OXFORD UNIVERSITY PRESS | Genome Biology and Evolution

Parallel evolution drives diversification in Streptococcus pneumoniae biofilms

Journal:	*Genome Biology and Evolution*
Manuscript ID	GBE-151009.R1
Manuscript Type:	Research Article
Date Submitted by the Author:	19-Mar-2016
Complete List of Authors:	Churton, Nicholas; University of Southampton, Centre for Biological Sciences, Faculty of Natural and Environmental Sciences; University of Southampton, Institute for Life Sciences; University of Southampton, Academic Unit of Clinical and Experimental Sciences Misra, Raju; Public Health England, Genomics Research Unit, Microbiology Services Howlin, Robert; University of Southampton, Centre for Biological Sciences, Faculty of Natural and Environmental Sciences; NIHR Southampton Respiratory Biomedical Research Unit; University of Southampton, Institute for Life Sciences Allan, Raymond; University of Southampton, Academic Unit of Clinical and Experimental Sciences, Faculty of Medicine; NIHR Southampton Respiratory Biomedical Research Unit; University of Southampton, Institute for Life Sciences, Faculty of Natural and Environmental Sciences; University Hospital Southampton NHS Foundation Trust, Southampton NIHR Wellcome Trust Clinical Research Facility Jeffries, Johanna; University of Southampton, Academic Unit of Clinical and Experimental Sciences; Public Health England; NIHR Southampton Respiratory Biomedical Research Unit; University of Southampton, Institute for Life Sciences Faust, Saul; University of Southampton, Academic Unit of Clinical and Experimental Sciences, Faculty of Medicine; NIHR Southampton Respiratory Biomedical Research Unit; University Hospital Southampton NHS Foundation Trust, Southampton NIHR Wellcome Trust Clinical Research Facility Gharbia, Saheer; Public Health England, Genomics Research Unit, Microbiology Services Edwards, Richard; University of New South Wales, School of Biotechnology and Biomolecular Sciences, Faculty of Natural and Environmental Sciences; University of Southampton, Centre for Biological Sciences Clarke, Stuart; University of Southampton, Academic Unit of Clinical and Experimental Sciences, Faculty of Medicine; Public Health England; NIHR Southampton Respiratory Biomedical Research Unit; University of Southampton, Institute for Life Sciences Webb, Jeremy; University of Southampton, Centre for Biological Sciences, Faculty of Natural and Environmental Sciences; NIHR Southampton Respiratory Biomedical Research Unit; University of Southampton, Institute for Life Sciences

Keywords:	Biofilm, Streptococcus pneumoniae, genomic diversification, parallel evolution

Parallel evolution in *Streptococcus pneumoniae* biofilms

Nicholas W. V. Churton[1,3,6], Raju V. Misra[2], Robert P. Howlin[1,5,6], Raymond N. Allan[3,5,6,8,] Johanna Jefferies[3,4,5,6], Saul N. Faust[3,5,8], Saheer E. Gharbia[2], Richard J. Edwards[1,6,7], Stuart C. Clarke[ϕ3,4,5,6] and Jeremy S. Webb*[ϕ1,5,6]

(1) Centre for Biological Sciences, Faculty of Natural and Environmental Sciences, University of Southampton, UK;

(2) Genomics Research Unit, Microbiology Services, Public Health England, Colindale, UK;

(3) Academic Unit of Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, UK;

(4) Public Health England, Southampton, UK;

(5) NIHR Southampton Respiratory Biomedical Research Unit, Southampton, UK

(6) Institute for Life Sciences, Faculty of Natural and Environmental Sciences, University of Southampton, UK

(7) School of Biotechnology and Biomolecular Sciences, University of New South Wales, Australia

(8) Southampton NIHR Wellcome Trust Clinical Research Facility, University Hospital Southampton NHS Foundation Trust, Southampton, United Kingdom,

* Corresponding author: J.S.Webb@soton.ac.uk

ϕ These authors contributed equally to the work

Key Words: Biofilm, *Streptococcus pneumoniae*, genomic diversification, parallel evolution.

Abstract

Streptococcus pneumoniae is a commensal human pathogen and the causative agent of various invasive and non-invasive diseases. Carriage of the pneumococcus in the nasopharynx is thought to be mediated by biofilm formation, an environment where isogenic populations frequently give rise to morphological colony variants, including small colony variant (SCV) phenotypes. We employed metabolic characterization and whole genome sequencing of biofilm-derived *S. pneumoniae* serotype 22F pneumococcal SCVs to investigate diversification during biofilm formation. Phenotypic profiling revealed that SCVs exhibit reduced growth rates, reduced capsule expression, altered metabolic profiles, and increased biofilm formation compared to the ancestral strain. Whole genome sequencing of 12 SCVs from independent biofilm experiments revealed that all SCVs studied had mutations within the DNA-directed RNA polymerase delta subunit (RpoE). Mutations included four large-scale deletions ranging from 51-264 bp, one insertion resulting in a coding frameshift, and seven nonsense single nucleotide substitutions that result in a truncated gene product. This work links mutations in the *rpo*E gene to SCV formation and enhanced biofilm development in *S. pneumoniae*, and therefore may have important implications for colonization, carriage and persistence of the organism. Furthermore, recurrent mutation of the pneumococcal *rpo*E gene presents an unprecedented level of parallel evolution in pneumococcal biofilm development.

Introduction

Streptococcus pneumoniae is an encapsulated bacterium that forms part of the human respiratory biota (Tocheva, et al. 2011). Carriage is asymptomatic but known to facilitate transmission of the pathogen and precede pneumococcal infection (Simell, et al. 2012). The resultant diseases can range from acute otitis media and pneumonia to invasive disease including septicemia and meningitis. Whilst much disease is preventable by vaccination (Gladstone, et al. 2011) it is still a major cause of morbidity and mortality worldwide (O'Brien, et al. 2009) requiring continuous surveillance.

Carriage of the pneumococcus is thought to be facilitated by biofilm formation (Marks, et al. 2012); biofilms are complex aggregations of bacteria which adhere to inert or living surfaces and are encased in an extracellular polymeric substance (Costerton, et al. 1995). Biofilms are a major contributor to chronic disease (Hall-Stoodley and Stoodley 2009) and it is estimated that around 65% of infections are a result of biofilm formation (Potera 1999). Furthermore, pneumococcal biofilm formation has been directly observed in chronic pneumococcal infections (Hall-Stoodley, et al. 2006; Nistico, et al. 2011).

Microorganisms have been used extensively to study rapid evolutionary changes in real time (Buckling, et al. 2009); large populations, short generation times and relatively simple genomes, coupled with the advent of next generation whole genome sequencing, allow us to address fundamentally important questions about adaption, mutation and morphological change within bacterial populations. Growth of respiratory pathogens within biofilms can result in rapid genetic diversification; colony morphology variants can emerge that exhibit increased adhesion, dispersal and/or recalcitrance to oxidative stress and antibiotics (Allegrucci and Sauer 2008; Boles, et al. 2004; Mitchell, et al. 2010; Webb, et al. 2004). Studies of pneumococcal biofilms have shown colony variants to be both phenotypically and genetically distinct from ancestral strains (Allegrucci and

Sauer 2007, 2008; Domenech, et al. 2009; Waite, et al. 2001). *S. pneumoniae* is a highly recombinant organism and prone to mutations that facilitate long term persistence in the host (Croucher, et al. 2011). This phenomenon of biofilm-mediated genetic diversification is thought to be a key factor in pneumococcal persistence within the host and is thought to act as a survival strategy, whereby species with the highest genetic diversity have the greatest ability to withstand perturbations within the environment (Boles, et al. 2004). The mechanisms by which bacteria attain such diversity remains unclear and are fundamental to understanding how biofilms persist despite continual antimicrobial therapy and is potentially relevant to future vaccine design (Jefferies, et al. 2011).

Cells growing within a biofilm are subjected to steep oxygen and pH gradients (Debeer, et al. 1994; Sternberg, et al. 1999), resulting in chemical gradients and micro-environmental niches within the biofilms (Debeer, et al. 1994). Adaptation to changes in environmental conditions is essential for survival within the biofilm (Ehrlich, et al. 2005). Recently, spontaneous mucoid variants have been shown to repeatedly arise in *Pseudomonas fluorescens* colonies (Kim, et al. 2014). These variants were shown to spatially position themselves on the surface of the colony to optimize access to oxygen and nutrients. The authors proposed that the mucoid variants underwent strong selection in mixed colony experiments where rapid increase in the mucoid phenotype resulted in dominance over the WT strain (Kim, et al. 2014). Furthermore whole genome sequencing of over 500 mucoid variants revealed a striking example of parallel evolution at the *rsm*E locus (Kim, et al. 2014). These data suggest that parallel evolution can underlie diversification in surface-attached communities of bacteria within biofilms.

In this work we use a clinical isolate of serotype 22F, which was chosen as a model serotype because its prevalence in pneumococcal carriage and disease has increased in both the UK and the US since the implementation of pneumococcal conjugate vaccines (Jacobs, et al. 2008; Tocheva, et al. 2011). We perform both conventional phenotyping and next generation sequencing to characterize the genetic diversity that can arise from biofilm-derived pneumococci. Mutations that

arise in serotype 22F pneumococcal variants were identified and related to phenotypic characteristics. We assess genotype-phenotype relationships and provide insight into genomic factors that are strongly favored under natural selection in pneumococcal biofilms.

Results

Three distinct *S. pneumoniae* colony variants were observed after three days of biofilm growth in micro-titre plates; small colony variants (SCVs) (<0.5 mm), medium colony variants (MCVs) (0.5 to 1 mm) and typical colony variants (TCVs) (>1 mm) (figure 1). Consistent with the ancestral strain, all variants stained Gram-positive, exhibited alpha hemolysis and remained optochin-sensitive. Colony-forming unit (CFU) enumeration of the biofilm biomass showed that the majority of colonies formed retained the wild type (WT) (TCV) morphology whereas SCVs and MCVs represented approximately 1% and 4% of the population respectively by day 9 (figure 1). Quantification of the colony diameter confirmed that SCVs and MCVs were significantly distinct from each other and the WT ancestral strain (WT vs. MCV: $T = 13.51$, $df = 16$, $p < 0.001$; WT vs. SCV: $T = 21.38$, $df = 16$, $p < 0.001$; MCV vs. SCV: $T = 9.73$ $df = 16$, $p < 0.001$). In contrast, colonies with the TCV phenotype did not differ significantly from the WT ancestral strain ($T = 0.28$, $df = 16$, $p > 0.05$) (figure 2). All variant morphologies had smooth margins. In order to determine whether this variation in morphology was a direct result of biofilm growth and not simply prolonged growth, planktonic WT cultures were grown for nine days, sub-culturing daily into fresh BHI to maintain cell viability. At days 1, 3, 6 and 9, cells were serially diluted and plated onto blood agar to assess morphology. Over the nine day time course the WT morphology remained consistent and the presence of small and medium sized colonies was not observed (data not shown).

Biofilm-derived colony variants were assessed for their ability to form biofilms compared to the ancestral strain. Confocal laser scanning microscopy (CLSM) revealed that the WT ancestor produced relatively flat biofilms with an average thickness of approximately 6 µm and small microcolony towers with maximum thickness of 12-15 µm (figure 3A). In contrast, SCVs produced structured biofilms with an average thickness of approximately 13 µm and well-defined microcolony towers with a maximum thickness of 25-35 µm (figure 3B). The 2-sample t-test revealed that SCV biofilms had significantly greater biomass ($\mu m^3/\mu m^2$) ($T = 3.26$, $df = 4$, $p < 0.05$), surface area (μm^2) (T

= 3.56, df = 4, $p < 0.05$), average thickness (µm) (T = 2.81, df = 4, $p < 0.05$) and maximum thickness (µm) (T = 5.47, df = 4, $p < 0.01$) compared to the ancestral strain (Figure 4). MCVs produced slightly larger biofilms than WT with an average thickness of approximately 10 µm and maximum thickness of 15-20 µm (figure 3C). TCVs produced biofilm comparable with WT; with an average thickness of 8 µm and small microcolony towers with a maximum thickness of 13-14 µm (figure 3D). Biofilms produced by the MCVs and TCVs did not differ significantly from WT. The percentage of dead cells within the biofilms was not significantly different between variants.

Phenotypic profiling using the API Rapid ID 32 Strep assay confirmed that all SCVs were *S. pneumoniae* and not contaminants. Profiling also revealed that eight of the 12 SCVs were unable to metabolize one or all of the following carbon substrates: D-trehalose, pullulan, maltose and D-saccharose. Furthermore, SCVs exhibited a reduced growth rate compared to the WT (T = 2.84, df = 37, $p < 0.01$) (figure 5), and a decrease in capsule staining compared to the WT (T = 7.79, df = 40, $p < 0.001$) (Figure 6) compared to the WT.

To identify mutations responsible for the phenotypic diversity observed in the SCVs, a total of 12 SCVs were harvested from either a 3-, 6- or 9-day biofilm from two independent biofilm experiments (five SCVs from experiment 1 and 7 SCVs from experiment 2) (table 1). The 12 SCVs were sequenced using next generation sequencing and compared to the 22F ancestral strain. Table 1 lists the confirmed mutations identified in this work. Notably all sequenced SCVs were shown to contain independent mutations within the DNA-directed RNA polymerase delta subunit (*rpo*E). With the exception of SCV5D3E1 and SCV7D9E1, these mutations were at different positions within the gene; SCV5D3E1 and SCV7D9E1 were isolated from separate biofilms, thus also occurred independently. Of the 12 SCVs sequenced, seven SCVs contained single nucleotide substitutions resulting in an early stop codon, one SCV had an insertion mutation resulting in a coding frameshift, and four SCVs contained large-scale deletions ranging from 51 to 264 bp (Table 1). PCR of the *rpo*E gene was employed to confirm the deletions seen in the whole genome analysis. Sequence analysis

identified short homologous regions flanking the deletions in three SCVs (SCV9D9E1, SCV1D3E2 and SCV2D6E2), suggesting underlying recombination events; there was no evidence to suggest that the deletion in SCV7D9E2 was a result of a recombination event. No change in the multi locus sequence typing profile of the SCVs was observed.

Discussion

Increasing evidence suggests that carriage of the pneumococcus in the nasopharynx of humans is mediated by biofilm formation (Blanchette-Cain, et al. 2013; Marks, et al. 2012). This work used conventional phenotyping and whole genome sequencing to characterize the genetic diversity that arises among biofilm-derived pneumococci. Three distinct colony morphologies were observed after three days of growth, including a small colony phenotype. This variation in colony morphology is consistent with previous studies involving other pneumococcal serotypes (Allegrucci and Sauer 2007, 2008) suggesting that this phenomenon is not an isolated occurrence. Whole genome sequencing revealed that all SCVs contained mutations within the DNA-directed RNA polymerase subunit gene (RpoE); including one insertion, seven substitutions and four large-scale deletions. In all cases these mutations would result in truncation of the *rpo*E gene product affecting the low complexity regions of the protein which have been linked to the specificity of DNA binding. Despite few SCV mutations overall, these parallel evolutionary events suggest that a strong selection pressure for *rpo*E-targeted mutations occurs within *S. pneumoniae* biofilms and may be important for biofilm development.

Of the four deletion events observed in this work, three are likely to be due to a single intra-molecular recombination event. In all four deletion events the first 8 bp of the deleted sequence commenced with the sequence pattern GA(C/A)GA(A/C)GA. This sequence pattern may be analogous to the *Chi* site seen in *Escherichia coli*, (Prudhomme, et al. 2002) and represent a recognition site for the pneumococcal recombinase machinery which in turn results in the deletions seen, however this hypothesis remains to be tested. Eight of the 12 sequenced SCVs contained a mutation only in the *rpo*E gene with no other additional changes to the genome. With the exception of SCV5D3E1 and SCV7D9E1, all mutations occurred at different positions in the gene. This observation supports the hypothesis that mutations in the *rpo*E gene are directly linked with the SCV phenotype. The consistent high frequency of mutations in a single gene, in multiple replicates of independent experiments, is clear evidence of biofilm-mediated parallel evolutionary events in real-

time. Parallel evolution of specific genes has been shown to be clinically relevant in determining the pathogenesis of *Burkholderia dolosa* (Lieberman, et al. 2011) and *P. fluorescens* (Kim, et al. 2014). Such a phenomenon has also been reported previously in *Pseudomonas aeruginosa* (McElroy, et al. 2014), however this study presents some of the best evidence of parallel evolutionary events in *S. pneumoniae*.

To date RpoE has not been studied in *S. pneumoniae.* Using *Bacillus subtilis* as a model organism, RpoE has been attributed to increasing transcriptional specificity by increasing binding of RNAP to promoter sequences (Achberger, et al. 1982; Achberger and Whiteley 1981). Additionally, RpoE has been attributed to increased efficiency of RNA synthesis and decreased affinity for nucleic acids due to enhanced recycling of RNA (Juang and Helmann 1994; Lopez de Saro, et al. 1999). Understanding of the RpoE function is based on gene deletion studies; *rpo*E deletion mutants have been shown to be viable suggesting that RpoE is non-essential (Lopez de Saro, et al. 1999), however, phenotypically *rpo*E mutant strains have be shown to display increased lag phase, altered cell morphology (Lopez de Saro, et al. 1999) and altered biofilm architecture (Xue, et al. 2010). In all studies to date disruption of the *rpo*E gene has been mediated by deletion or alteration of the sequence (Lopez de Saro, et al. 1995; Weiss, et al. 2014; Xue, et al. 2010). Our experiments provide evidence that mutations affecting the *rpo*E gene occur spontaneously during biofilm development. Their recurrence suggests an advantage in terms of growth and/or survival within biofilms.

S. pneumoniae is a nutritionally fastidious facultative anaerobe but the availability of a glucose source within the human nasopharynx is thought to be in low concentrations (Philips, et al. 2003). As biofilm formation is thought to represent a survival strategy in a nutritionally limited environment (Domenech, et al. 2012), possibly, the reduction in carbon substrate utilization on the API strip may reflect selection for strains with reduced metabolic capacity and growth rate because they are better adapted to survive under biofilm conditions. Alternatively, there may be selection

against the maintenance of energetically costly pathways not required under biofilm conditions. Further experimentation would be required to determine if either of these hypotheses are correct.

We hypothesize that the mutations seen here can benefit the survival of *S. pneumoniae* in biofilms by altering gene expression in favor of colonization rather than virulence. In Gram-positive bacteria RpoE has been shown to be important for rapid changes in gene expression in order to adapt to changing environmental conditions (Rabatinova, et al. 2013; Xue, et al. 2012; Xue, et al. 2011; Xue, et al. 2010). The reduction in growth rate seen in SCVs in this study is consistent with RpoE studies in *B. subtilis* (Lopez de Saro, et al. 1999). The slower growth that we observed may be beneficial to a sub-population of SCVs in the biofilm because it may allow bacteria to avoid direct competition between neighboring cells under conditions of limited nutrient availability. A number of studies have shown biofilm variants to have reduced virulence (Sanchez, et al. 2011), which may similarly reflect a change in gene expression that promotes survival within biofilms. Interestingly, RpoE has been shown to be up-regulated in response to quorum sensing molecule autoinducer-2 in *Streptococcus mutans* (Sztajer, et al. 2008).

Of note, additional point mutations were also observed in choline binding protein J, the streptococcal histidine triad protein and the iron compound ABC uptake transporter permease protein (PiuB). This observation is relevant as mutations within cell surface proteins may inform future protein-based pneumococcal vaccine studies. Protein-based pneumococcal vaccine design should be cautious not to target cell surface proteins which have a tendency to mutate, as this may result in vaccine-escape. It is interesting to note that we observed no mutations within the capsule genes of the biofilm-derived variants. This differs from previous studies that implicated a 7 kb deletion in the cps3DSU operon in the formation of SCV variants in serotype 3 (Allegrucci and Sauer 2007). Whether biofilm-mediated mutations are a result of natural selection acting on random mutational events or a process of targeted or adaptive mutation remains to be determined (Banas, et al. 2007). In this work, the lack of SCV phenotypes in the planktonic population and the variety of

*rpo*E mutations seen would suggest a strong selection pressure for *rpo*E-targeted mutations occurs within *S. pneumoniae* biofilms and may be important for biofilm development.

This work has shown that diversification under biofilm conditions resulted in the emergence of SCVs with parallel mutations within the *rpo*E gene. Further work is investigating the transcriptional and translational effect of *rpo*E mutation on biofilm development and its clinical significance, as well as comparisons of genomic diversification between disease and carriage isolates of *S. pneumoniae* 22F. This study presents an unprecedented level of biofilm-mediated parallel evolution in *S. pneumoniae* and suggests a strong positive selection for mutations in *rpo*E, which may identify *rpo*E as an important gene for biofilm formation, carriage and pathogenicity of the pneumococcus.

Materials and Methods

Bacterial strains

The clinical isolate (ID: sp_3016) of *S. pneumoniae* 22F (ST433) was obtained from the third year (October 2009 - March 2010) of an on-going nasopharyngeal carriage study in children aged 4 years and under at University Hospital Southampton, UK. Isolates were stored at -80 ^{o}C in glycerol stock consisting of 50% brain heart infusion (BHI) broth (Oxoid) and 50% glycerol (Sigma). Capsular serotyping of the reference isolate was performed using genomic DNA and multiplex PCR (Pai, et al. 2006) and multi locus sequence typing (MLST) was performed using Qiagen Genomic Services and ST's assigned using the MLST website (www.mlst.net) prior to biofilm experimentation.

Growth curves

Isolates were grown on Columbia blood agar (CBA) plates (Oxoid) overnight, colonies were used to inoculate 10 ml of BHI broth, and subsequently incubated at 37 ^{o}C with 5% CO_2 for 10 hours. The optical density (OD_{600}) was measured every hour and 100 µl of culture taken every two hours to perform viable counts. Growth curves for each strain were performed in triplicate on separate days.

Biofilm culture and colony variation

1 x 10^8 CFU/ml was inoculated into MatTek culture plates (MatTek Corporation Ashland, MA) or 6-well mico-titre plates (Nunc) as described previously (Hall-Stoodley, et al. 2008). Triplicate biofilms were grown under static conditions at 37 ^{o}C with 5% CO_2 for 1, 3, 6 and 9 days in two independent experiments. Half of the culture media was removed daily and replaced with equal quantities of fresh 1:5 BHI pre-warmed to 37 ^{o}C. At each time point biofilms were washed twice with fresh pre-warmed 1:5 BHI and the biomass was harvested using a cell scraper, vortexed, diluted to 10^{-6} and 100 µl was plated onto CBA (Oxoid) in triplicate to assess changes in colony morphology compared to the inoculum. The growth rate was assessed for SCVs using the equation µ = ((log10 N - log10 N0)

2.303) / (t - t0) and CFU data from the exponential growth phase (between 4 and 8 hours of growth) from triplicate growth experiments.

Assessment of colony morphology

Colony morphology was assessed based on size and regularity of the colony surface. Variants were defined as follows; small colony variants (SCVs) (<0.5 mm), medium colony variants (MCVs) (0.5-1 mm) and typical colony variants (TCVs) (>1 mm). Colonies were visualized using a Leica MZ 16F stereomicroscope and images were taken using a Leica digital camera with microscope and a 5 mm graticule scale. Colony diameters were quantified using the ImageJ analysis software (http://rsbweb.nih.gov/ij/). A total of nine colony variants from triplicate biofilms were measured. Per experiment, up to 12 phenotypic variants of each morphology type were selected, per time point, from the triplicate biofilms. Variants were sub-cultured and stored at -80 ^{o}C in glycerol stock for future phenotypic and genetic analysis. Additional phenotypic characterization and whole genome sequencing was performed on 12 SCVs harvested from two independent experiments. Five SCVs were characterized from experiment 1 (3 × 3-day-old biofilms and 2 × 9-day-old biofilms) and seven SCVs from experiment 2 (1 × 3-day-old biofilms, 3 × 6-day-old biofilms and 3 × 9-day-old biofilms).

Visualization of the biofilm

Biofilms were imaged using confocal laser scanning microscopy (CLSM) and *Bac*Light live/dead stain (Life Sciences). Triplicate biofilms of each variant type were cultured in MatTek culture plates (MatTek Corp. Ashland, US) under static biofilm conditions for 3 days, washed twice with fresh warmed 1:5 strength BHI, and then stained according to the manufacturer's instructions. All images were taken using a 63 x objective on a Leica TCS SP5 confocal laser scanning microscope on a Leica DMI6000 inverted microscope frame. Five fields of view were taken per triplicate biofilm. Z-Scans were performed every 0.3 µm on each field of view. Biofilm formation was quantified using

COMSTAT 1 and the program Matlab (R2012a) (Heydorn, et al. 2000) to determine the average thickness and maximum thickness, total biomass and surface area of the WT and SCV biofilms.

API Rapid ID 32 Strep assay

API Rapid ID 32 Strep assay (bioMérieux) was performed according to the manufacturer's instructions.

Capsule quantification

Quantification of the capsule was adapted and performed according to Hammerschidt *et al.* (2005) and Rukke *et al.* (2012). SCVs were grown on CBA plates (Oxoid) overnight at 37 ^{o}C and 5% CO_2. The growth was transferred to 10 ml of BHI and cultured to an OD equating to approximately 10^8 CFU/ml (OD_{600} ~ 0.3-4). 5 ml cultures were centrifuged (2500 *xg* at 4 ^{o}C for 10 minutes) to pellet the cells. The supernatant was discarded, the pellet washed twice in 0.5 ml PBS, then re-suspended into 5 ml of sterile distilled water. 250 µl of re-suspended cells was transferred to a sterile microcentrifuge tube and 1 ml of Stain-all solution [20 mg Stains-all (Sigma-Aldrich), 60 µl glacial acetic acid in 100 ml of 50% formamide] added. Absorbance was measured at 640 nm and subtracted from the negative control (250 µl of sterile distilled water stained with 1 ml of Stains-all). Capsule quantification was measured in triplicate, with each measurement performed on separate days.

Genomic DNA extraction

Genomic DNA for whole genome sequencing was extracted from 10 ml cultures of BHI containing approximately 10^9 CFU/ml using Qiagen Genomic tip 100/G in accordance with the manufacturer's instruction for Gram-positive bacteria. Genomic DNA for PCR was extracted using the QIAamp DNA mini kit (QIAGEN) according to the manufacturer's instructions with minor modifications. Frozen stocks of each isolate were streaked onto CBA (Oxoid) and incubated at 37 ^{o}C in 5 % CO_2. Colony growth was transferred using a sterile swab to 200 µl of lysis buffer (10 mM Tris pH8, 100 mM ethylenediaminetetraacetic acid pH8, 0.5% w/v sodium dodecyl sulphate) and incubated at 37 ^{o}C for

1 hour. 20 µl proteinase K was added and samples were incubated at 56 °C for 1 hour. The following steps were performed as described in the Qiagen QIAamp mini kit protocol.

Whole genome sequencing

Biofilm-derived variants and the ancestral WT 22F strain underwent shot gun paired end whole genome sequencing using the Roche GS Junior™ and/or the Illumina MiSeq™. The ancestral 22F strain was sequenced both on both the Roche GS junior™ 454 sequencer and on the Illumina MiSeq™ using the 150 base, pair-end protocol. The 12 SCV samples were sequenced on the Illumina MiSeq™ (2 x 150 or 2 x 250 bp). For Roche GS junior™ 454 paired-end sequencing, genomic DNA was sheared and GS Titanium Library Paired End Adaptors were added to the fragments, emPCR was performed using the Roche GS Junior™ Titanium emPCR Kit (Lib L). Samples were loaded onto the Roche GS Junior™ PicoTiterPlate according to the manufacturer's instructions and the Roche GS Junior™ Titanium Sequencing Kit. For Illumina MiSeq™ sequencing, all samples were sheared and libraries were prepared using the Nextera library prep kit, Illumina MiSeq™ reagent cartridge and sequencing reagents were used according to the manufacturer's instructions. Paired-end sequencing of the WT strain acted as the template for comparing the variants. Illumina sequence data was assembled *de novo* using Velvet (Zerbino 2010) and the Velvet optimizer script generating assemblies with an average coverage of 62×. Assemblies were assessed for quality using the assemblathon script and satisfactory n50 scores were generated (mean = 40945 bp) (Earl, et al. 2011). For the ancestral 22F strain, Roche GS junior™ 454 sequence data was assembled *de novo* using Newbler GSassembler with Illumina sequence data to generate a more comprehensive reference sequence. The resultant assembly of the ancestral strain had an 86x depth of coverage. The *de novo* assembled scaffolds of the ancestral strain were aligned against strain D39 reference in Mauve (Darling, et al. 2010), in order to arrange the scaffold. All gaps were removed to create a pseudoreference from which all annotations were based. Annotation of the ancestral strain *de novo* assembly was achieved using RAST, the online annotation service (Overbeek, et al. 2014). Sequence

data are available from the NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra/) under the accession number: SRP071227.

Multi-locus sequence typing (MLST) of biofilm-derived variants

All sequences were assembled *de novo* as described above; the assembled contigs were uploaded to the Centre for Genomic Epidemiology online service (Larsen, et al. 2012) to assess for changes in MLST from the ancestral strain ST433.

Mutation analysis

Point mutations were identified by mapping variant genome sequences against the annotated ancestral strain reference. Illumina fastq reads were mapped against the ancestral 22F strain sequence using STAMPY (Lunter and Goodson 2011). The output sequence alignment mapping (SAM) files were converted to binary alignment mapping (BAM) files using SAMtools and variant discovery was performed using mpileup, bcftools and gatk annotator (Li, et al. 2009; McKenna, et al. 2010) to generate variant call files (.vcf), which contained the list of nucleotide variants, their position and quality metrics. Variants were filtered based on quality score (q-value) and those with q-values <20 were rejected. Alternatively variant call files (.vcf) were generated using SNPtree1.1 (Leekitcharoenphon, et al. 2012). Each mutation was manually curated by mapping the Illumina fastq files against the WT reference to generate bam files (.bam) and indexed bam files (.bai) using SAMtools. Bam files were subsequently visualized using Integrative Genomics Viewer (Robinson, et al. 2011; Thorvaldsdottir, et al. 2013) to determine coverage and percentage confidence for each variant. Furthermore the percentage confidence of each mutation was calculated as the coverage of the identified variant base call divided by the total base call coverage at the variant position. Mutations underwent rigorous selection criteria to avoid mis-calling. Only variant positions that were present in both in the mapping data and *de novo* assemblies (>20x coverage and a percentage confidence >90%) were classed as confidently identified mutations. Functional effects of the

mutations were postulated using the ExPASy research portal (Artimo, et al. 2012) and the EMBL SMART database (Letunic, et al. 2012; Schultz, et al. 1998).

Polymerase chain reaction of *rpo*E

To confirm the validity of the *rpo*E deletions in the SCVs, DNA from each SCV underwent PCR for the *rpo*E gene. The *rpo*E gene is 588 bp in length and primers were designed to amplify an amplicon of 609 bp, encompassing the full *rpo*E gene. Forward primer 5'-GAGGAGAAACGCTTTGGAATTAGAAG-3' and reverse primer 5'-GCTAACTCTTATTCCTCGCTGGTTTC-3' were designed. PCR program consisted of three stages; stage 1: 94 ^{o}C for 5 minutes, stage 2: 30 cycles of 94 ^{o}C for 30 seconds 50 ^{o}C for 30 seconds 72 ^{o}C for 30 seconds, stage 3: 72 ^{o}C for 10 minutes and stage 4: Pause at 4 ^{o}C. PCR products were run out on a 1% agarose gel containing GelRed^{TM} for 45 minutes at 90 V and visualized under ultraviolet light. Bioline HyperLadder^{TM} II was used to assess the size of the bands.

Statistical analysis

All statistical analysis was performed on Minitab^{TM} 16.2 Statistical Software (www. minitab.com). Normality of the data was determined using the Anderson-darling test and equality of variance was determined using the F-test. Normality and equal variance was achieved after square root transformation of the growth rate and capsule staining data. Prior to analysis data from all 12 sequenced SCVs were pooled and the 2-sample t-test was used to determine a significant difference between the SCV phenotype and the WT 22F strain.

Ethics statement

Ethical approval for the collection of nasopharyngeal swabs from children aged four years and under with written informed consent from parents/guardians was obtained from Southampton and South West Hampshire Research Ethics Committee 'B' [REC 06/Q1704/105]. Ethical approval was not required for further use of the bacterial isolates from the swabs.

Acknowledgements

This work is funded by the BBSRC and Pfizer UK with additional funding from the EPSRC. Thanks are due to Dr. Nicola Thorne, Dr. Lisa Cross and Dr. Chloe Bishop for sequencing services at the Department of Bioanalysis and Horizon Technologies Public Health England, Colindale, UK. Thanks are also due to Alice Still and Jenna Alnajar for additional sequencing services at the Academic Unit of Clinical and Experimental Sciences, University of Southampton, UK. Thanks are also due to Dr. David Johnston and the Southampton Biomedical Imaging Unit for support imaging the biofilms, the staff and laboratories of the Southampton National Institute of Health Research (NIHR) Wellcome Trust Clinical Research Facility and NIHR Respiratory Biomedical Research Unit. Parts of this work were used in a poster presentation at the 15th Conference in Genomics and Proteomics of Human Pathogens at Public Health England, Colindale, UK.

References

Achberger EC, Hilton MD, Whiteley HR 1982. The effect of the delta subunit on the interaction of Bacillus subtilis RNA polymerase with bases in a SP82 early gene promoter. Nucleic Acids Res 10: 2893-2910.

Achberger EC, Whiteley HR 1981. The Role of the Delta-Peptide of the Bacillus-Subtilis Rna-Polymerase in Promoter Selection. Journal of Biological Chemistry 256: 7424-7432.

Allegrucci M, Sauer K 2007. Characterization of colony morphology variants isolated from Streptococcus pneumoniae biofilms. Journal of Bacteriology 189: 2030-2038. doi: 10.1128/jb.01369-06

Allegrucci M, Sauer K 2008. Formation of Streptococcus pneumoniae non-phase-variable colony variants is due to increased mutation frequency present under biofilm growth conditions. Journal of Bacteriology 190: 6330-6339. doi: 10.1128/jb.00707-08

Artimo P, et al. 2012. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Research 40: W597-W603. doi: Doi 10.1093/Nar/Gks400

Banas JA, et al. 2007. Evidence that accumulation of mutants in a biofilm reflects natural selection rather than stress-induced adaptive mutation. Applied and Environmental Microbiology 73: 357-361. doi: 10.1128/aem.02014-06

Blanchette-Cain K, et al. 2013. Streptococcus pneumoniae biofilm formation is strain dependent, multifactorial, and associated with reduced invasiveness and immunoreactivity during colonization. MBio 4: e00745-00713. doi: 10.1128/mBio.00745-13

Boles BR, Thoendel M, Singh PK 2004. Self-generated diversity produces "insurance effects" in biofilm communities. Proceedings of the National Academy of Sciences of the United States of America 101: 16630-16635. doi: 10.1073/pnas.0407460101

Buckling A, Craig Maclean R, Brockhurst MA, Colegrave N 2009. The Beagle in a bottle. Nature 457: 824-829. doi: 10.1038/nature07892

Costerton JW, Lewandowski Z, Caldwell DE, Korber DR, Lappinscott HM 1995. Microbial Biofilms. Annual Review of Microbiology 49: 711-745.

Croucher NJ, et al. 2011. Rapid Pneumococcal Evolution in Response to Clinical Interventions. Science 331: 430-434. doi: 10.1126/science.1198545

Darling AE, Mau B, Perna NT 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5: e11147. doi: 10.1371/journal.pone.0011147

Debeer D, Stoodley P, Roe F, Lewandowski Z 1994. EFFECTS OF BIOFILM STRUCTURES ON OXYGEN DISTRIBUTION AND MASS-TRANSPORT. Biotechnology and Bioengineering 43: 1131-1138.

Domenech M, Garcia E, Moscoso M 2012. Biofilm formation in Streptococcus pneumoniae. Microb Biotechnol 5: 455-465. doi: 10.1111/j.1751-7915.2011.00294.x

Domenech M, Garcia E, Moscoso M 2009. Versatility of the capsular genes during biofilm formation by Streptococcus pneumoniae. Environmental Microbiology 11: 2542-2555. doi: 10.1111/j.1462-2920.2009.01979.x

Earl D, et al. 2011. Assemblathon 1: a competitive assessment of de novo short read assembly methods. Genome Res 21: 2224-2241. doi: 10.1101/gr.126599.111

Ehrlich GD, Hu FZ, Shen K, Stoodley P, Post JC 2005. Bacterial plurality as a general mechanism driving persistence in chronic infections. Clinical Orthopaedics and Related Research: 20-24. doi: 10.1097/01.blo.0000175124.52590.c0

Gladstone RA, Jefferies JM, Faust SN, Clarke SC 2011. Continued control of pneumococcal disease in the UK - the impact of vaccination. J Med Microbiol 60: 1-8. doi: jmm.0.020016-0 [pii]

10.1099/jmm.0.020016-0

Hall-Stoodley L, et al. 2006. Direct detection of bacterial biofilms on the middle-ear mucosa of children with chronic otitis media. JAMA 296: 202-211. doi: 10.1001/jama.296.2.202

Hall-Stoodley L, et al. 2008. Characterization of biofilm matrix, degradation by DNase treatment and evidence of capsule downregulation in Streptococcus pneumoniae clinical isolates. Bmc Microbiology 8. doi: 10.1186/1471-2180-8-173

Hall-Stoodley L, Stoodley P 2009. Evolving concepts in biofilm infections. Cellular Microbiology 11: 1034-1043. doi: 10.1111/j.1462-5822.2009.01323.x

Hammerschmidt S, et al. 2005. Illustration of pneumococcal polysaccharide capsule during adherence and invasion of epithelial cells. Infect Immun 73: 4653-4667. doi: 73/8/4653 [pii] 10.1128/IAI.73.8.4653-4667.2005

Heydorn A, et al. 2000. Quantification of biofilm structures by the novel computer program COMSTAT. Microbiology 146 (Pt 10): 2395-2407.

Jacobs MR, Good CE, Bajaksouzian S, Windau AR 2008. Emergence of Streptococcus pneumoniae Serotypes 19A, 6C, and 22F and Serogroup 15 in Cleveland, Ohio, in Relation to Introduction of the Protein-Conjugated Pneumococcal Vaccine. Clinical Infectious Diseases 47: 1388-1395. doi: 10.1086/592972

Jefferies JM, Clarke SC, Webb JS, Kraaijeveld AR 2011. Risk of red queen dynamics in pneumococcal vaccine strategy. Trends Microbiol 19: 377-381. doi: 10.1016/j.tim.2011.06.001

Juang YL, Helmann JD 1994. The delta subunit of Bacillus subtilis RNA polymerase. An allosteric effector of the initiation and core-recycling phases of transcription. J Mol Biol 239: 1-14. doi: 10.1006/jmbi.1994.1346

Kim W, Racimo F, Schluter J, Levy SB, Foster KR 2014. Importance of positioning for microbial evolution. Proc Natl Acad Sci U S A 111: E1639-1647. doi: 10.1073/pnas.1323632111

Larsen MV, et al. 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol 50: 1355-1361. doi: 10.1128/JCM.06094-11

Leekitcharoenphon P, et al. 2012. snpTree--a web-server to identify and construct SNP trees from whole genome sequence data. BMC Genomics 13 Suppl 7: S6. doi: 10.1186/1471-2164-13-S7-S6

Letunic I, Doerks T, Bork P 2012. SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res 40: D302-305. doi: 10.1093/nar/gkr931

Li H, et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079. doi: DOI 10.1093/bioinformatics/btp352

Lieberman TD, et al. 2011. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. Nature Genetics 43: 1275-U1148. doi: Doi 10.1038/Ng.997

Lopez de Saro FJ, Woody AY, Helmann JD 1995. Structural analysis of the Bacillus subtilis delta factor: a protein polyanion which displaces RNA from RNA polymerase. J Mol Biol 252: 189-202.

Lopez de Saro FJ, Yoshikawa N, Helmann JD 1999. Expression, abundance, and RNA polymerase binding properties of the delta factor of Bacillus subtilis. J Biol Chem 274: 15953-15958.

Lunter G, Goodson M 2011. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res 21: 936-939. doi: 10.1101/gr.111120.110

Marks LR, Reddinger RM, Hakansson AP 2012. High levels of genetic recombination during nasopharyngeal carriage and biofilm formation in Streptococcus pneumoniae. MBio 3. doi: 10.1128/mBio.00200-12

McElroy KE, et al. 2014. Strain-specific parallel evolution drives short-term diversification during *Pseudomonas aeruginosa* biofilm formation. Proc Natl Acad Sci U S A 111: E1419-1427. doi: 10.1073/pnas.1314340111

McKenna A, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20: 1297-1303. doi: 10.1101/gr.107524.110

Mitchell G, et al. 2010. *Staphylococcus aureus* sigma B-dependent emergence of small-colony variants and biofilm production following exposure to *Pseudomonas aeruginosa* 4-hydroxy-2-heptylquinoline-N-oxide. Bmc Microbiology 10. doi: 33

10.1186/1471-2180-10-33
Nistico L, et al. 2011. Adenoid reservoir for pathogenic biofilm bacteria. Journal of Clinical Microbiology 49: 1411-1420. doi: JCM.00756-10 [pii]

10.1128/JCM.00756-10
O'Brien KL, et al. 2009. Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. Lancet 374: 893-902.
Overbeek R, et al. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Research 42: D206-D214. doi: Doi 10.1093/Nar/Gkt1226
Pai R, Gertz RE, Beall B 2006. Sequential multiplex PCR approach for determining capsular serotypes of Streptococcus pneumoniae isolates. Journal of Clinical Microbiology 44: 124-131. doi: 10.1128/jcm.44.1.124-131.2006
Philips BJ, Meguer JX, Redman J, Baker EH 2003. Factors determining the appearance of glucose in upper and lower respiratory tract secretions. Intensive Care Med 29: 2204-2210. doi: 10.1007/s00134-003-1961-2
Potera C 1999. Microbiology - Forging a link between biofilms and disease. Science 283: 1837-+. doi: DOI 10.1126/science.283.5409.1837
Prudhomme M, Libante V, Claverys JP 2002. Homologous recombination at the border: Insertion-deletions and the trapping of foreign DNA in *Streptococcus pneumoniae*. Proc Natl Acad Sci U S A 99: 2100-2105. doi: DOI 10.1073/pnas.032262999
Rabatinova A, et al. 2013. The delta subunit of RNA polymerase is required for rapid changes in gene expression and competitive fitness of the cell. J Bacteriol 195: 2603-2611. doi: 10.1128/JB.00188-13
Robinson JT, et al. 2011. Integrative genomics viewer. Nat Biotechnol 29: 24-26. doi: 10.1038/nbt.1754
Rukke HV, Hegna IK, Petersen FC 2012. Identification of a functional capsule locus in *Streptococcus mitis*. Mol Oral Microbiol 27: 95-108. doi: 10.1111/j.2041-1014.2011.00635.x
Sanchez CJ, et al. 2011. Streptococcus pneumoniae in biofilms are unable to cause invasive disease due to altered virulence determinant production. PLoS One 6: e28738. doi: 10.1371/journal.pone.0028738
Schultz J, Milpetz F, Bork P, Ponting CP 1998. SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci U S A 95: 5857-5864.
Simell B, et al. 2012. The fundamental link between pneumococcal carriage and disease. Expert Rev Vaccines 11: 841-855. doi: 10.1586/erv.12.53
Sternberg C, et al. 1999. Distribution of bacterial growth activity in flow-chamber biofilms. Applied and Environmental Microbiology 65: 4108-4117.
Sztajer H, et al. 2008. Autoinducer-2-regulated genes in Streptococcus mutans UA159 and global metabolic effect of the luxS mutation. J Bacteriol 190: 401-415. doi: 10.1128/JB.01086-07
Thorvaldsdottir H, Robinson JT, Mesirov JP 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14: 178-192. doi: 10.1093/bib/bbs017
Tocheva AS, et al. 2011. Declining serotype coverage of new pneumococcal conjugate vaccines relating to the carriage of *Streptococcus pneumoniae* in young children. Vaccine 29: 4400-4404. doi: 10.1016/j.vaccine.2011.04.004
Waite RD, Struthers JK, Dowson CG 2001. Spontaneous sequence duplication within an open reading frame of the pneumococcal type 3 capsule locus causes high-frequency phase variation. Molecular Microbiology 42: 1223-1232.
Webb JS, Lau M, Kjelleberg S 2004. Bacteriophage and phenotypic variation in Pseudomonas aeruginosa biofilm development. Journal of Bacteriology 186: 8066-8073. doi: 10.1128/jb.186.23.8066-8073.2004
Weiss A, Ibarra JA, Paoletti J, Carroll RK, Shaw LN 2014. The delta Subunit of RNA Polymerase Guides Promoter Selectivity and Virulence in Staphylococcus aureus. Infect Immun 82: 1424-1435. doi: 10.1128/IAI.01508-14

Xue X, Li J, Wang W, Sztajer H, Wagner-Dobler I 2012. The global impact of the delta subunit RpoE of the RNA polymerase on the proteome of Streptococcus mutans. Microbiology 158: 191-206. doi: 10.1099/mic.0.047936-0

Xue X, et al. 2011. Lack of the delta subunit of RNA polymerase increases virulence related traits of Streptococcus mutans. PLoS One 6: e20075. doi: 10.1371/journal.pone.0020075

Xue X, Tomasch J, Sztajer H, Wagner-Dobler I 2010. The delta subunit of RNA polymerase, RpoE, is a global modulator of Streptococcus mutans environmental adaptation. J Bacteriol 192: 5081-5092. doi: 10.1128/JB.00653-10

Zerbino DR 2010. Using the Velvet de novo assembler for short-read sequencing technologies. Curr Protoc Bioinformatics Chapter 11: Unit 11 15. doi: 10.1002/0471250953.bi1105s31

Figure 1: Colony Forming Unit enumeration of biofilm-derived pneumococcal variants. Cells were harvested from triplicate pneumococcal biofilms from two independent experiments at time points 1, 3, 6 and 9 days and plated onto CBA to assess for changes in colony morphology. Variants were defined as follows; small colony variants (SCVs) (<0.5 mm), medium colony variants (MCVs) (0.5-1 mm) and typical colony variants (TCVs) (>1 mm). Data represents the mean CFU counts from the two biofilm experiment. Error bars represent standard error. Images of colonies were taken on CBA under 6 x objective on a *Leica* dissection Stereomicroscope. Scale bar = 1 mm.

Figure 2: Colony quantification of serotype 22F biofilm-derived colony variants. Diameter values of the three distinct biofilm-derived colony variant populations harvested from pneumococcal biofilms quantified using ImageJ analysis software. A total of nine colony variants, from triplicate biofilms were measured. Numbers signify mean values.

Figure 3: Biofilm formation of the biofilm-derived colony variants. Three-day old biofilms of (A) the WT ancestral strain, (B) small colony variant, (C) medium colony variant and (D) typical colony variant were visualized on a Leica TCS SP5 confocal laser scanning microscope using *Bac*Light live/dead stain. Green cells indicate live cells and red cells indicate dead cells. Z-Scans were performed every 0.3 µm on each field of view. White bars represent 25 µm.

Figure 4: COMSTAT analysis of biofilm-derived colony variants. Biofilm formation of the biofilm-derived variants quantified using COMSTAT 1 and the program Matlab (R2012a) (Heydorn, et al. 2000) for triplicate 3-day-old biofilms of each variant type, grown in MatTek culture plates under static conditions. Graph depicts (A) the mean average thickness and maximum thickness, (B) mean biomass and(C) mean surface area. Error bars represent 95% confidence intervals.

Figure 5: Mean growth rate of biofilm-derived SCV phenotype. The growth rate was assessed for SCVs from the exponential growth phase (between 4 and 8 hours of growth) from triplicate growth experiments. SCV rate was calculated using pooled data from all 12 sequenced SCVs. A 2-sample t-

test was used to compare SCV rate to the WT rate. Numbers signify mean values. Error bars represent 95% confidence intervals.

Figure 6: Capsule quantification of the SCV phenotype. Capsule quantification was determined by staining the pneumococcal capsule with Stains-all solution. The absorbance was measured at optical density 640 nm and subtracted from the negative control. Capsule quantification was calculated using pooled data from all 12 sequenced SCVs; graph depicts mean absorbance. Error bars represent 95% confidence intervals. Graph represents pooled data from all 12 sequenced SCVs.

Table 1: SCV Mutations Identified Using Next Generation Sequencing

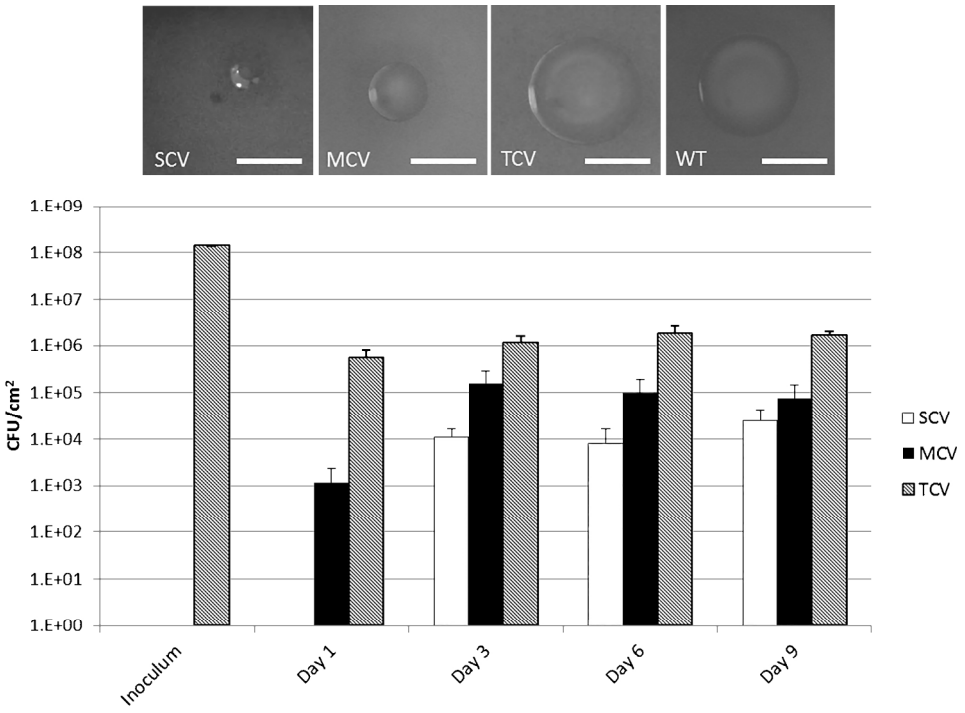

Figure 1: Colony Forming Unit enumeration of biofilm-derived pneumococcal variants. Cells were harvested from triplicate pneumococcal biofilms from two independent experiments at time points 1, 3, 6 and 9 days and plated onto CBA to assess for changes in colony morphology. Variants were defined as follows; small colony variants (SCVs) (<0.5 mm), medium colony variants (MCVs) (0.5-1 mm) and typical colony variants (TCVs) (>1 mm). Data represents the mean CFU counts from the two biofilm experiment. Error bars represent standard error. Images of colonies were taken on CBA under 6 x objective on a Leica dissection Stereomicroscope. Scale bar = 1 mm.
766x549mm (72 x 72 DPI)

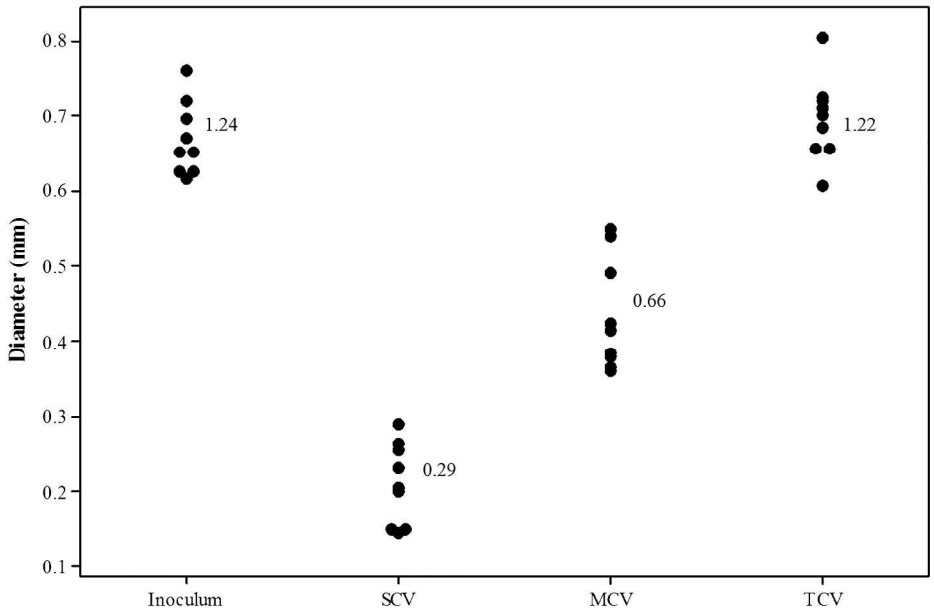

Figure 2: Colony quantification of serotype 22F biofilm-derived colony variants. Diameter values of the three distinct biofilm-derived colony variant populations harvested from pneumococcal biofilms quantified using ImageJ analysis software. A total of nine colony variants, from triplicate biofilms were measured. Numbers signify mean values.
1032x705mm (72 x 72 DPI)

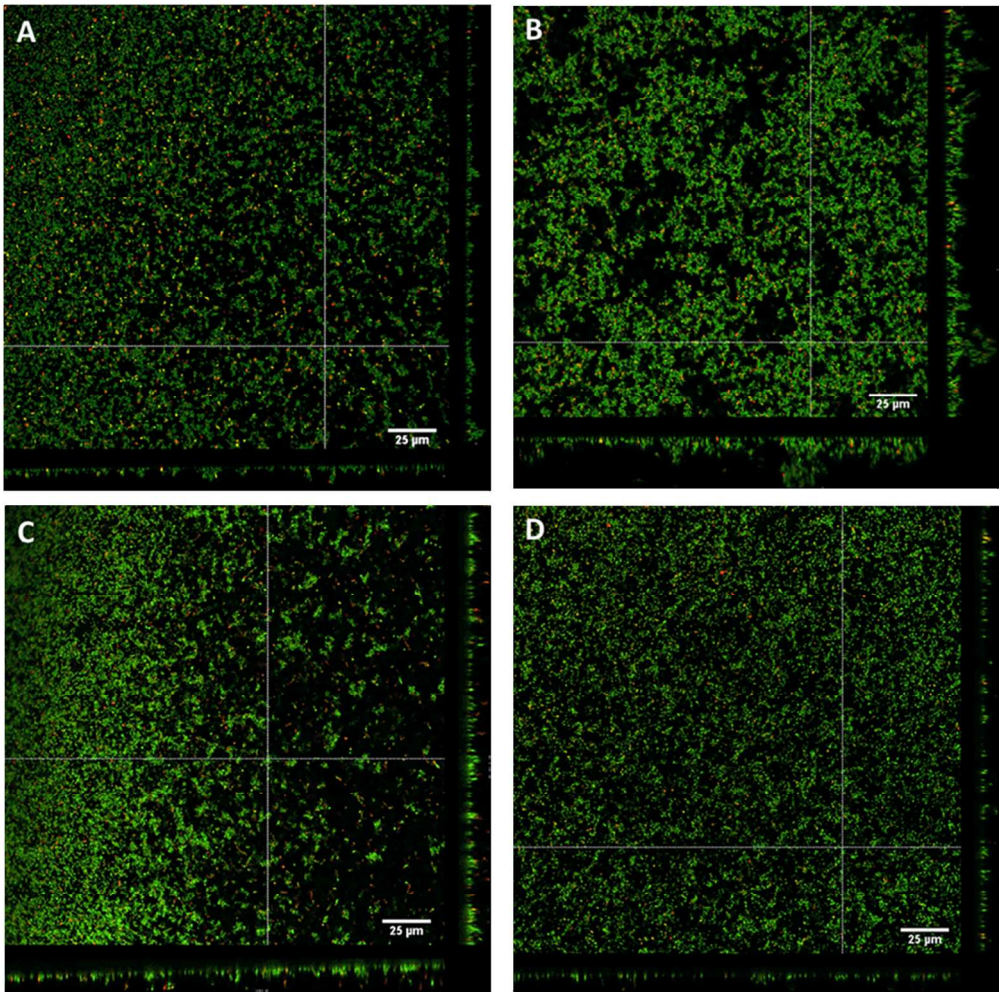

Figure 3: Biofilm formation of the biofilm-derived colony variants. Three-day old biofilms of (A) the WT ancestral strain, (B) small colony variant, (C) medium colony variant and (D) typical colony variant were visualized on a Leica TCS SP5 confocal laser scanning microscope using BacLight live/dead stain. Green cells indicate live cells and red cells indicate dead cells. Z-Scans were performed every 0.3 µm on each field of view. White bars represent 25 µm.
633x629mm (72 x 72 DPI)

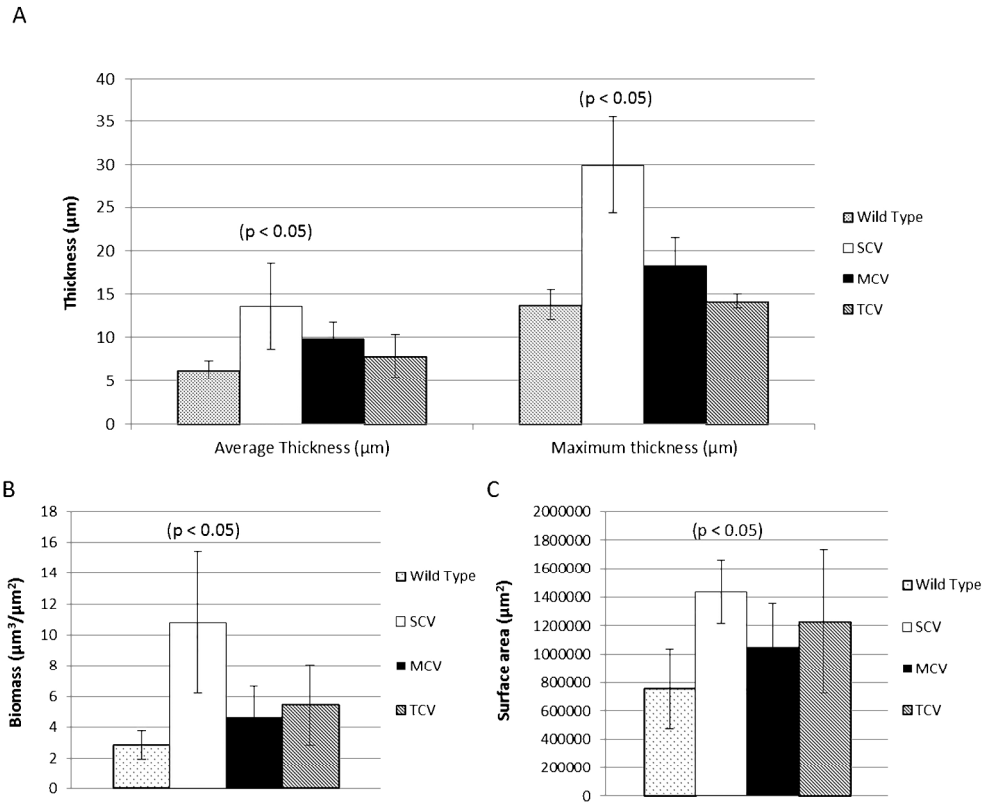

Figure 4: COMSTAT analysis of biofilm-derived colony variants. Biofilm formation of the biofilm-derived variants quantified using COMSTAT 1 and the program Matlab (R2012a) (Heydorn, et al. 2000) for triplicate 3-day-old biofilms of each variant type, grown in MatTek culture plates under static conditions. Graph depicts (A) the mean average thickness and maximum thickness, (B) mean biomass and(C) mean surface area. Error bars represent 95% confidence intervals.
810x703mm (72 x 72 DPI)

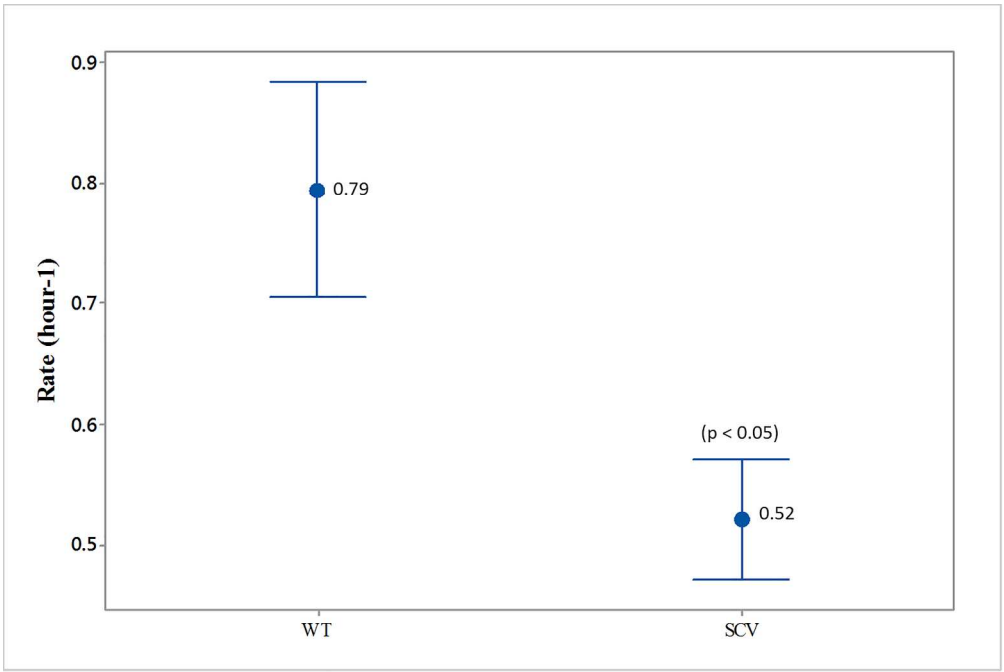

Figure 5: Mean growth rate of biofilm-derived SCV phenotype. The growth rate was assessed for SCVs from the exponential growth phase (between 4 and 8 hours of growth) from triplicate growth experiments. SCV rate was calculated using pooled data from all 12 sequenced SCVs. A 2-sample t-test was used to compare SCV rate to the WT rate. Numbers signify mean values. Error bars represent 95% confidence intervals. 1016x678mm (72 x 72 DPI)

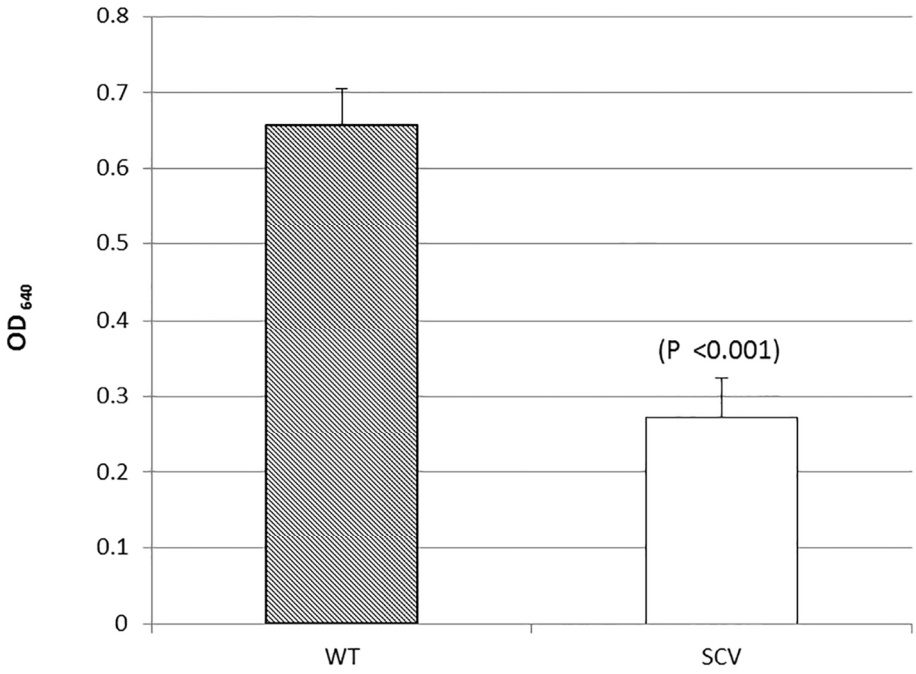

Figure 6: Capsule quantification of the SCV phenotype. Capsule quantification was determined by staining the pneumococcal capsule with Stains-all solution. The absorbance was measured at optical density 640 nm and subtracted from the negative control. Capsule quantification was calculated using pooled data from all 12 sequenced SCVs; graph depicts mean absorbance. Error bars represent 95% confidence intervals. Graph represents pooled data from all 12 sequenced SCVs.
579x480mm (72 x 72 DPI)

Variant*	Reference Position	WT	Variant	Gene Description	Effect	Coverage	Percentage confidence
SCV1D3E1	423010	G	T	DNA-directed RNA polymerase delta subunit	E173Stop	49x	98
SCV5D3E1	422977	G	T	DNA-directed RNA polymerase delta subunit	E163Stop	69x	100
SCV3D9E1	328131	A	C	Choline binding protein J	Q118P	30x	96.7
SCV3D9E1	422921	C	A	DNA-directed RNA polymerase delta subunit	S144Stop	44x	100
SCV3D9E1	908199	T	G	Hydrolase, putative	L26V	27x	96.4
SCV7D9E1	422977	G	T	DNA-directed RNA polymerase delta subunit	E163Stop	243x	100
SCV9D9E1	422795-423059	A	-	DNA-directed RNA polymerase delta subunit	Deletion	n/a	n/a
SCV9D9E1	1620361	C	A	Iron compound ABC uptake transporter permease protein PiuB	A278D	47x	97.4
SCV1D3E2	422765-422894	A	-	DNA-directed RNA polymerase delta subunit	Deletion	n/a	n/a
SCV2D6E2	422799-423063	G	-	DNA-directed RNA polymerase delta subunit	Deletion	n/a	n/a
SCV4D6E2	422947	G	T	DNA-directed RNA polymerase delta subunit	E153Stop	197x	99.5
SCV4D6E2	591574	G	T	Beta-galactosidase	A146S	129x	98.5
SCV4D6E2	1396435	C	T	Glycine/D-amino acid oxidases family	R412H	137x	100
SCV6D6E2	423020	-	G	DNA-directed RNA polymerase delta subunit	Insertion	92x	93.9
SCV5D9E2	423028	G	T	DNA-directed RNA polymerase delta subunit	E180Stop	146x	100
SCV6D9E2	422962	G	T	DNA-directed RNA polymerase delta subunit	E158Stop	122x	99.2
SCV7D9E2	423028-423090	G	-	DNA-directed RNA polymerase delta subunit	Deletion	n/a	n/a
SCV7D9E2	1035274	G	A	Streptococcal histidine triad protein	Q111Stop	106x	100

Footnote: * D3=Day 3 biofilm, D6=Day 6 Biofilm, D9=Day 9 biofilm; E1=Experiment 1, E2=Experiment 2.

Table 1: Small Colony Variant Mutations Identified Using Next Generation Sequencing
1000x611mm (72 x 72 DPI)