



Capacity To Utilize Raffinose Dictates Pneumococcal Disease Phenotype

Vikrant Minhas,^a Richard M. Harvey,^a Lauren J. McAllister,^a Torsten Seemann,^{b,c} Anna E. Syme,^b  Sarah L. Baines,^c James C. Paton,^a Claudia Trappetti^a

^aResearch Centre for Infectious Diseases, Department of Molecular and Biomedical Science, University of Adelaide, Adelaide, Australia

^bMelbourne Bioinformatics, The University of Melbourne, Melbourne, Australia

^cDepartment of Microbiology and Immunology, The University of Melbourne, Melbourne, Australia

ABSTRACT *Streptococcus pneumoniae* is commonly carried asymptotically in the human nasopharynx, but it also causes serious and invasive diseases such as pneumonia, bacteremia, and meningitis, as well as less serious but highly prevalent infections such as otitis media. We have previously shown that closely related pneumococci (of the same capsular serotype and multilocus sequence type [ST]) can display distinct pathogenic profiles in mice that correlate with clinical isolation site (e.g., blood versus ear), suggesting stable niche adaptation within a clonal lineage. This has provided an opportunity to identify determinants of disease tropism. Genomic analysis identified 17 and 27 single nucleotide polymorphisms (SNPs) or insertions/deletions in protein coding sequences between blood and ear isolates of serotype 14 ST15 and serotype 3 ST180, respectively. SNPs in raffinose uptake and utilization genes (*rafR* or *rafK*) were detected in both serotypes/lineages. Ear isolates were consistently defective in growth in media containing raffinose as the sole carbon source, as well as in expression of raffinose pathway genes *aga*, *rafG*, and *rafK*, relative to their serotype/ST-matched blood isolates. Similar differences were also seen between serotype 23F ST81 blood and ear isolates. Analysis of *rafR* allelic exchange mutants of the serotype 14 ST15 blood and ear isolates demonstrated that the SNP in *rafR* was entirely responsible for their distinct *in vitro* phenotypes and was also the determinant of differential tropism for the lungs versus ear and brain in a mouse intranasal challenge model. These data suggest that the ability of pneumococci to utilize raffinose determines the nature of disease.

IMPORTANCE *S. pneumoniae* is a component of the commensal nasopharyngeal microflora of humans, but from this reservoir, it can progress to localized or invasive disease with a frequency that translates into massive global morbidity and mortality. However, the factors that govern the switch from commensal to pathogen, as well as those that determine disease tropism, are poorly understood. Here we show that capacity to utilize raffinose can determine the nature of the disease caused by a given pneumococcal strain. Moreover, our findings provide an interesting example of convergent evolution, whereby pneumococci belonging to two unrelated serotypes/lineages exhibit SNPs in separate genes affecting raffinose uptake and utilization that correlate with distinct pathogenic profiles *in vivo*. This further underscores the critical role of differential carbohydrate metabolism in the pathogenesis of localized versus invasive pneumococcal disease.

KEYWORDS *Streptococcus pneumoniae*, carbohydrate metabolism, otitis media, pneumonia, single nucleotide polymorphisms, virulence

Citation Minhas V, Harvey RM, McAllister LJ, Seemann T, Syme AE, Baines SL, Paton JC, Trappetti C. 2019. Capacity to utilize raffinose dictates pneumococcal disease phenotype. mBio 10:e02596-18. <https://doi.org/10.1128/mBio.02596-18>.

Editor Larry S. McDaniel, University of Mississippi Medical Center

Copyright © 2019 Minhas et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to James C. Paton, james.paton@adelaide.edu.au, or Claudia Trappetti, claudia.trappetti@adelaide.edu.au.

This article is a direct contribution from a Fellow of the American Academy of Microbiology. Solicited external reviewers: Jeffrey Weiser, NYU Lagone Health; Andrew Camilli, Tufts University School of Medicine.

Received 22 November 2018

Accepted 29 November 2018

Published 15 January 2019

Streptococcus pneumoniae (the pneumococcus) is one of the world's foremost bacterial pathogens, killing 1 to 2 million people each year. In spite of this, it is considered part of the "normal" nasopharyngeal microflora, asymptotically colonizing up to 65% of individuals; these carriers are the principal reservoirs for transmission of *S. pneumoniae* in the community (1, 2). In a small proportion of carriers, which nevertheless translates into globally significant numbers, *S. pneumoniae* invades from its nasopharyngeal reservoir to cause disease: e.g., by aspiration into the lungs to cause pneumonia, by direct or indirect invasion of the blood (bacteremia) or central nervous system (meningitis), or by ascension of the eustachian tube to access the middle ear and cause otitis media (OM) (1, 2). However, the molecular mechanisms whereby pneumococci transition from a commensal lifestyle to cause either localized or invasive disease are poorly understood.

The pneumococcus is a genetically plastic and diverse species, comprising at least 98 capsular serotypes, superimposed on more than 12,000 clonal lineages (sequence types [STs]) recognizable by multilocus sequence typing (3). It has a core genome of roughly 1,500 genes, with the remaining 30% of the genome present as accessory regions (ARs), present in some but not all clonal lineages. Individual *S. pneumoniae* strains can differ markedly in their virulence phenotypes, including their capacity to colonize the nasopharynx, spread from person to person, or progress to either localized or invasive infections. Capsule switching experiments have shown that both serotype and genetic background (i.e., ST) influence virulence (4, 5), but strain complexity has complicated attempts to examine whether there is any association between a given clonal lineage or serotype and propensity to cause localized rather than invasive infections.

Previous studies in our laboratory have shown that *S. pneumoniae* clinical isolates belonging to the same serotype and ST may display distinct virulence phenotypes in mice, in accordance with their original site of isolation in humans (blood versus ear). After intranasal (i.n.) challenge, serotype 3 blood isolates belonging to ST180, ST232, and ST233 did not stably colonize the nasopharynx, but spread to the blood in the majority of mice; none spread to the ear. In contrast, ear isolates colonized the nasopharynx at higher levels than the respective ST-matched blood isolates and also spread to the ear compartment; none caused bacteremia (6). In a separate study, serotype 14 (ST15) blood and ear isolates all exhibited a similar capacity to colonize the nasopharynx, but significant differences were observed between bacterial loads in other host niches. Blood isolates caused pneumonia in most animals, whereas ear isolates were not detected in the lungs of any of the mice 24 h post-intranasal challenge. Conversely, ST15 ear isolates, but not blood isolates, were able to spread to the brain, and in the ear compartment, the bacterial load and proportion of infected mice were significantly greater for mice challenged with ear rather than blood isolates (7). Thus, strains within a clonal lineage appear to be exhibiting stable niche adaptation.

Although members of the same serotype and ST type have very closely related genetic backbones, they are not necessarily identical and may have acquired distinct ARs or other genetic changes, such as single nucleotide polymorphisms (SNPs) or insertions or deletions (indels). In the present study, we have compared the genomes of representative serotype 14 ST15 and serotype 3 ST180 blood and ear isolates to determine whether such differences can account for their distinct virulence phenotypes. We show that SNPs in loci responsible for uptake and utilization of the trisaccharide raffinose are the ultimate determinant of disease progression.

RESULTS

Genetic differences between serotype/ST-matched blood and ear isolates. In the first instance, draft genomes of serotype 14 ST15 strains 4559 (blood isolate) and 947 (ear isolate) were assembled from PacBio and MiSeq data and then compared (see Materials and Methods). The only differences in ARs were the presence of a 35-kb prophage and a 3.2-kb plasmid in 4559 and not 947, but these ARs were not present in other serotype 14 ST15 blood isolates in our collection (results not shown). Seven-

TABLE 1 Genes containing indels or SNPs that led to amino acid changes, identified from the whole-genome variant calling analysis between 4559 and 947^a

Locus tag in 947	Gene	Product	Change in aa sequence in 4559 relative to 947
0862	<i>pfkA</i>	ATP-dependent 6-phosphofructokinase	S212G
1153	<i>glgA</i>	Glycogen synthase	E174G
1345	<i>pncB</i>	Nicotinate phosphoribosyltransferase	N434D
1631	<i>scrR</i>	HTH-type transcriptional regulator	ΔL27-G28
1803	<i>rafR</i>	HTH-type transcriptional regulator	D249G
1255	<i>pyrP</i>	Uracil permease	V65A
1737	<i>piuA</i>	Fe ³⁺ import ATP-binding protein	G141V
2020		ABC transporter ATP-binding protein	Y508N
0330	<i>cpsE</i>	CPS glycosyltransferase	L43P
1139	<i>iga</i>	Immunoglobulin A1 protease	Premature stop 1905 (4559) due to indel
1594	<i>nanB</i>	Sialidase B	Premature stop 362 (947) due to indel
1741	<i>pfbA</i>	Plasmin and fibronectin-binding protein A	T318M
0945	<i>coiA</i>	Competence protein	E78K
1141	<i>addA</i>	ATP-dependent helicase/nuclease subunit A	I980M
1060		Acetyl transferase	C101G
1194		Cytosolic protein containing multiple CBS domains	Premature stop 104 (947) due to SNP
1731		Hypothetical protein (no Pfam match)	H32P

^aResults for the raffinose pathway gene *rafR* are in boldface.

teen SNPs and indels present within protein coding sequences of 4559 and 947 resulting in a change in the predicted amino acid sequence are listed in Table 1. The genes affected included those predicted to be involved in metabolism and energy production, transcriptional regulation, transporters, and putative virulence factors. Among the latter category, an SNP resulting in a L43P substitution was identified in *cpsE*, which encodes the glycosyl transferase that initiates assembly of the capsular polysaccharide (CPS) repeat unit. However, we have previously shown that there is no difference in total CPS production between 4559 and 947 (7). The SNP in the putative plasmin and fibronectin-binding protein gene *pfbA* is also a conservative T318M substitution. The protein encoded by *iga* is truncated in 4559 compared to 947, but only by four amino acids. On the other hand, the *nanB* sequence in 947 has a premature stop codon that truncates the protein by 330 amino acids (47% of the 4559 protein), presumably inactivating the gene product. Mutagenesis studies have previously shown that NanB contributes to colonization of both the upper and lower respiratory tract of mice, albeit to a lesser extent than the major neuraminidase NanA (8). Nevertheless, 4559 and 947 colonize the nasopharynx equally well (7). Interestingly, SNPs were identified in two metabolic genes, coding for ATP-dependent 6-phosphofructokinase (*pfkA*) and a glycogen synthase (*glgA*), as well as in two helix-turn-helix (HTH)-type transcriptional regulators, *scrR* and *rafR*, involved in metabolism of sucrose and raffinose, respectively. Given the importance of carbohydrate metabolism to *S. pneumoniae* (9), we employed a phenotypic microarray to compare the capacity of 4559 and 947 to metabolize over 100 different carbohydrates (see Materials and Methods). The only difference observed between the ear and blood isolates was a reduced capacity of the former (947) to grow in medium containing raffinose as the sole carbon source (data not shown).

Genetic differences between ear/blood isolate pairs that are common to two unrelated serotypes/ST lineages would be strong candidates for determinants of tissue tropism. Genomic comparisons were therefore also made between two serotype 3 ST180 ear and blood isolates (strains 180/2 and 180/15, respectively), which like the serotype 14 ST15 isolates, have previously been shown to exhibit distinct tissue tropism in mice in accordance with clinical isolation site (6). There were no differences in ARs between the two strains, while SNPs and indels impacting the deduced amino acid sequence for 27 genes were identified (Table 2). Interestingly, there were no affected genes in common with those in Table 1. However, an I227T SNP was detected in the serotype 3 *rafK* gene, encoding the ATP-binding protein component of the raffinose

TABLE 2 Genes containing indels or SNPs that led to amino acid changes, identified from the whole-genome variant calling analysis between 180/2 and 180/15^a

Locus tag in 180/2	Gene	Product ^b	Change in aa sequence of 180/15 relative to 180/2
100	<i>purN</i>	Phosphoribosyl-glycinamide	G81A
254	<i>rpsJ</i>	30S ribosomal protein S10	Y58D
285		Hypothetical protein	A81S
314	<i>cdsA</i>	Phosphatidate cytidyltransferase	M14I
335	<i>adhP</i>	Alcohol dehydrogenase 1	M210V
403	<i>fabK</i>	Enoyl-[acyl-carrier-protein] reductase	I1029T
449		Nitronate monooxygenase	I55M
512	<i>glnA</i>	Glutamine synthetase	F22L
645	<i>nhaK</i>	Sodium, potassium, lithium and rubidium/H ⁺ antiporter	G190D
741		VanZ family protein	C149W
996		Hypothetical protein	H38R
1121	<i>clcA</i>	H ⁺ /Cl ⁻ exchange transporter	M131I
1138	<i>ptsH</i>	Phosphocarrier protein HPr	I14V
1172		Formate/nitrate transporter	A211E
1194	<i>glnP</i>	Glutamine transport system permease protein	S662A
1234		SpF43_sRNA	Y31C
1306	<i>alaS</i>	Alanine-tRNA ligase	E18A
1387	<i>apbE</i>	FAD:protein FMN transferase	M52I
1404		LPXTG cell wall anchor domain-containing protein	ΔK112-Q119; G125K, E126T, P127E, E130V, K131N, I133D; ΔQ135-P178
1491	<i>rafK</i>	Raffinose import ATP-binding protein	I227T
1616	<i>dnaB</i>	DNA helicase	C375R
1760	<i>fepD_2</i>	Ferric enterobactin transport system permease protein	S248G
1863	<i>rpoC</i>	DNA-directed RNA polymerase subunit beta	D76E
1878	<i>acyP</i>	Acyolphosphatase	V4I
1887	<i>rsgA</i>	Small ribosomal subunit biogenesis	G40S
2045	<i>aspS</i>	Aspartate tRNA ligase	E51V
2100	<i>dltD</i>	D-Alanyl-lipoteichoic acid biosynthesis protein	D151E

^aResults for the raffinose pathway gene *rafK* are in boldface.

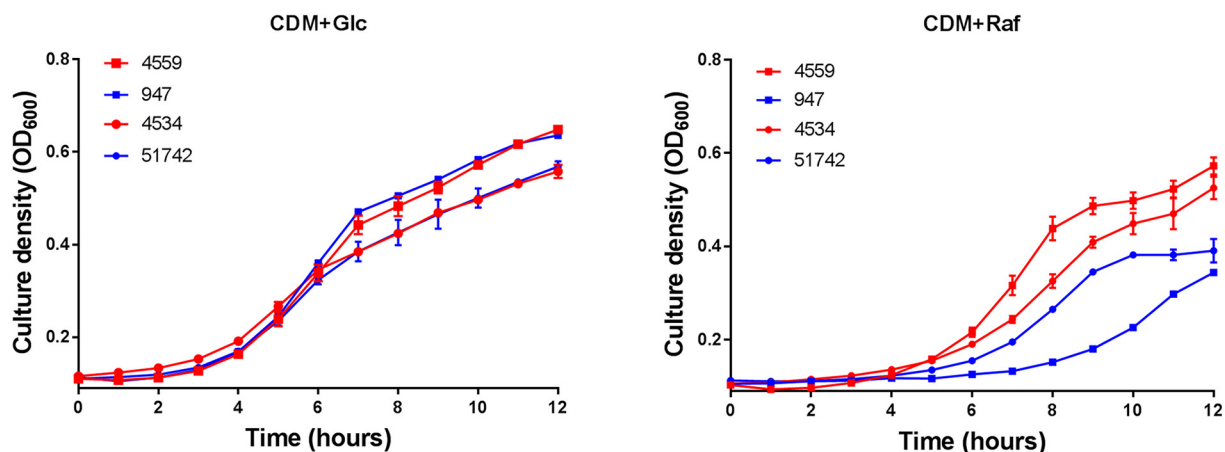
^bFAD, flavin adenine dinucleotide; FMN, flavin mononucleotide.

ABC transporter. *RafK* is known to be essential for activation of other *raf* operon genes, and the SNP identified in ST180 isolates is located in the conserved regulatory domain motif 1 (10). Thus, potential defects in raffinose uptake/metabolism appear to be a common feature of ear isolates from both serotypes/lineages.

Blood isolates utilize raffinose more efficiently than ear isolates. In view of the SNPs in genes associated with raffinose metabolism between ear and blood isolates in two unrelated serotypes/STs and the fact that the serotype 14 ear and blood isolates differed only in their ability to metabolize raffinose on phenotypic microarray analysis, *in vitro* growth phenotypes were further investigated. Strains 4559 and 947, as well as another pair of serotype 14 ST15 blood and ear isolates (4534 and 51742, respectively), were grown in a chemically defined medium (CDM) with either glucose or raffinose as the sole carbon source (designated CDM+Glc and CDM+Raf, respectively) (Fig. 1). In CDM+Glc, there were no significant differences in growth rates between blood and ear isolates. However, in CDM+Raf, the two blood isolates grew at a higher rate and to a higher final culture density (optical density at 600 nm [OD₆₀₀]) than either of the serotype 14 ST15 ear isolates. Similarly, there was no significant difference in growth rates of the serotype 3 ST180 ear and blood isolates (180/15 and 180/2, respectively) in CDM+Glc, but the blood isolate grew better than the ear isolate in CDM+Raf (Fig. 1). Thus, defective growth in raffinose appears to be a common defect in ear isolates relative to serotype/ST-matched blood isolates.

The raffinose uptake/utilization operon in *S. pneumoniae* comprises genes encoding transcriptional regulators (*rafR* and *rafS*), an α -galactosidase (*aga*), the ABC transporter substrate-binding protein and two cognate permeases (*rafE*, *rafF*, and *rafG*), a sucrose phosphorylase (*gtfA*), and a protein of unknown function (*rafX*), as well as the ATP binding protein component of the transporter (*rafK*), which is independently located in the genome (11) (Fig. 2). To determine if the difference in ability to utilize raffinose

Serotype 14 ST15



Serotype 3 ST180

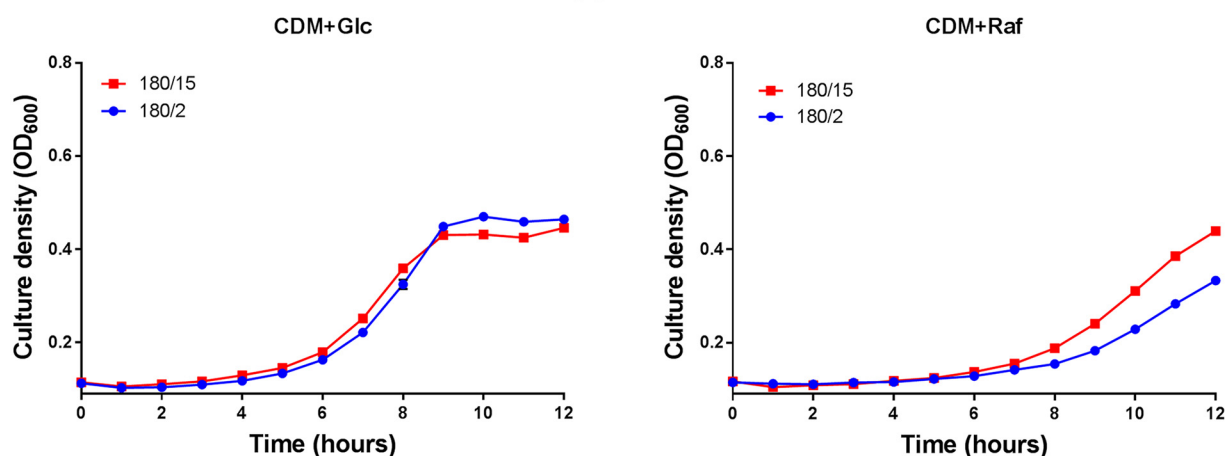


FIG 1 Differential growth of blood and ear isolates in raffinose. *S. pneumoniae* serotype 14 ST15 blood isolates 4559 and 4534 and ear isolates 947 and 51742 were grown in 200 μ l CDM supplemented with 0.5% glucose (CDM+Glc) or 0.5% raffinose (CDM+Raf). Similar growth studies were also performed for serotype 3 ST180 strains 180/15 (blood isolate) and 180/2 (ear isolate). OD₆₀₀ was measured every hour for 12 h. Data are mean OD₆₀₀ \pm standard deviation (SD) from triplicate assays.

between the blood and ear isolates corresponded with raffinose operon gene expression, *S. pneumoniae* serotype 14 ST15 strains 4559, 947, 4534, and 51742 and serotype 3 ST180 strains 180/2 and 180/15 were grown to the same OD₆₀₀ (0.2) in CDM+Glc and then washed and resuspended in CDM+Raf and incubated for a further 30 min. RNA

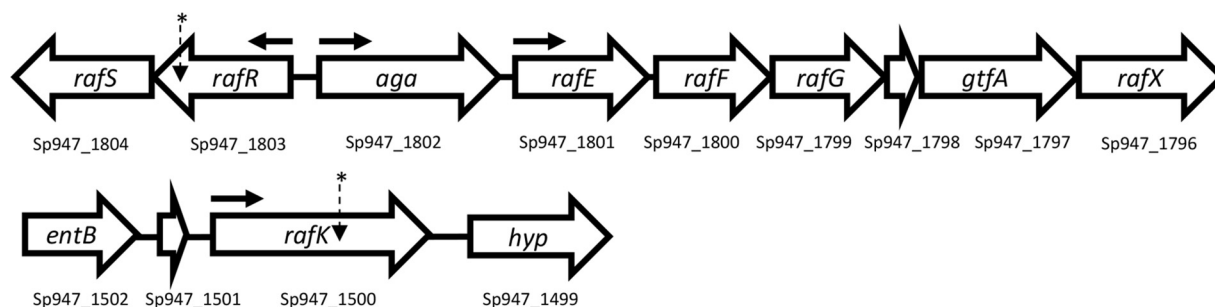


FIG 2 Genetic loci encoding raffinose uptake and utilization in *S. pneumoniae*. The numbers below each gene refer to the locus tags in the serotype 14 ST15 947 genome. The locations of SNPs in serotype 14 ST15 and serotype 3 ST180 isolates are indicated with asterisks; horizontal arrows show the locations of promoters.

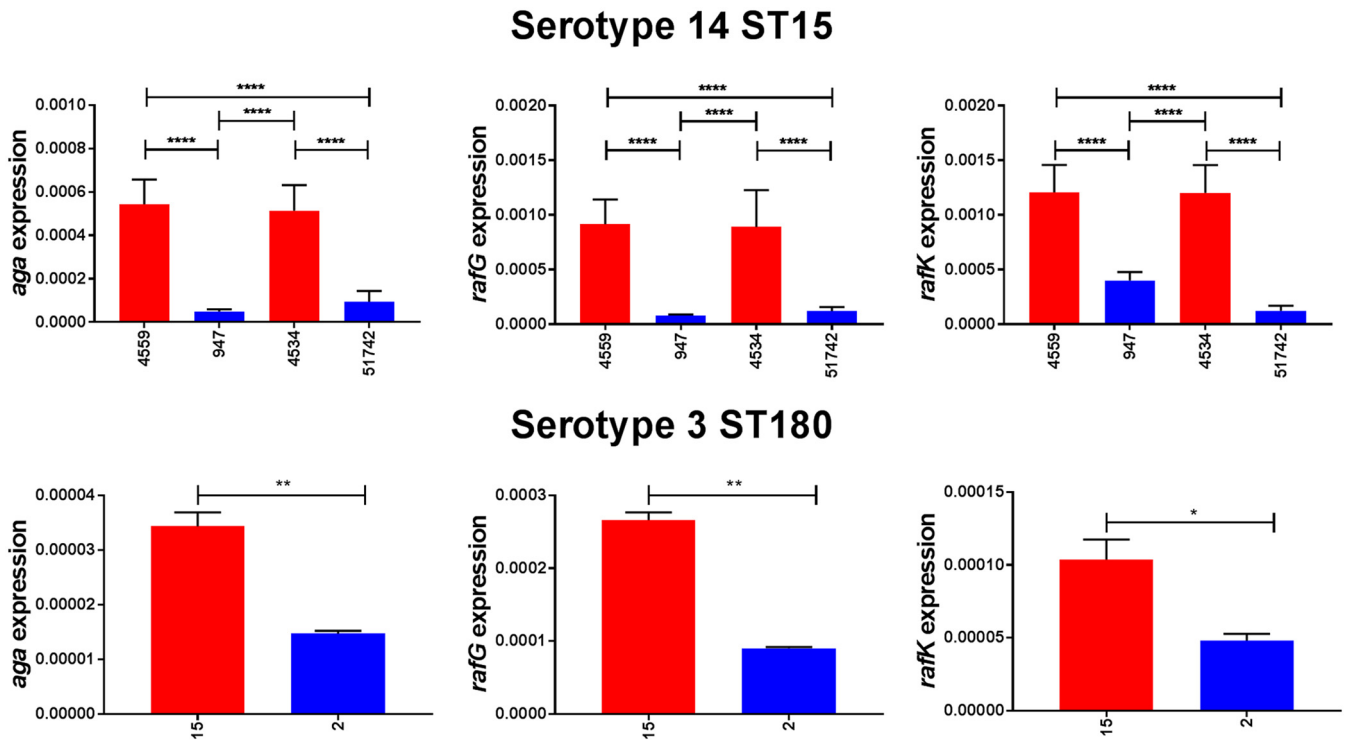


FIG 3 Expression of raffinose pathway genes by serotype 14 and 3 blood and ear isolates. The indicated strains were grown in CDM+Glc to an OD_{600} of 0.2, washed and resuspended in CDM+Raf, and then incubated at 37°C for a further 30 min. RNA was then extracted, and levels of *aga*, *rafG*, and *rafK* mRNA were analyzed by qRT-PCR using 16S rRNA as an internal control (see Materials and Methods). The data presented are the means \pm SD from three independent experiments. *, $P < 0.05$, **, $P < 0.01$, and ****, $P < 0.0001$, by unpaired *t* test.

was then extracted, and levels of *aga*, *rafG*, and *rafK* mRNA, representative of each of the three *rafR*-regulated transcriptional units, were then measured relative to 16S rRNA by quantitative real-time reverse transcription-PCR (qRT-PCR). In every case, expression levels for all three genes were significantly greater in the blood isolates than in the respective ear isolates (Fig. 3).

As further confirmation, blood and ear isolates belonging to serotype 23F ST81 were also tested for growth in CDM+Glc and CDM+Raf, as well as for expression of *aga*, *rafG*, and *rafK* (Fig. 4). Again, the blood isolate grew to a higher OD_{600} than the ear isolate in CDM+Raf, but not in CDM+Glc. Moreover, expression of all three *raf* genes was significantly higher in the blood isolate than in the ear isolate.

The SNP in 947 *rafR* is responsible for its raffinose phenotype. In order to test whether the distinct *in vitro* and *in vivo* phenotype of 947 relative to 4559 was attributable to the SNP in *rafR*, allelic-exchange mutagenesis was performed in 4559 and 947, generating a 4559 derivative with its *rafR* allele replaced by that from 947 (designated 4559^{947rafR}) and a 947 derivative expressing the 4559 *rafR* allele (947^{4559rafR}) (see Materials and Methods). Growth assays in CDM+Glc showed no significant differences in growth rates between 4559, 947, 4559^{947rafR}, and 947^{4559rafR}. However, in CDM+Raf, growth of 4559^{947rafR} was at least as poor as that of 947, while growth of 947^{4559rafR} was similar to that of 4559 (Fig. 5A). Expression of *aga*, *rafG*, and *rafK* was then examined in 4559, 947, 4559^{947rafR}, and 947^{4559rafR} by qRT-PCR after 30 min of growth in CDM+Raf. For all three genes, expression levels in 947^{4559rafR} were indistinguishable from those in 4559, while expression in 4559^{947rafR} was essentially the same as that in 947 (Fig. 5B). Thus, exchange of *rafR* alleles between 4559 and 947 significantly impacts both growth phenotype and *raf* operon gene expression in CDM+Raf.

Virulence phenotypes of 4559 and 947 and their *rafR* exchange mutants. In order to determine whether the marked difference in virulence phenotypes of 4559 and

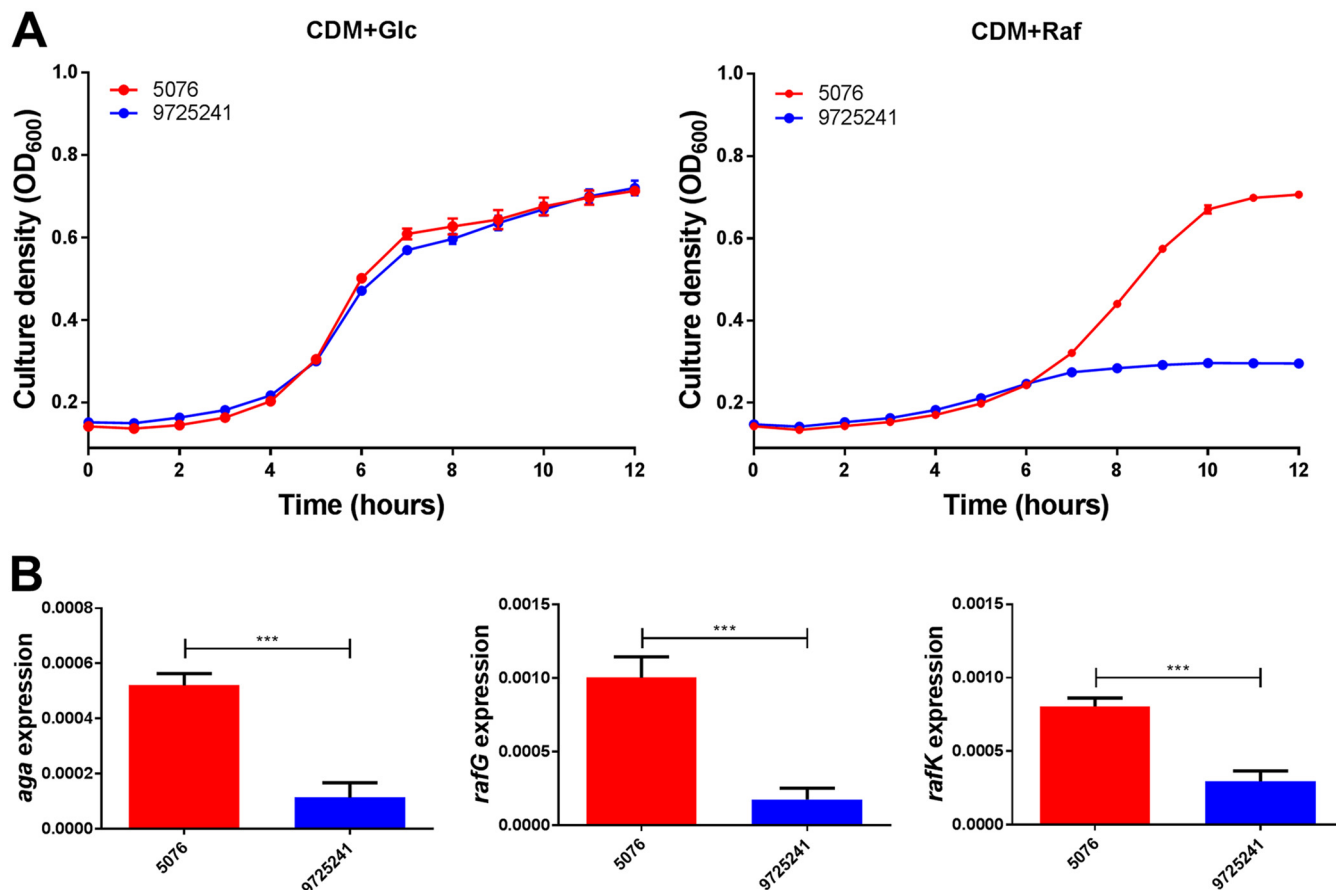


FIG 4 Growth phenotype and raffinose pathway gene expression in serotype 23F ST81 blood and ear isolates. (A) Growth of blood isolate 5076 and ear isolate 9725241 in CDM+Glc or CDM+Raf was monitored by OD₆₀₀ for 12 h. Data are mean OD₆₀₀ ± SD from triplicate assays. (B) The indicated strains were grown in CDM+Glc to an OD₆₀₀ of 0.2, washed and resuspended in CDM+Raf, and then incubated at 37°C for a further 30 min. RNA was then extracted, and levels of *aga*, *rafG*, and *rafK* mRNA were analyzed by qRT-PCR using 16S rRNA as an internal control. The data presented are the means ± SD from three independent experiments. ***, $P < 0.001$ by unpaired t test.

947 is also directly attributable to the SNP in *rafR*, 4559, 947, 4559^{947rafR}, and 947^{4559rafR} were tested in a murine intranasal challenge model. Groups of Swiss mice were challenged with 10⁸ CFU of each strain, and bacterial loads were quantitated in various tissues 24 h postchallenge (Fig. 6). No significant differences in bacterial numbers in the nasopharynx were seen between any groups (Fig. 6), and no bacteria were detected in the blood of any mice (data not presented). However, 4559 was better able than 947 to persist in the lungs of infected mice, with significantly higher geometric mean (GM) bacterial load ($P < 0.0001$) and a significantly greater proportion of infected animals (14/16 versus 6/16; $P < 0.01$) (Fig. 6). On the other hand, bacterial loads of 947 in the ear were significantly greater than that for mice challenged with 4559 ($P < 0.01$), and the proportion of infected mice was also significantly greater (16/16 versus 7/16; $P < 0.001$) (Fig. 6). A similar trend was also seen in the brain (Fig. 6), in accordance with our previous report (4).

Exchanging the *rafR* alleles has a striking impact on virulence phenotype. In the lungs, both the GM CFU and proportion of infected mice for the group challenged with 4559^{947rafR} were significantly lower than those for the 4559 group ($P < 0.0001$ and $P < 0.001$, respectively). Indeed, the virulence phenotype of 4559^{947rafR} was indistinguishable from that of 947. Conversely, the GM bacterial load and proportion of infected mice for the 947^{4559rafR} group were significantly greater than those for the 947 group ($P < 0.01$ and $P < 0.05$, respectively); there were no significant differences in these parameters between the 947^{4559rafR} and 4559 groups. In the ear, both the GM

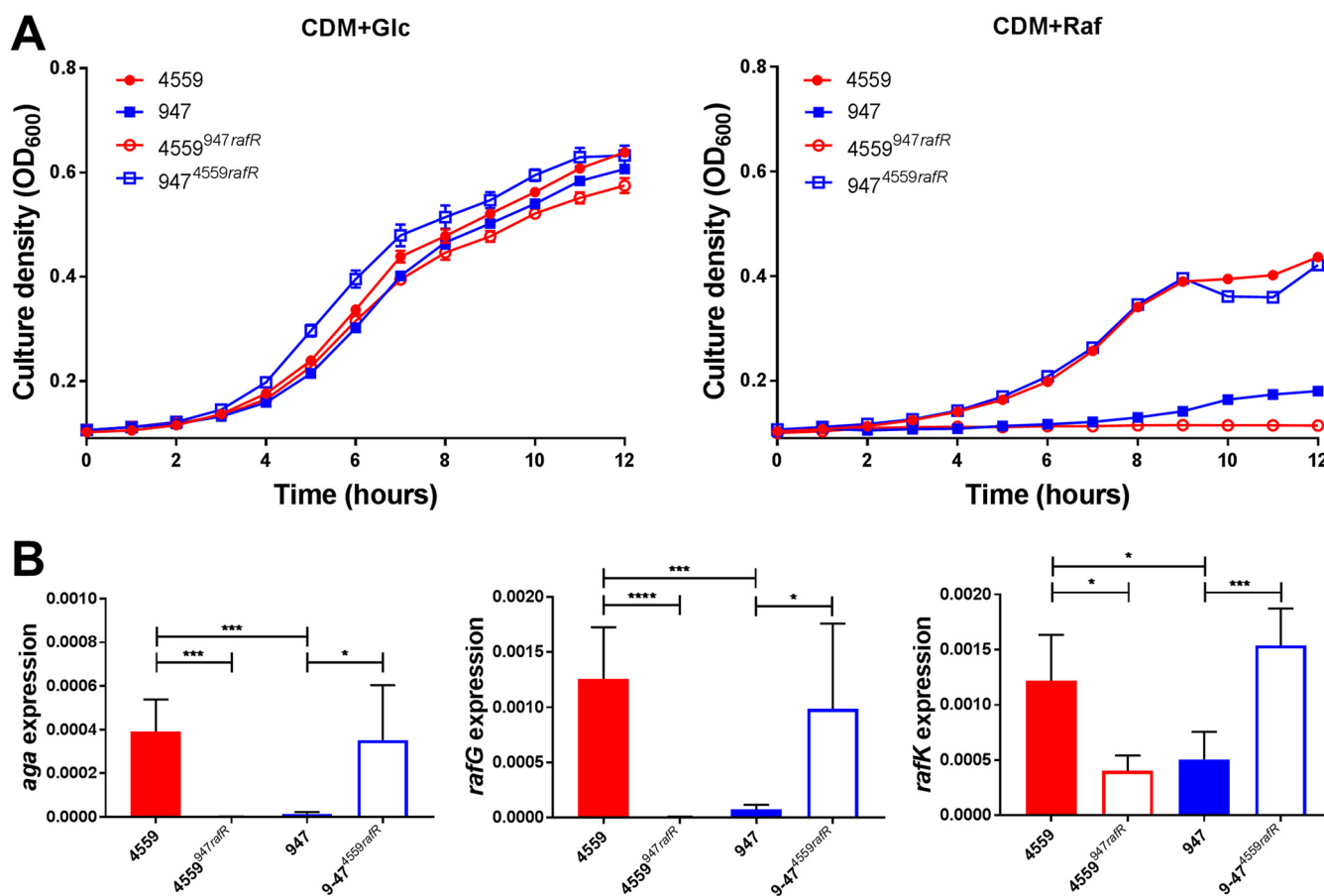


FIG 5 Growth phenotype and raffinose operon gene expression in *rafR* exchange mutants. (A) *S. pneumoniae* strains 4559, 947, 4559^{947rafR}, and 947^{4559rafR} were grown in CDM+Glc or CDM+Raf, and OD₆₀₀ was monitored for 12 h. Data are mean OD₆₀₀ ± SD from triplicate assays. (B) The indicated strains were grown in CDM+Glc to an OD₆₀₀ of 0.2, washed and resuspended in CDM+Raf, and then incubated at 37°C for a further 30 min. RNA was then extracted, and levels of *aga*, *rafG*, and *rafK* mRNA were analyzed by qRT-PCR. Data are the means ± SD from three independent experiments. *, $P < 0.05$, ***, $P < 0.001$, and ****, $P < 0.0001$, by unpaired *t* test.

CFU and proportion of infected mice for the group challenged with 4559^{947rafR} were significantly greater than those for the 4559 group ($P < 0.05$ and $P < 0.01$, respectively). Conversely, both the GM CFU and proportion of infected mice for the group challenged with 947^{4559rafR} were significantly lower than those for the 947 group ($P < 0.01$ in both cases). Moreover, there was no significant difference in either GM bacterial loads or proportions of infected mice between the 4559^{947rafR} and 947 groups or between the 947^{4559rafR} and 4559 groups (Fig. 6). A similar pattern is seen in the brain; the GM CFU for the 4559^{947rafR} group was significantly greater than those for either the 4559 or 947^{4559rafR} groups ($P < 0.01$ in both cases). Moreover, there was no significant difference in either GM bacterial loads or proportions of infected mice between the 4559^{947rafR} and 947 groups or between the 947^{4559rafR} and 4559 group (Fig. 6). Collectively, these data show that swapping the *rafR* allele between 4559 and 947 leads to a switch in their respective virulence profiles, and thus, the D49G SNP in *rafR* is entirely responsible for the observed difference in tissue tropisms between the serotype 14 ST15 blood and ear isolates.

Mutagenesis of *rafK* in serotype 3 ST180 blood and ear isolates. Attempts to construct *rafK* exchange mutants of serotype 3 ST180 blood and ear isolates (180/15 and 180/2, respectively) analogous to the *rafR* exchange mutants constructed for the serotype 14 strains were not successful. Thus, the impact of the SNP in *RafK* could not be directly tested. However, we were able to delete the native *rafK* genes from both type 3 strains (designated 180/15 Δ *rafK* and 180/2 Δ *rafK*, respectively). Both mutants

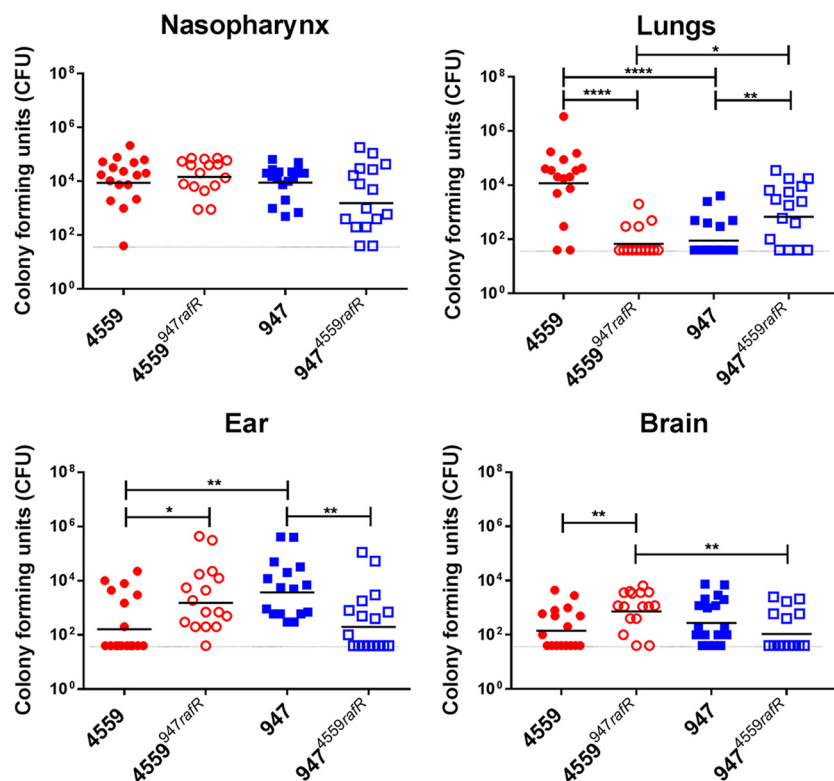


FIG 6 Virulence phenotype of *rafR* exchange mutants. Groups of 16 mice were infected intranasally with 10^8 CFU of the indicated strain. At 24 h, all mice from each group were euthanized and numbers of pneumococci in the indicated tissues/sites were quantitated (see Materials and Methods). Viable counts (total CFU per tissue) are shown for each mouse at each site; horizontal bars indicate the geometric mean (GM) CFU for each group; the broken line indicates the threshold for detection. Differences in GM bacterial loads between groups are indicated by asterisks: *, $P < 0.05$, **, $P < 0.01$, and ****, $P < 0.0001$, by unpaired t test.

were incapable of growth in CDM+Raf, and expression of *aga* and *rafG* was virtually undetectable by qRT-PCR; *rafK* expression was also undetectable, as expected (Fig. 7). Thus, the ear isolate 180/2 exhibits a phenotype that is intermediate between that of the blood isolate 180/15 and either of the two $\Delta rafK$ mutants, consistent with partial functionality of the ear isolate RafK. *S. pneumoniae rafK* deletion mutants have previously been shown to be outcompeted by the wild type in the murine lung and nasopharynx (10, 12). Similarly, in the present study, bacterial loads in the lungs, blood, ear, and brain were also lower for mice challenged with the 180/2 and 180/15 $\Delta rafK$ mutants relative to those challenged with the respective wild types at 48 h after intranasal challenge (result not presented). This indicates that even the intermediate level of raffinose pathway gene expression exhibited by ear isolate 180/2 contributes to virulence.

DISCUSSION

Pneumococci are strictly fermentative bacteria, relying solely on carbohydrate metabolism for energy and growth (13). However, carbohydrate availability differs between host niches, and so the ability to respond to and utilize distinct carbohydrates is crucial for pneumococcal fitness *in vivo*. The *S. pneumoniae* genome encodes 21 phosphotransferase systems (PTSs) and up to 8 ATP binding cassette (ABC) transporters for the import of carbohydrates (9, 14), accounting for roughly 30% of all transport systems. Previous studies have shown that several of these carbohydrate transporters, present in both the core and accessory genome, impact pneumococcal virulence. For example, a sucrose PTS and ABC transporter system of serotype 4 pneumococci have been shown to play roles in murine colonization and pneumonia, respectively (15),

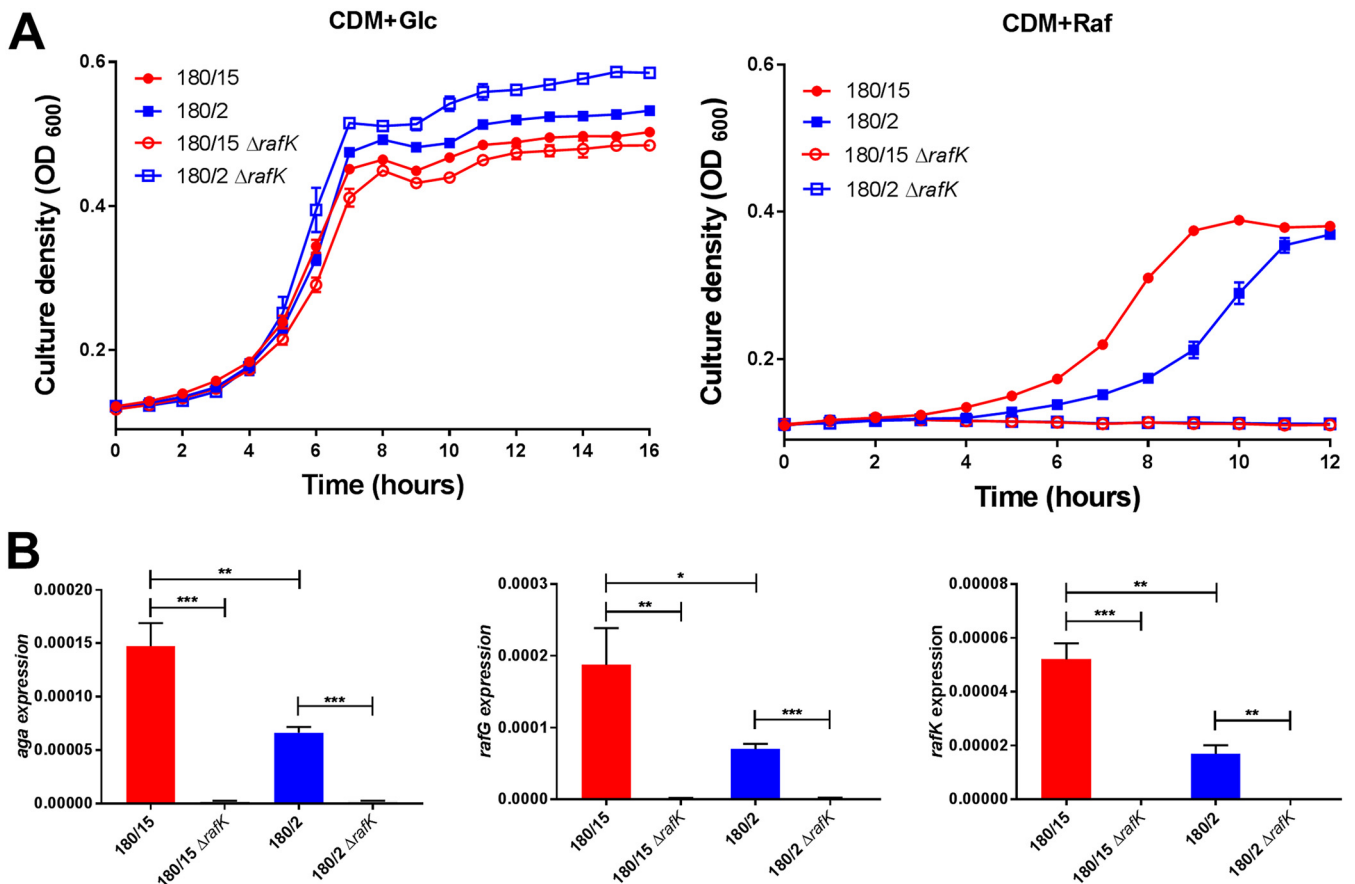


FIG 7 Growth phenotype and raffinose operon gene expression in $\Delta rafK$ mutants. (A) *S. pneumoniae* serotype 3 strains 180/15, 180/2, 180/15 $\Delta rafK$, and 180/2 $\Delta rafK$ were grown in CDM+Glc or CDM+Raf, and OD₆₀₀ was monitored for 12 h. Data are mean OD₆₀₀ \pm SD from triplicate assays. (B) The indicated strains were grown in CDM+Glc to an OD₆₀₀ of 0.2, washed and resuspended in CDM+Raf, and then incubated at 37°C for a further 30 min. RNA was then extracted, and levels of *aga*, *rafG*, and *rafK* mRNA were analyzed by qRT-PCR. Data are the means \pm SD from three independent experiments. *, $P < 0.05$, **, $P < 0.01$, and ***, $P < 0.001$, by unpaired *t* test.

while transporters for carbohydrates such as glucose, galactose, and mannose were shown to impact invasive pneumococcal disease (16–18).

The present study further underscores the critical role played by differential carbohydrate metabolism in pneumococcal pathogenesis. It demonstrates that reduced capacity to utilize raffinose does not simply reduce pneumococcal virulence, but rather changes the nature of disease caused. In multiple serotypes/ST lineages, ear isolates had defective growth in CDM+Raf and reduced expression of raffinose pathway genes relative to their serotype/ST-matched blood isolates. Exchange of *rafR* alleles between ear and blood isolates of serotype 14 ST15 reversed these *in vitro* phenotypes. Moreover, *rafR* exchange caused blood isolates to now cause otitis media and meningitis rather than pneumonia, after intranasal challenge, and conversely cause ear isolates to now target the lungs. This striking switch in *in vitro* and *in vivo* behaviors was attributable to a single, nonconservative SNP (D249G) in RafR, identifying this residue as one of critical functional importance. Significantly, the region of RafR from amino acids 226 to 268 comprises a conserved signature sequence for the AraC/XylS family of transcriptional regulators (11).

Interestingly, in spite of exhibiting similarly distinct *in vitro* and *in vivo* phenotypes, the serotype 3 ST180 blood and ear isolates did not share the SNP in *rafR*, but rather had an SNP in *rafK*, which encodes the ATPase required for raffinose uptake via the ABC transport system encoded by *rafEFG*. RafK-mediated uptake of raffinose has previously been shown to be essential for induction of the *raf* operons in *S. pneumoniae* D39 (10). Attempts to construct *rafK* exchange mutants in this lineage (analogous to the serotype

14 ST15 *rafR* exchange mutants) were not successful. However, *rafK* deletion mutants of both 180/2 and 180/15 were obtained. Whereas the wild-type ear isolate 180/2 exhibited reduced growth in CDM+Raf and expression of *aga*, *rafG*, and *rafK* relative to the wild-type blood isolate 180/15, both *rafK* deletion mutants were unable to grow in CDM+Raf at all, and expression of any of the *raf* operon transcripts was undetectable. Clearly, the RafK allele carried by 180/2 retains partial function. The I227T SNP that distinguishes the RafK alleles of 180/2 and 180/15 is located in the conserved regulatory domain motif 1. This domain is believed to be involved in the interaction between RafK and the enzyme dihydrolipoamide dehydrogenase (DLDH), which has been shown to modulate raffinose uptake and *raf* operon expression in *S. pneumoniae* D39 (10). In the murine model, both *rafK* deletion mutants exhibited reduced bacterial loads in multiple host niches relative to their respective wild-type strains, consistent with previous reports (10, 12).

Our findings provide an interesting example of convergent evolution, whereby pneumococci belonging to two unrelated serotypes/lineages exhibit SNPs in separate genes, each affecting raffinose uptake and utilization, which in turn correlate with distinct pathogenic profiles in both mice and humans (the latter by inference from the clinical isolation site). In *S. pneumoniae* D39, induction of expression of the *raf* operon gene *aga* required the presence of raffinose; reduced but nevertheless significant *aga* expression also occurred in a *rafR* knockout mutant (11). Thus, *raf* operon expression in pneumococci can be impacted either by defects in raffinose import (e.g., due to a defective RafK), such that insufficient exogenous raffinose (if present) is internalized to induce *raf* expression, or by functional defects in the transcriptional activator RafR, such that baseline levels of expression induced by the presence of raffinose are not further upregulated. The *raf* operons are part of the core genome of *S. pneumoniae*, and BLASTX analysis of available genomes shows that there is between 1% and 3% deduced amino acid sequence variation within any of the *raf* genes. Thus, SNPs are widespread, but it is not known which (if any) of these other SNPs impact the capacity to import or utilize raffinose or the virulence phenotype.

Notwithstanding the results presented above, the precise mechanism whereby differential raffinose uptake/utilization determines the virulence phenotype is uncertain. Raffinose is a plant-derived trisaccharide present in many staple foods, particularly beans and soy (19, 20). Although humans are unable to metabolize it, dietary raffinose is known to be absorbed by the intestinal epithelium (21), raising the possibility of at least small amounts being present on mucosal surfaces. As part of the present study, we confirmed that expression of *aga*, *rafG*, and *rafK* was not detectable by qRT-PCR when pneumococci are grown *in vitro* in media lacking raffinose. However, expression of all three genes was detected in RNA extracts of mouse lung tissue 6 h after intranasal challenge with either of the serotype 14 ST15 blood or ear isolates (results not shown). Since raffinose is the only known inducer of the *raf* operon in *S. pneumoniae*, this finding is strongly indicative of the presence of bioavailable raffinose in the murine lung. A potential complicating factor is that RafEFG is reported to be also capable of importing stachyose (14), while RafK has been reported to also energize uptake of sialic acid and maltotetraose via unrelated transporters (12). Thus, the SNPs observed in the present study could have pleiotropic effects. However, no differences in metabolism of these sugars between the serotype 14 ST15 blood and ear isolates were observed using phenotypic microarray analysis, and the serotype 3 ST180 strains were unable to grow in CDM with stachyose or sialic acid as the sole carbon source. Moreover, there was no significant difference in the growth rates of the ST180 blood and ear isolates when grown in CDM with maltotetraose (results not presented). A particularly intriguing finding of the present study was that lower raffinose uptake/utilization by the ear isolates provided an advantage over blood isolates in the ear compartment. Interestingly, exogenous raffinose has recently been shown to promote biofilm formation by *Streptococcus mutans* by promoting aggregation of extracellular DNA into the biofilm matrix. Biofilm formation was unaffected by deletion of the α -galactosidase gene *agaL*, indicating that the effect was unrelated to metabolism of any internalized raffinose (22). Thus, it is conceivable that the reduced

TABLE 3 *S. pneumoniae* strains used in this study

Strain	Description	Source	Reference
4559	Serotype 14 ST15	Blood	7
947	Serotype 14 ST15	Ear	7
4534	Serotype 14 ST15	Blood	7
51742	Serotype 14 ST15	Ear	7
4559 ^{947rafR}	4559 expressing 947 <i>rafR</i> gene		This study
947 ^{4559rafR}	947 expressing 4559 <i>rafR</i> gene		This study
180/15	Serotype 3 ST180	Blood	6
180/2	Serotype 3 ST180	Ear	6
5076	Serotype 23F ST81	Blood	This study
9725241	Serotype 23F ST81	Ear	This study

capacity of *S. pneumoniae* ear isolates to assimilate (and thereby deplete) raffinose from the middle ear mucosa may similarly promote pneumococcal biofilm formation in that niche, leading to otitis media. Further studies are in progress in our laboratory to elucidate the precise molecular mechanism whereby fine-tuning of levels of expression of raffinose uptake and utilization genes can have such a profound impact on pathogenic profiles of clinical isolates of *S. pneumoniae*.

MATERIALS AND METHODS

Bacterial strains and growth conditions. The *S. pneumoniae* strains used in this study are listed in Table 3. Cells were routinely grown in casein-based, semisynthetic liquid medium (C+Y) (23) or serum broth (SB) as required. Growth assays were performed using a chemically medium (CDM) comprising RPMI 1640 medium (Sigma), supplemented with amino acids, vitamins, choline, and catalase as described previously (24), with either 0.5% glucose or 0.5% raffinose. Bacteria were plated on Columbia agar supplemented with 5% (vol/vol) horse blood (BA) with or without gentamicin (40 µg/ml), kanamycin (500 µg/ml), or streptomycin (150 µg/ml) (as required) and incubated at 37°C in 5% CO₂ overnight. For gene expression analyses, strains were grown in CDM+Glc medium to an OD₆₀₀ of 0.2, before being incubated in CDM+Raf for 30 min.

Genome sequencing. *S. pneumoniae* strains were grown to mid-exponential phase in Todd-Hewitt broth supplemented with 1% yeast extract. Genomic DNA (gDNA) was extracted using the Qiagen genomic DNA buffer set with 100/g Genomic Tips according to the manufacturer's instructions, except mutanolysin (20 U) and sodium deoxycholate (0.1%) were included to aid cell lysis. The gDNA was sequenced at the Ramaciotti Centre for Genomics (University of New South Wales, Sydney, Australia) on an Illumina MiSeq (250-bp paired-end reads), as well as a PacBio RSII instrument using one SMRT cell per strain, a 20-kb insert library, and the P6 polymerase and C4 sequencing chemistry.

Bioinformatic analyses. The Artemis Comparison Tool was used to compare genomes (25). MiSeq reads of 4559 and 947 and 180/15 and 180/2 were mapped to the assembled reference genome of the opposing strain with BOWTIE2 version 2.2.6 (26). Variant calling was then performed using SAMTools version 0.1.18 (27), and variants were mapped to coding sequences of the reference strain using BEDTools version 2.25.0 (28). Single nucleotide polymorphisms (SNPs) and insertions/deletions (indels) were filtered for those with scores of 100 or greater. Artemis was used to visualize SNPs and indels (29). Sanger sequencing was performed to confirm the SNPs in *rafR* and *rafK* (Australian Genome Research Facility, Adelaide).

Phenotypic microarrays. Carbon phenotype microarray analysis was performed on the serotype 14 ST 15 strains, using the PM microplates PM1 and PM2A (Biolog, Inc.), which tested for the catabolism of 190 different carbon sources. Each well of the microarrays contained a different carbon source. Briefly, cells were suspended in the provided buffer (as per the manufacturer's instructions) to an A₅₉₀ of 0.37. One hundred microliters of this suspension was added to the wells, and the A₅₉₀ was measured after 17 h of incubation at 37°C. Catabolism was measured through the reduction of a colorless tetrazolium dye by NADH, produced during catabolic activity. Absorbance values above 0.65 after subtraction of that for the zero carbon source blank were considered positive.

Growth assays. Each tested strain was grown in CDM supplemented with either 0.5% Glc (CDM+Glc), 0.5% Raf (CDM+Raf), or no sugar (CDM) and then incubated at 37°C for 12 h in 96-well flat bottom plates (Costar). The OD₆₀₀ was measured every 15 min using a SPECTRAMax M2 spectrophotometer (Millennium Science). All experiments were conducted in triplicate and repeated at least two times.

qRT-PCR. Differences in levels of gene expression were assayed by one-step relative quantitative real-time RT-PCR (qRT-PCR) in a Roche LC480 real-time cyclor essentially as described previously (30). The specific primers used for the various genes are listed in Table 4 and were used at a final concentration of 200 nM per reaction. As an internal control, primers specific for 16S rRNA were employed. Amplification data were analyzed using the comparative critical threshold cycle (2^{-ΔΔCT}) method (31).

Mutagenesis. The *rafR* gene swap between serotype 14 ST15 4559 and 947 strains, to produce 4559^{947rafR} and 947^{4559rafR}, was achieved via allelic exchange mutagenesis utilizing the Janus cassette, as described previously (32, 33). This involved a three-step process in which endogenous *rpsL* (which confers

TABLE 4 Oligonucleotide primers used in this study

Primer	Sequence (5'→3')	Reference
<i>rafR</i> Flank F	GCGAACGTAGGTTACAATCGT	This study
<i>rafR</i> R j tail	GGAAAGGGGCCAGGTCTCTAGCATGTGCTACCTCCTACC	This study
<i>rafR</i> F j tail	CATTATCCATAAAAATCAAAGGGGAAATCCTACCAAGCTGTCTACC	This study
<i>rafR</i> Flank R	CGAACGTAGTTCAGTGGTAGAA	This study
Janus F	CCGTTTGATTTTAAATGGATAATG	33
Janus R	AGAGACCTGGGCCCTTTCC	33
<i>aga</i> F	AAGGTCAGAATGGTCCACAG	This study
<i>aga</i> R	GCTGGAAAAATCAGCCATAAA	This study
<i>rafG</i> F	CCTATGGCAGCCTACTCCATC	This study
<i>rafG</i> R	GGGTCTGTGGAATCGCATAGG	This study
<i>rafK</i> F	AACGACGTAGCTCCAAAAGA	This study
<i>rafK</i> R	GCTGGTTTACGTTCCAAGAA	This study
16s rRNA F	GGTGAGTAACGCGTAGGTAA	34
16s rRNA R	ACGATCCGAAAACCTTCTTC	34
<i>rafR</i> sanger	AGTAGAAGAGCTGGTGTGTTG	This study
<i>rafR</i> sanger	TCTGTGACTAAGCCAGTTTC	This study
<i>rafK</i> Flank F	AGGACTTGGTTCTTGTGTGAG	This study
<i>rafK</i> R ery tail	TTGTTTATGTAATCACTCCTTCTACCATGAGGTGAACTCC	This study
<i>rafK</i> F ery tail	CGGGAGGAAATAATTCTATGAGATCAGTTAATCTAGGGAGAG	This study
<i>rafK</i> Flank R	CTCAAAGGCAACTGGACAAC	This study

streptomycin sensitivity) was first replaced with the streptomycin-resistant *rpsL1* allele by direct transformation of the blood and ear isolates. The Janus cassette (comprising a kanamycin resistance marker and a dominant counterselectable *rpsL⁺* marker) was then used to replace the native *rafR* gene by direct transformation with a linear PCR product comprising the Janus cassette flanked by sequences 5' and 3' to *rafR* (selecting on kanamycin). In the final step, the Janus cassette in Kan^r/Strep^s transformants is replaced by transformation with the alternative *rafR* allele and flanking sequences, counterselecting on streptomycin (loss of the Janus cassette reinstates the Strep^r phenotype). Gene swap constructs were confirmed by Sanger DNA sequencing (AGRF, Adelaide). The *rafK* gene was also deleted from serotype 3 ST180/2 and ST180/15 by direct transformation with a linear DNA fragment comprising an erythromycin resistance cassette flanked by sequences 5' and 3' to *rafK* generated by overlap PCR, essentially as previously described (34). The primers used are listed in Table 4. Mutant constructs were confirmed by PCR.

Animal studies. Animal experiments were approved by the University of Adelaide Animal Ethics Committee. Groups of outbred 6-week-old female Swiss (CD-1) mice were anesthetized by intraperitoneal injection of pentobarbital sodium (Nembutal; Rhone-Merieux) and challenged intranasally (i.n.) with 50 μ l of bacterial suspension containing approximately 1×10^8 CFU in SB (7). The challenge dose was confirmed retrospectively by serial dilution and plating on BA. Mice were euthanized by CO₂ asphyxiation at 24 h, and then tissue samples (lungs, nasopharynx, brain, ear, and blood) were harvested and pneumococci enumerated in tissue homogenates as described previously via serial dilution and plating on plates containing BA plus gentamicin (35).

Data availability. Genome sequences have been deposited with ENA under accession no. SAMEA5092021, SAMEA5092022, SAMEA5092023, and SAMEA5092024, for strains 947, 4559, 180/2, and 180/15, respectively.

ACKNOWLEDGMENTS

We acknowledge the contribution of the Antibiotic Resistant Sepsis Pathogens Framework Initiative consortium (<https://data.bioplatforms.com/organization/pages/bpa-sepsis/consortium>) in the generation of genome sequence data used in this publication. The Initiative is supported by funding from Bioplatforms Australia through the Australian Government National Collaborative Research Infrastructure Strategy (NCRIS). We also thank Kimberley McLean for assistance with animal experiments.

This work was supported by National Health and Medical Research Council (NHMRC) Program Grant 1071659 and NHMRC Senior Principal Research Fellowship 1043070 to J.C.P. as well as Australian Research Council DECRA Fellowship DE140100963 and a University of Adelaide Beacon Fellowship to C.T. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

REFERENCES

- Kadioglu A, Weiser JN, Paton JC, Andrew PW. 2008. The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. *Nat Rev Microbiol* 6:288–301. <https://doi.org/10.1038/nrmicro1871>.
- Weiser JN, Ferreira DM, Paton JC. 2018. *Streptococcus pneumoniae*: transmission, colonization and invasion. *Nat Rev Microbiol* 16:355–367. <https://doi.org/10.1038/s41579-018-0001-8>.

3. Enright MC, Spratt BG. 1998. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. *Microbiology* 144:3049–3060. <https://doi.org/10.1099/00221287-144-11-3049>.
4. McAllister LJ, Ogunniyi AD, Stroehrer UH, Leach AJ, Paton JC. 2011. Contribution of serotype and genetic background to virulence of serotype 3 and serogroup 11 pneumococcal isolates. *Infect Immun* 79:4839–4849. <https://doi.org/10.1128/IAI.05663-11>.
5. Kelly T, Dillard JP, Yother J. 1994. Effect of genetic switching of capsular type on virulence of *Streptococcus pneumoniae*. *Infect Immun* 62:1813–1819.
6. Trappetti C, van der Maten E, Amin Z, Potter AJ, Chen AY, van Mourik PM, Lawrence AJ, Paton AW, Paton JC. 2013. Site of isolation determines biofilm formation and virulence phenotypes of *Streptococcus pneumoniae* serotype 3 clinical isolates. *Infect Immun* 81:505–513. <https://doi.org/10.1128/IAI.01033-12>.
7. Amin Z, Harvey RM, Wang H, Hughes CE, Paton AW, Paton JC, Trappetti C. 2015. Isolation site influences virulence phenotype of serotype 14 *Streptococcus pneumoniae* strains belonging to multilocus sequence type 15. *Infect Immun* 83:4781–4790. <https://doi.org/10.1128/IAI.01081-15>.
8. Manco S, Hernon F, Yesilkaya H, Paton JC, Andrew PW, Kadioglu A. 2006. Pneumococcal neuraminidases A and B both have essential roles during infection of the respiratory tract and sepsis. *Infect Immun* 74:4014–4020. <https://doi.org/10.1128/IAI.01237-05>.
9. Buckwalter CM, King SJ. 2012. Pneumococcal carbohydrate transport: food for thought. *Trends Microbiol* 20:517–522. <https://doi.org/10.1016/j.tim.2012.08.008>.
10. Tyx RE, Roche-Hakansson H, Hakansson AP. 2011. Role of dihydrolipoamide dehydrogenase in regulation of raffinose transport in *Streptococcus pneumoniae*. *J Bacteriol* 193:3512–3524. <https://doi.org/10.1128/JB.01410-10>.
11. Rosenow C, Maniar M, Trias J. 1999. Regulation of the alpha-galactosidase activity in *Streptococcus pneumoniae*: characterization of the raffinose utilization system. *Genome Res* 9:1189–1197. <https://doi.org/10.1101/gr.9.12.1189>.
12. Marion C, Aten AE, Woodiga SA, King SJ. 2011. Identification of an ATPase, MsmK, which energizes multiple carbohydrate ABC transporters in *Streptococcus pneumoniae*. *Infect Immun* 79:4193–4200. <https://doi.org/10.1128/IAI.05290-11>.
13. Paixão L, Caldas J, Kloosterman TG, Kuipers OP, Vinga S, Neves AR. 2015. Transcriptional and metabolic effects of glucose on *Streptococcus pneumoniae* sugar metabolism. *Front Microbiol* 6:1041. <https://doi.org/10.3389/fmicb.2015.01041>.
14. Bidossi A, Mulas L, Decorosi F, Colomba L, Ricci S, Pozzi G, Deutscher J, Viti C, Oggioni MR. 2012. A functional genomics approach to establish the complement of carbohydrate transporters in *Streptococcus pneumoniae*. *PLoS One* 7:e33320. <https://doi.org/10.1371/journal.pone.0033320>.
15. Iyer R, Camilli A. 2007. Sucrose metabolism contributes to in vivo fitness of *Streptococcus pneumoniae*. *Mol Microbiol* 66:1–13. <https://doi.org/10.1111/j.1365-2958.2007.05878.x>.
16. Hava DL, Camilli A. 2002. Large-scale identification of serotype 4 *Streptococcus pneumoniae* virulence factors. *Mol Microbiol* 45:1389–1406.
17. Orihuela CJ, Radin JN, Sublett JE, Gao G, Kaushal D, Tuomanen EI. 2004. Microarray analysis of pneumococcal gene expression during invasive disease. *Infect Immun* 72:5582–5596. <https://doi.org/10.1128/IAI.72.10.5582-5596.2004>.
18. Ogunniyi AD, Mahdi LK, Trappetti C, Verhoeven N, Merkmans D, Van der Hoek MB, Plumptre CD, Paton JC. 2012. Identification of genes that contribute to the pathogenesis of invasive pneumococcal disease by in vivo transcriptomic analysis. *Infect Immun* 80:3268–3278. <https://doi.org/10.1128/IAI.00295-12>.
19. Díaz-Batalla L, Widholm JM, Fahey GC, Jr, Castaño-Tostado E, Paredes-López O. 2006. Chemical components with health implications in wild and cultivated Mexican common bean seeds (*Phaseolus vulgaris* L.). *J Agric Food Chem* 54:2045–2052. <https://doi.org/10.1021/jf051706l>.
20. Kumar V, Rani A, Goyal L, Dixit AK, Manjaya JG, Dev J, Swamy M. 2010. Sucrose and raffinose family oligosaccharides (RFOs) in soybean seeds as influenced by genotype and growing location. *J Agric Food Chem* 58:5081–5085. <https://doi.org/10.1021/jf903141s>.
21. Lobley RW, Burrows PC, Warwick R, Dawson DJ, Holmes R. 1990. Simultaneous assessment of intestinal permeability and lactose tolerance with orally administered raffinose, lactose and L-arabinose. *Clin Sci (Lond)* 79:175–183. <https://doi.org/10.1042/cs0790175>.
22. Nagasawa R, Sato T, Senpuku H. 2017. Raffinose induces biofilm formation by *Streptococcus mutans* in low concentrations of sucrose by increasing production of extracellular DNA and fructan. *Appl Environ Microbiol* 83:e00869-17. <https://doi.org/10.1128/AEM.00869-17>.
23. Lacks S, Hotchkiss RD. 1960. A study of the genetic material determining an enzyme in *Pneumococcus*. *Biochim Biophys Acta* 39:508–518. [https://doi.org/10.1016/0006-3002\(60\)90205-5](https://doi.org/10.1016/0006-3002(60)90205-5).
24. Kloosterman TG, Bijlsma JJ, Kok J, Kuipers OP. 2006. To have neighbour's fare: extending the molecular toolbox for *Streptococcus pneumoniae*. *Microbiology* 152:351–359. <https://doi.org/10.1099/mic.0.28521-0>.
25. Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J. 2005. ACT: the Artemis Comparison Tool. *Bioinformatics* 21:3422–3423. <https://doi.org/10.1093/bioinformatics/bti553>.
26. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
27. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
28. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
29. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* 16:944–945. <https://doi.org/10.1093/bioinformatics/16.10.944>.
30. Mahdi LK, Ogunniyi AD, LeMessurier KS, Paton JC. 2008. Pneumococcal virulence gene expression and host cytokine profiles during pathogenesis of invasive disease. *Infect Immun* 76:646–657. <https://doi.org/10.1128/IAI.01161-07>.
31. Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-ΔΔCT} method. *Methods* 25:402–408. <https://doi.org/10.1006/meth.2001.1262>.
32. Harvey RM, Hughes CE, Paton AW, Trappetti C, Tweten RK, Paton JC. 2014. The impact of pneumolysin on the macrophage response to *Streptococcus pneumoniae* is strain-dependent. *PLoS One* 9:e103625. <https://doi.org/10.1371/journal.pone.0103625>.
33. Sung CK, Li H, Claverys JP, Morrison DA. 2001. An *rpsL* cassette, Janus, for gene replacement through negative selection in *Streptococcus pneumoniae*. *Appl Environ Microbiol* 67:5190–5196. <https://doi.org/10.1128/AEM.67.11.5190-5196.2001>.
34. Trappetti C, McAllister LJ, Chen A, Wang H, Paton AW, Oggioni MR, McDevitt CA, Paton JC. 2017. Autoinducer 2 signaling via the phosphotransferase FruA drives galactose utilization by *Streptococcus pneumoniae*, resulting in hypervirulence. *mBio* 8:e02269-16. <https://doi.org/10.1128/mBio.02269-16>.
35. Trappetti C, Ogunniyi AD, Oggioni MR, Paton JC. 2011. Extracellular matrix formation enhances the ability of *Streptococcus pneumoniae* to cause invasive disease. *PLoS One* 6:e19844. <https://doi.org/10.1371/journal.pone.0019844>.