Identifying genetic mutations which aid *Streptococcus pneumoniae* in intra-host transmission from the lung to the blood

Isobel Maclean[1], Daniel Neill[2] and Angharad Green[2]

[1] Advanced Biological Sciences MSc (Microbiology), School of Life Sciences, Liverpool, UK, L69 7ZB; [2] Department of Clinical Infection, Microbiology and Immunology, University of Liverpool, Liverpool, L69 7BE.

***Streptococcus pneumoniae* (pneumococcus) is a bacterial species normally found living in the throat of humans. The bacteria normally cause no harm, however they can spread through the body to the lungs and blood. Here, the bacteria can cause the development of diseases such as pneumonia and septicaemia (sepsis) respectively. The processes by which the pneumococcus spreads and causes these diseases are not understood. This study aims to identify genetic adaptations that are acquired by the bacteria during infection of the lungs and blood. Mice were infected with *S. pneumoniae*, which resulted in the development of pneumonia, and in some cases, sepsis. Bacterial samples were taken and analysed to reveal genetic mutations present within the population, and their frequency. The study found three different genes that are potentially beneficial to the pneumococcus during infection of the lungs and blood. The functions of these genes were researched and are related to bacterial metabolism. In conclusion, the proteins that these genes encode for may be potential targets for future treatments or vaccines, as they enable the bacteria to cause disease.**

Abstract

Streptococcus pneumoniae or pneumococcus is a Gram-positive bacterium which colonises the human nasopharynx, mostly asymptomatically. The pneumococcus possesses the ability to infect diverse host niches within the human body, which can result in the development of invasive diseases such as pneumonia, septicaemia, and meningitis. The incidence of fatality for these diseases is as high as 11-30% within the U.S and Europe. Experimental evolution was performed via the passage of 10 independent lineages, using a mouse pneumonia model, to identify acquisition of genetic mutations in lung and blood populations of pneumococcus. Samples of the pneumococcal population in the lung and the blood were taken at passage numbers 1, 5, 10, 15 and 20. Bioinformatic analyses show the genetic variants present within the blood and lung isolates, and these variants were compared. The results show evidence of niche specific adaptation to the blood environment. Genetic variants in *ilvD*, *pyk* and *lctO*, were present within multiple independent lineages within the blood isolates showing evidence of parallel evolution. The gene products of *pyk* and *lctO* play roles in pyruvate metabolism. *IlvD* is involved in amino acid biosynthesis and iron acquisition. Iron acquisition is a key challenge for bacterial pathogens when colonising the blood, as most iron is bound to haemoglobin. Genes *ilvD* and *pyk* were shown to have significantly higher gene expression within *S. pneumoniae* D39 when bacteria were grown in blood mimicking conditions as compared to lung mimicking conditions. In conclusion these genetic variants may be potential targets for new treatments or vaccines.

Introduction

Streptococcus pneumoniae (pneumococcus) is a Gram-positive bacterium which acts as an opportunistic pathogen within the human host. Pneumococcus causes diseases such as otitis media, and invasive diseases; pneumonia, septicaemia, and meningitis (Aprianto *et al*., 2018).

A common complication of community acquired pneumococcal pneumonia is the development of bacteraemia or septicaemia (sepsis) (Ceccato & Torres, 2018). This can often progress to septic shock, which has a high mortality rate even in treated cases. In studies in Europe and the US, case fatality rates for invasive pneumococcal disease and sepsis are reported to be as high as 11-30% (Askim *et al*., 2016; Chavanet, 2012). The mechanisms that allow pneumococcus to colonise diverse host niches are not well understood, for example the mechanisms by which pneumococcus is able to penetrate the epithelial barrier of the alveoli and enter the blood is still unclear.

The hypothesis of this study is that pneumococcal colonisation of the blood from the lung is associated with a bottleneck of just one cell, this has already been demonstrated in a previous study (Kono *et al*., 2016). Due to reduced diversity within the pneumococcal population, as pneumococcus replicates within the blood, the new environmental pressures may result in the selection of a novel, advantageous genotype. Only the genotype with the most advantageous mutations will survive. As well as the accumulation of new genetic variants, changes in gene expression patterns may also aid in colonisation of the blood. The aim of this study is to identify and determine the function of these genes in which variants are selected during the lung to blood transition. This will provide an insight into the pathogenicity of pneumococcal sepsis, potentially contributing to the development of new treatments.

Methodology

In vivo experimental evolution

Ancestor strain *S. pneumoniae* D39 (serotype 2) was grown overnight from a single colony. The following day, the bacteria were passaged to new growth medium, and permitted to grow to mid-exponential stage. Ten mice were inoculated, establishing 10 distinct pneumococcal lineages. Infection led to the development of pneumonia. After each passage, bacteria were recovered from the lungs. A small number of bacteria were swabbed onto gentamicin BAB agar, and colonies were confirmed as *S. pneumoniae* by an optochin disk. The pneumococcal population was then resuspended in solution and stored in three tubes at -80°C, to be used for analysis of the genome. One was used in re-infection of a new mouse (passage number 2). The passage process was repeated 20 times for each of the 10 lineages as shown in Figure 1. Samples of pneumococcal population from the blood in these infected mice were stored, alongside the samples kept from the lungs.

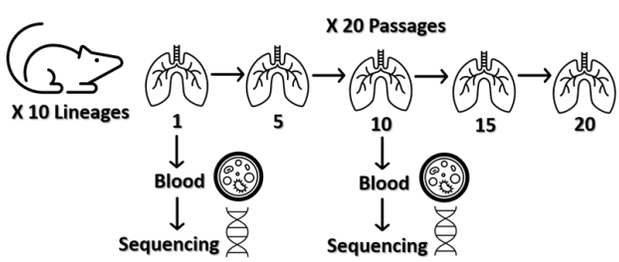

Figure 1. Experimental evolution mouse pneumonia model. Demonstration of how *S. pneumoniae* was passaged a total of 20 times via a mouse pneumonia model, for 10 separate independent lineages. Samples were taken from the lung after passage numbers 1, 5, 10, 15 and 20. If mice had developed a blood infection, a sample of the pneumococcal population from the blood was also stored. All samples were sequenced.

Data Analysis

Genetic variants that arose within the lungs were compared to those that arose within the blood. To begin the comparison, a spreadsheet was created using Microsoft Excel which included genetic variants exclusively present in lung evolved lineages/ blood evolved lineages, and the frequency of the genetic variant within the pneumococcal population.

Genes of interest were compiled into a spreadsheet and were narrowed down to a final list for lung-specific variants, and a separate list for blood-specific variants. The variants on the final list were researched using literature searches, Uniprot (The UniProt, 2018), KEGG (Kanehisa, 2019), PneumoBrowse (Slager *et al*., 2018) and PneumoExpress (Aprianto *et al*., 2018).

Aprianto *et al*., (2018) used *S. pneumoniae* D39 to quantify relative expression levels of the transcriptome in 22 different infection relevant conditions, including lung-mimicking conditions (LMC) and blood-mimicking conditions (BMC). PneumoExpress data was used to perform a one-way ANOVA Sidak's multiple comparison test, selecting LMC and BMC for gene variants *ilvD*, *lctO* and *pyk*. Statistical analyses were performed using GraphPad Prism version is 8.4.3.

Results

Table 1 was constructed from raw data. In particular, variants in *pyk*, *lctO* and *ilvD* are of interest. This is because they are all present in 3 different lineages, suggesting evidence of parallel evolution. The function of these genes, as shown in the table, include pyruvate metabolism and iron acquisition.

Figure 2A shows expression levels (TPM) of the gene "*ilvD*" in *S. pneumoniae* strain D39 in a range of infection relevant conditions, including lung mimicking conditions (LMC), blood mimicking conditions (BMC), nose mimicking conditions (NMC) and throughout the time course of an infection, measured in minutes post infection (mpi). The graphs show that expression of *ilvD* was highest for BMC, and at 30 mpi.

Figure 2B and 2C show the same graph as described above for genes "*pyk*" and "*lctO*" respectively. In Figure 2B expression levels of *pyk* were significantly higher within BMC than LMC. Overall, the highest levels of expression were seen for the infection conditions "Infection, 60 mpi". Figure 2C shows expression levels of *lctO* were relatively low and were comparable in BMC and LMC.

Lineage	Lung or Blood	Variant	Frequency in blood	Frequency in lung	Mutation	Protein	Function
2	Both	rny → I14I (ATC→ATT)	100%	100%	G ⟶ A	Ribonuclease Y	Endoribonuclease activity that initiates mRNA decay
4	Both	P0_01974 → A608S (GCA→TCA)	74.3%	30.3%	C ⟶ A	Lanthionine biosynthesis protein (LanM)	Catalytic activity
4	Both	sdhB → G14R (GGA→AGA)	72.2%	15.9%	C ⟶ T	L-serine dehydratase	Gluconeogenesis and 4 iron, 4 sulfur cluster binding
7	Both	modC → L160L (TTG→TTA)	61.5%	23.7%	C ⟶ T	Molybdenum import ATP-binding protein	Part of the ABC transporter complex ModABC involved in molybdenum import
7	Both	P0_00735 → P90A (CCA→GCA)	61.4%	18.8%	C ⟶ G	Hypothetical protein	N/A
7	Both	ettA_1 → K62N (AAG→AAT)	59.5%	16.9%	C ⟶ A	ABC transporter ATP -binding protein	ATP binding and ATPase activity
Multiple	Blood	pyk → F57L (TTC→TTA)	21.5%	N/A	C ⟶ A	Pyruvate kinase	Involved in synthesis of pyruvate (glycolysis pathway)
Multiple	Both	lctO → V326G (GTC→GGC)	10.5%	49.3% (lineage 2)	T ⟶ G	L lactate oxidase	L-lactate dehydrogenase activity - catalyses production of pyruvate
Multiple	Blood	ilvD → L278I (CTT→ATT)	9.6%	N/A	G ⟶ T	Dihydroxy-acid dehydratase	Amino acid synthesis and metal ion binding Iron-sulfur (4Fe-4S)

Table 1. Table constructed from raw data. In particular, variants in *pyk*, *lctO* and *ilvD* are of interest. This is because they are all present in 3 different lineages, suggesting evidence of parallel evolution. The function of these genes, as shown in the table, include pyruvate metabolism and iron acquisition.

P values were obtained demonstrating that the expression of *ilvD* ($p < 0.0001$) and *pyk* ($p = 0.0079$) were both significantly higher within BMC than LMC. The gene *lctO* was slightly more expressed within BMC than LMC, however the difference was not significant, $P>0.05$.

Discussion

The results of this study show evidence of niche specific adaptations within *S. pneumoniae* to the blood environment. It was found that two genes were both involved in pyruvate metabolism, *pyk* and *lctO*. Gene *ilvD* is also linked to pyruvate as it is involved in the pathway by which pyruvate is converted into valine. The results show that all three of these gene variants are present within multiple lineages within blood isolates and are all linked to or involved in pyruvate metabolism. There are many enzymes involved in pyruvate metabolism and it has been shown that the loss of these enzymes within the pyruvate node can affect pneumococcal virulence (Echlin *et al*., 2020).

The gene *pyk* encodes for the enzyme pyruvate kinase, the function of this enzyme is in the metabolic pathway glycolysis. In glycolysis, pyruvate is synthesised from D-glyceraldehyde 3-phosphate. Pyruvate kinases are present in many bacterial species, for example Gram positive bacterium *Staphylococcus aureus*, due to the unstable nature of bacterial enzymes, pyruvate kinase has not been extensively studied. A study by Zoraghi *et al*., (2010) has suggested that pyruvate kinase may be a novel target for antimicrobial treatments. The results of the study showed that within *S. aureus*, pyruvate kinase was essential to survival of the bacterium. It was also found that pyruvate kinase activity was higher during growth phase of the bacterium. These data suggest pyruvate kinase may be a potential target for antimicrobial therapies. In future studies, similar experiments should be performed using *S. pneumoniae* pyruvate kinase, to assess if the enzyme is essential to pneumococcal survival within the blood.

Gene *lctO* encodes for enzyme L-lactate oxidase which oxidises L-lactate to produce pyruvate, a key role in the generation of energy (ATP) in an aerobic environment (The UniProt, 2018). *LctO* is closely linked to *SpxB*, as they are both part of the central metabolism of pneumococcus in aerobic conditions. *SpxB* encodes for the enzyme pyruvate oxidase, this enzyme decarboxylates pyruvate, releasing ATP. In addition to this function *SpxB* the main producer of hydrogen peroxide, which inhibits the growth of neighbouring bacterial species (Redanz *et al*., 2018). *SpxB* was not identified as a gene variant within the blood isolates in my dataset. This may indicate the production of hydrogen peroxide is not beneficial to the colonisation of the blood, and that if *lctO* is under selection, it is likely to be due to another function of *lctO*, such as the generation of pyruvate.

A genetic variant in *ilvD* occurred within multiple lineages, in blood isolates only. The expression of this gene was significantly higher within BMC than LMC (Aprianto *et al*., 2018). The gene encodes for the protein dihydroxy acid dehydratase (DHAD). This enzyme catalyses the fourth

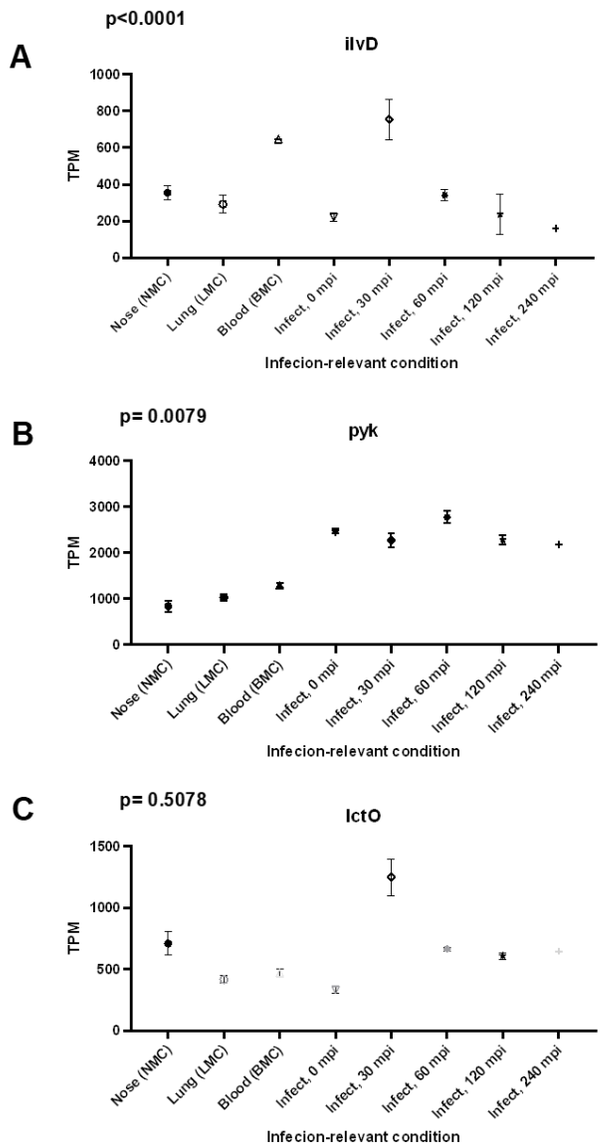

Figure 2. Expression of gene variants in *Streptococcus pneumoniae* D39 in different infection-relevant conditions. Using data provided by PneumoExpress each graph represents a different gene (titled). Graphs show the transcripts per million reads (TPM) of the genes in *S. pneumoniae* D39 for each infection relevant condition. P values gained from output of one-way ANOVA multiple comparison test in which blood (blood-mimicking conditions) were compared to lung (lung-mimicking conditions).

step in the biosynthesis of isoleucine and valine. The DHAD synthesises L-valine from pyruvate, indicating that *ilvD* may be linked to pyruvate metabolism, which is a key role of the gene products mentioned above, for genes *pyk* and *lctO*.

The molecular function of *ilvD* also includes 4 iron, 4 sulfur cluster binding and metal ion binding (UniProt, 2018). Enzymes encoded for by *ilvD* (DHAD) contain an [Fe-S] cluster as a co-factor in the active site (Rahman, *et al*., 2018). This indicates that *ilvD* is involved in the acquisition of iron. Within the blood, iron is bound to haemoglobin with a high affinity. This can create a challenge for bacteria when colonising the blood, as bacterial metabolism requires the use of free iron. *IlvD* may provide an advantage to *S. pneumoniae* in colonisation of the blood by enabling the acquisition of free iron. To confirm this theory, a mutant strain of *S. pneumoniae* without gene *ilvD* may be created using knockout methods.

Conclusion

This study has found evidence of niche specific adaptations to the blood environment, with pyruvate metabolism and iron acquisition being the function of these genes. The research focused on genes *ilvD*, *lctO* and *pyk*. These three genes occurred in multiple lineages, and at multiple passage timepoints. This shows evidence of parallel evolution, and that these genes were under selection within the blood environment. Further research is required, for example, the creation of knockout mutants of *S. pneumoniae* without *ilvD*, *lctO* and *pyk* would confirm if their function is essential in colonisation of the blood. These genes and their gene products may be potential targets for treatments in pneumococcal sepsis.

References

Aprianto, R., Slager, J., Holsappel, S & Veening, J.-W. 2018. High-resolution analysis of the pneumococcal transcriptome under a wide range of infection-relevant conditions. *Nucleic Acids Research*, 46 (19), 9990-10006.

Askim, Å., Mehl, A., Paulsen, J., *et al*. 2016. Epidemiology and outcome of sepsis in adult patients with Streptococcus pneumoniae infection in a Norwegian county 1993-2011: an observational study. *BMC Infectious Diseases*, 16 223-223.

Ceccato, A. & Torres, A. (2018) Sepsis and community-acquired pneumonia. *Annals of Research Hospitals*, 2(7).

Chavanet, P. (2012). Pneumococcus infections: Is the burden still as heavy? *Médecine et Maladies Infectieuses*, 42(4), 149-153.

Echlin, H., Frank, M., Rock, C. & Rosch, J. W. 2020. Role of the pyruvate metabolic network on carbohydrate metabolism and virulence in Streptococcus pneumoniae. *Molecular Microbiology*, 114(4), 356-552.

Kanehisa, M. 2019. Toward understanding the origin and evolution of cellular organisms. *Protein Sci,* 28(11),1947-1951.

Kono, M., Zafar, M. A., Zuniga, *et al*. 2016. Single Cell Bottlenecks in the Pathogenesis of *Streptococcus pneumoniae*. *PLOS Pathogens*, 12 (10), e1005887.

Rahman, M. M., Andberg, M., Koivula, A., Rouvinen, J. & Hakulinen, N. 2018. The crystal structure of D-xylonate dehydratase reveals functional features of enzymes from the Ilv/ED dehydratase family. *Scientific Reports*, 8(1), 865.

Redanz, S., Cheng, X., Giacaman, R. A., *et al*. 2018. Live and let die: Hydrogen peroxide production by the commensal flora and its role in maintaining a symbiotic microbiome. *Molecular Oral Microbiology*, 33 (5), 337-352.

Slager, J., Aprianto, R & Veening, J. W. 2018. Deep genome annotation of the opportunistic human pathogen Streptococcus pneumoniae D39. Nucleic Acids Research. 46(19), 9971-9989.

The UniProt. 2018. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*, 47(D1), D506-D515.

Zoraghi, R., See, R. H., Gong, H., *et al.* 2010. Functional Analysis, Overexpression, and Kinetic Characterization of Pyruvate Kinase from Methicillin-Resistant *Staphylococcus aureus*. *Biochemistry*, 49(35), 7733-7747.