequence were assembled, these largely represented gaps in sequence assembly due to sequence variation below the 80% sequence identity cutoff. The genome sequence of INV104 (Genbank accession no. FQ312030), ATCC 700669 (Genbank accession no. FM211187), G34 (Genbank accession no. CP0001015) and D39 (Accession no. NC_005353) were used in comparisons. Alignments were performed using the Artemis Comparison Tool (ACT) [80]. The sequences of strains 1 and 1861 were submitted to the sequence read archive at NCBI and have accession numbers SPX030816.2 and SPX030825.2, respectively.

**Construction of PPI-1 variable region mutants**

Mutants were constructed using the primers in Table S1. Mutants requiring the deletion of peczT were performed in two steps. D39ΔPPI-1 and D39ΔPPI-1 were constructed from D39ΔpezT. The construct for D39ΔpezT was generated by overlap extension PCR, performed essentially as described below, from products amplified using primers a - ag and ee - g for the flanking products and J214-J215 for the amplification of ermB from pVex891 [81]. The construct for D39ΔPPI-1 was generated by restriction endonuclease treatment and subsequent ligation using primers t - ed and ee - g for the flanking products amplified from D39 and

---

**Figure 6.** Competition between PPI-1v derivatives and wild type in the nasopharynx, blood and lungs. Competitive index (CI) within nasal lavage, nasal tissue, blood and lung samples of individual mice are indicated at 24 h (A–D) and 48 h (E–H) post-challenge. CIs are expressed as log-transformed ratios of the first strain relative to the second strain for mice where pneumococci CFU were recovered from the relevant niche. Values are pooled from two replicate experiments that used 5 mice per group at each time point. Statistical differences between the log-transformed geometric mean CI and a hypothetical value of 0 (ratio of 1:1) in each niche were analyzed using the one-sample t-test (*, P<0.05; **, P<0.01; ***, P<0.001). doi:10.1371/journal.pone.0019650.g006
RHcatF – RHcatR for amplification of cmr. The construct for D39 was generated by restriction endonuclease treatment and subsequent ligation using primers 5′ – 3′ for the flanking products amplified from strain 1 DNA and cmr. The construct for D39 was generated by restriction endonuclease treatment and subsequent ligation using primers 5′ – 3′ for the flanking products amplified from strain 1 DNA and J293a – J254a for amplification of spe. Generation of competent S. pneumoniae cells and subsequent transformation was performed using the complete transformation medium (CTM) method [82,83].

Statistical analyses

Differences in median survival times and differences in the geometric mean number of pneumococci in each niche between groups were analyzed by the unpaired t-test (two tailed). Differences in the relative expression levels of genes between niche cultures were performed using the unpaired t-test (two tailed). Differences between the competitive index of a test sample and CI = 1 were analyzed using log-transformed values by one-sample t-test. All analyses were performed using GraphPad Prism version 5.01. P< 0.05 was considered significant.

References

12. Hiller NL, Janto B, Hogg JS, Boissy R, Yu S, et al. (2007) Comparative genomic analysis of the pneumococcal MLST database which is located at Imperial College London and is funded by the Wellcome Trust.

Supporting Information

Table S1 List of oligonucleotides used in this study. (DOC)

Table S2 List of genes present in ARs associated with hypervirulence. (XLS)

Acknowledgments

We would like to thank Andrew Lawrence from the Women’s and Children’s Hospital, North Adelaide, South Australia, for the provision of strain 4496 in this study. We acknowledge BiGGSi (the Bacterial Microarray Group at St. George’s Hospital, University of London) for supply of the microarray slides and advice. We also acknowledge the use of the pneumococcal MLST database which is located at Imperial College London and is funded by the Wellcome Trust.

Author Contributions

Conceived and designed the experiments: RMH UHS ADO JCP. Performed the experiments: RMH UHS ADO. Analyzed the data: RMH UHS ADO JCP. Contributed reagents/materials/analysis tools: HCS-V AJL. Wrote the paper: RMH UHS ADO JCP.
44. Khoo SK, Loll B, Chan WT, Shoeman RL, Ngoo L, et al. (2007) Molecular and
24. Smith BT, Walker GC (1998) Mutagenesis and more: