Diverse *Streptococcus pneumoniae* strains drive a MAIT cell response through MR1-dependent and cytokine-driven pathways

Ayako Kurioka[1*], Bonnie van Wilgenburg[1*], Reza Rezaei Javan[1*], Ryan Hoyle[1*], Andries J. van Tonder[1], Caroline L. Harrold[1], Tianqi Leng[1], Prabhjeet Phalora[1], Lauren J. Howson[2], Dawn Shepherd[2], Vincenzo Cerundolo[2], Angela B. Brueggemann[1,3], Paul Klenerman[1]

[1]The Peter Medawar Building for Pathogen Research, Nuffield Department of Medicine, University of Oxford, Oxford, UK

[2]MRC Human Immunology Unit, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK

[3]Department of Medicine, Imperial College London, London, UK

*These authors contributed equally

Abstract

Mucosal Associated Invariant T (MAIT) cells represent an innate T cell population of emerging significance. These abundant cells can recognize ligands generated by microbes utilizing the riboflavin synthesis pathway, presented via the MHC-like molecule MR1 and binding of specific T cell receptors (TCR). They also possess a functional programme (shared by innate T cell populations expressing CD161) allowing microbial sensing in a cytokine-dependent, TCR-independent manner. *Streptococcus pneumoniae* is a major human pathogen that is also associated with commensal carriage, thus host control at the mucosal interface is critical. The recognition of *S. pneumoniae* strains by MAIT cells has not been defined, nor have the genomics and transcriptomics of the riboflavin operon (Rib genes). We examined the expression of Rib genes in *S. pneumoniae* at rest and in response to metabolic stress and linked this to MAIT cell activation *in vitro*. We observed robust recognition of *S. pneumoniae* strains at rest and following stress, using both TCR-dependent and TCR-independent pathways. The pathway used was highly dependent on the antigen-presenting cell, but was maintained across a wide range of clinically-relevant strains. The riboflavin operon was highly conserved across a range of 571 *S. pneumoniae* from 39 countries dating back to 1916, and different versions of the riboflavin operon were also identified in related *Streptococcus* species. These data indicate an important functional relationship between MAIT cells and *S. pneumoniae*, which may be tuned by local factors, including the metabolic state of the organism and the antigen-presenting cell that it encounters.

48 ## Author Summary

49

50 *Streptococcus pneumoniae* is the leading cause of bacterial pneumonia, causes
51 invasive diseases such as meningitis and bacteraemia, and is associated with
52 significant morbidity and mortality, particularly in children and the elderly. Here, we
53 demonstrate that a novel T cell population called Mucosal-associated invariant T
54 (MAIT) cells is able to respond to a diverse range of *S. pneumoniae* strains. We
55 found that this response was dependent on the T cell receptor (which recognises
56 metabolites of the bacterial riboflavin biosynthesis pathway), cytokines, and the type
57 of antigen-presenting cell. A population genomics approach was also used to assess
58 the prevalence and diversity of the genes encoding the riboflavin biosynthesis
59 pathway among a large and diverse collection of *S. pneumoniae*. These genes were
60 highly conserved across a range of 571 *S. pneumoniae* from 39 countries dating
61 back to 1916, and was also present in other related *Streptococcus* species. Given
62 the low levels of MAIT cells in neonates and MAIT cell decline in the elderly, both of
63 whom are at the highest risk of invasive pneumococcal disease, further
64 understanding of the functional role of MAIT cells in host defense against this major
65 pathogen may novel therapeutics or vaccines to be designed.

66

67

68

69

70

Introduction

Streptococcus pneumoniae (the 'pneumococcus') is the most common cause of community-acquired pneumonia and is associated with significant morbidity and mortality, especially among young children and older adults (1,2). *S. pneumoniae* also causes invasive diseases like meningitis and bacteraemia, and upper respiratory tract infections like otitis media and sinusitis, among all age groups (3). Antimicrobial-resistant strains are widespread and pose problems in the treatment of infections, which led the World Health Organisation to include *S. pneumoniae* on their recent list of priority pathogens (4). The currently available pneumococcal conjugate vaccines prompt an immune response to polysaccharide capsules (differentiated as serotypes) and they are highly effective at preventing invasive pneumococcal disease due to vaccine-serotype strains; however, current vaccines only protect against a small number of the possible serotypes which has led to increases in the rates of disease caused by non-vaccine-serotype pneumococci are problematic (5,6). Therefore, pneumococcal disease remains a serious problem and a better understanding of the host defense against *S. pneumoniae* may allow for the design of novel therapeutics or vaccines.

There is an increasing appreciation of the role played by unconventional T cells in orchestrating early cellular events in response to invading pathogens (7). Mucosal-associated invariant T (MAIT) cells are a recently described innate T cell population, which are abundant in mucosal tissues including the lung, as well as the blood and liver (8–10). These cells conventionally express a semi-invariant T cell receptor

(TCR) consisting of a Vα7.2-Jα33/Jα12/Jα20 chain, paired with a limited repertoire of Vβ chains (11). This TCR is able to recognize ligands presented by the conserved major histocompatibility complex (MHC)-related protein 1 (MR1) (8). MR1 binds vitamin-B based precursors derived from the riboflavin biosynthesis pathway, which is conserved across various bacteria and fungi, but does not occur in humans (10). The ligands presented by MR1 are derived from either early intermediates in riboflavin synthesis (5-A-RU) that form adducts with other small metabolites (5-OP-RU), or the direct precursors of riboflavin (e.g. ribityllumazine (RL)-6,7-diMe) (10,12,13). Human MAIT cells are also characterized by a high expression of the C-type lectin-like receptor CD161 and the interleukin (IL)-18 receptor, and are able to respond to innate cytokines even in the absence of prior activation or TCR signaling (14,15). Upon activation, these cells produce various immunomodulatory cytokines including IFNγ, TNFα, IL-17, and Granzyme B.

MAIT cells are critical for the control of bacterial infections in mice, particularly in the lungs (16–18). For instance, aerosol infection models with *Mycobacterium bovis* Bacillus Calmette-Guérin (BCG) and *Francisella tularensis* live vaccine strain (LVS) demonstrated that MAIT cells were essential for the early control of the bacterial burden in the lung (18,19). Indeed, MAIT cells are required to coordinate both innate and adaptive arms of the immune response during pulmonary infections, as early GM-CSF production from lung MAIT cells in response to *F. tularensis* was required for the differentiation of dendritic cells, without which there was a delayed recruitment of activated CD4+ T cells (20). Thus, the rapid and innate activation of

MAIT cells in response to pulmonary bacteria is critical for bridging the innate and adaptive system, shaping the local adaptive immune response.

Despite their role in pulmonary infection models, it remains unclear whether MAIT cells play a role in the defense against *S. pneumoniae* infection. Our first aim was to determine whether MAIT cells can respond to *S. pneumoniae,* and if so, to elucidate the mechanism driving such activation. We found that MAIT cells responded to *S. pneumoniae* in an MR1-dependent manner in the presence of macrophages but not monocytes, and was dependent on co-stimulation provided by innate cytokines. *In vitro* experiments and RNA-sequencing experiments also suggested that temperature and riboflavin availability affected *S. pneumoniae*, which in turn affected the responsiveness of MAIT cells. Secondly, using a population-level genomics approach, we found that the riboflavin synthesis pathway is ubiquitous and highly conserved amongst *S. pneumoniae*, which may have implications in the design of immunotherapy approaches or vaccines that target MAIT cell activity. Riboflavin operon genes were also found among other non-pneumococcal *Streptococcus* species (spp.), including *S. agalactiae* (group B streptococci), which suggests that perhaps the immunological observations made here could be extended to other human-associated *Streptococcus* spp. infections.

Results

S. pneumoniae possess a highly-conserved riboflavin synthesis operon that is upregulated with heat stress

The genes encoding the riboflavin biosynthetic enzymes of *S. pneumoniae* (*ribD*, *ribE*, *ribA* and *ribH*) were found to be clustered together in the same orientation in a predicted 3.4 Kb operon structure (Figure 1A). The prevalence and sequence diversity of the coding regions of the riboflavin genes were investigated in a large, global and historical genome dataset of *S. pneumoniae* isolated between 1916 and 2008 from people of all ages residing in 36 different countries. 561 (98.2%) of the *S. pneumoniae* genomes contained the riboflavin operon. Nine of the ten genomes that lacked the operon were of a single multilocus sequence type (note: $ST^{serotype}$), $ST13^{14/nontypable}$, and the other belonged to $ST695^{4}$ (Supplementary Table 1). All genes in the riboflavin operon were found to be highly conserved: nucleotide and amino acid sequence identity were >99% (Table 1). The dN/dS analysis revealed a higher prevalence of synonymous versus non-synonymous mutations, supporting the importance of maintaining the riboflavin operon (Table 1).

Sequencing was performed on RNA extracted from a pneumococcus that was incubated at a higher temperature than normal (40°C vs. 37°C). Differential expression analysis revealed that all the riboflavin operon genes were significantly up-regulated after 2-4 hours of incubation under heat stress as compared to the control (Figure 1B). Subsequently, the riboflavin operon was found to be significantly

down-regulated after 5–6 hours of incubation. The concurrent increase and decrease in the expression of the four riboflavin genes suggested that these genes are transcriptionally coupled.

MAIT cells are activated by *S. pneumoniae*

To determine whether MAIT cells were able to respond to *S. pneumoniae,* 10 Pneumococcal Molecular Epidemiological Network (PMEN) reference strains were used to probe the activation of MAIT cells in the presence of the monocytic cell line, THP1. Peripheral blood mononuclear cells (PBMCs) were cultured with paraformaldehyde (PFA)-fixed *S. pneumoniae* and THP1 cells overnight, or *Escherichia coli (E. coli)* as a positive control (Figure 2A). There was a clear production of interferon-γ (IFNγ) from MAIT cells across all strains although there was variability in the responses: 7 out of 10 strains reached statistical significance (PMEN2, PMEN13, PMEN14, PMEN34, PMEN35, PMEN36, and PMEN39). Similarly, CD69 expression was induced by all 10 strains as measured by geometric mean fluorescence intensity (gMFI), and reached significance for 7 strains (PMEN9, PMEN12, PMEN13, PMEN34, PMEN35, PMEN36, and PMEN39). In comparison, there was negligible activation of non-MAIT cells (CD161−CD8+ T cells; Figure 2B) across all tests as measured by IFNγ or CD69 expression, suggesting that *S. pneumoniae* specifically activated MAIT cells.

188 **MAIT cell activation by *S. pneumoniae* in the presence of monocytes is not**
189 **MR1-dependent**
190
191 We have previously found that the response of MAIT cells to *E. coli* is dependent on
192 both MR1 as well as innate cytokines IL-12 and IL-18 (14). To investigate whether
193 the response of MAIT cells to *S. pneumoniae* was dependent on MR1, cytokines, or
194 both, we cultured PFA-fixed *S. pneumoniae* with PBMCs and THP1 cells in the
195 presence of anti-MR1, anti-IL-12, and anti-IL-18 blocking antibodies (Figure 3A). As
196 expected, IFNγ expression in response to *E. coli* could be blocked significantly by
197 anti-IL-12 and anti-IL-18 blocking antibodies, which was further reduced and
198 abrogated by the addition of an anti-MR1 blocking antibody, showing that both the
199 TCR and cytokine signaling pathways contribute to MAIT cell activation by *E. coli*.
200 Surprisingly, we found that blockade of MR1 had no effect on pneumococcal MAIT
201 cell activation. Instead, blocking IL-12 and IL-18 completely abrogated MAIT cell
202 activation across all strains tested. This suggested that despite the fact that *S.*
203 *pneumoniae* possesses the riboflavin synthesis pathway (Table 1, Figure 1A),
204 activation of MAIT cells by *S. pneumoniae* in the presence of THP1 cells was
205 dependent on cytokines, and not activation through antigen presentation by MR1.
206
207 Next, to investigate whether *S. pneumoniae* can activate MAIT cells through the
208 MR1 pathway alone, Jurkat cells engineered to express the MAIT cell TCR Vα7.2-
209 Jα33 were cultured with fixed *S. pneumoniae* strains and cultured overnight in the
210 presence of THP1 cells (Figure 3B). There was no significant change in the
211 expression of CD69 by MAIT-Jurkats in the presence of any of the *S. pneumoniae*

strains, although the cells were able to respond to the *E. coli* positive control. Thus, in the presence of monocytic cells, there was very little activation of MAIT cells by *S. pneumoniae* through the MR1 pathway.

Given the upregulation of the riboflavin synthesis pathway in *S. pneumoniae* upon heat stress at 40˚C (Figure 1B), we tested whether changing environmental factors such as temperature and availability of riboflavin in the growth media would trigger activation of MAIT cells through the MR1 pathway. *S. pneumoniae* strain PMEN34 was grown for 16 hours in Todd Hewitt Broth (THB) media supplemented with yeast (THB-Y) at 36˚C and then either transferred to 40˚C incubation, or transferred to riboflavin-free assay media (RAM) at 36˚C or 40˚C for 4 hours, before the bacteria were fixed (Figure 3C). Although there was a slight increase in the fraction of MAIT cells expressing IFNγ when bacteria were cultured in the absence of riboflavin in RAM, regardless of temperature, this increase was not dependent on MR1 as the addition of an anti-MR1 blocking antibody did not reduce the frequency of IFNγ-expressing MAIT cells. Thus, the complete absence of riboflavin, increased incubation temperature or the combination of both did not lead to MR1-dependent activation of MAIT cells.

We also tested whether using live *S. pneumoniae* strain PMEN34 or the supernatant of *S. pneumoniae* growth culture, instead of fixed bacteria, would stimulate MAIT cells through the MR1-pathway (Figure 3D); however, these responses were small and could not be significantly blocked by an anti-MR1 blocking antibody. We enriched for CD8+ T cells and added these to the assay instead of PBMCs in case

other cells were interfering with MAIT cell activation through MR1, but these responses were similar to those using PBMCs and not affected by MR1-blocking. Thus, in the presence of monocytes, MAIT cells are activated mainly through innate cytokines rather than MR1, regardless of temperature or riboflavin availability.

MR1-dependent activation of MAIT cells by *S. pneumoniae* in the presence of macrophages

We next tested whether monocyte-derived macrophages may be able to present the MR1-ligand to activate MAIT cells more effectively through MR1, as alveolar macrophages play an important role in the immune response to *S. pneumoniae* (21,22). Furthermore, we investigated whether temperature or the abundance of riboflavin in the media affects MR1-dependent activation of MAIT cells in the presence of macrophages. For this, *S. pneumoniae* strain PMEN34 was grown for 16 hours in THB media, with or without yeast extract (THB-Y and THB, respectively) at either 36˚C or 40˚C. Given that the riboflavin synthesis pathway was upregulated by a short incubation at 40˚C (Figure 1), we also transferred half of the bacteria grown overnight at 36˚C to 40˚C for 4 hours. The bacteria were fixed immediately and added to PBMCs and monocyte-derived macrophages overnight (Figure 4A, B).

Surprisingly, we found that when using monocyte-derived macrophages, *S. pneumoniae* induced IFNγ expression from MAIT cells that was reduced by an anti-MR1 blocking antibody, suggesting that the response was MR1-dependent, in contrast to the response in the presence of monocytes. The addition of the anti-MR1

blocking antibody significantly reduced the frequency of IFNγ-expressing MAIT cells regardless of the temperature and media in which the *S. pneumoniae* were grown. Interestingly, there was a clear increase in activation induced by bacteria grown in THB media without the addition of yeast, compared to bacteria grown in the presence of yeast. Degranulation was also induced by *S. pneumoniae* grown in THB media and was blocked by the anti-MR1-blocking antibody to a varying degree (Figure 4C, D).

Next, in order to compare the dependency of MAIT cell activation by *S. pneumoniae* on MR1, IL-12 and IL-18 in the context of macrophages, *S. pneumoniae* strain PMEN34 was grown for 16 hours in THB or THB-Y media at either 36˚C or 40˚C. Half of the bacteria grown overnight at 36˚C were transferred to 40˚C for an additional 4 hours. Then the bacteria were fixed and added to PBMCs and macrophages overnight in the presence of anti-MR1, anti-IL-12, and anti-IL-18 antibodies (Figure 4E, F). We found that there was a significant effect of blocking MR1 on IFNγ production from MAIT cells in the presence of *S. pneumoniae* cultured in THB media, regardless of temperature, but this was almost completely abrogated in the presence of anti-IL-12 and anti-IL-18 blocking antibodies. There was a similar effect of blocking antibodies on MAIT cell IFNγ production in response to *S. pneumoniae* cultured in THB-Y media overnight at 40˚C. Furthermore, although we have previously found that degranulation in response to *E. coli* was completely dependent on MR1 and not on cytokines (23), degranulation of MAIT cells in response to *S. pneumoniae* was, in each case, further reduced by the addition of anti-IL-12 and anti-IL-18 compared to blocking MR1 alone. In particular,

degranulation in response to *S. pneumoniae* grown at 40˚C for the last 4 hours in THB media was significantly reduced by anti-MR1 blocking, which was further reduced by the addition of anti-IL-12 and anti-IL-18 blocking antibodies (Figure 4F). This suggests that degranulation in response to *S. pneumoniae* requires co-stimulation from cytokines.

Finally, we set out to confirm that *S. pneumoniae* were able to activate MAIT cells in the presence of macrophages in an MR1-dependent manner. We cultured Jurkat cells that express the MAIT cell TCR Vα7.2-Jα33 with fixed *S. pneumoniae* grown in THB or THB-Y media at different temperatures in the presence of monocyte-derived macrophages (Figure 4G). This showed that *S. pneumoniae* grown in THB media significantly increased the expression of CD69 in MAIT-Jurkat cell lines, compared to the Jurkat cells cultured with the sterility control, although the response was relatively weak. Thus, in the presence of monocyte-derived macrophages, *S. pneumoniae* were able to activate MAIT cells in an MR1-dependent manner, but were highly dependent on the presence of cytokines to boost the response.

Riboflavin operons are also present in non-pneumococcal *Streptococcus* spp.

A bioinformatic investigation of 824 genomes of 69 different *Streptococcus* spp. revealed that the riboflavin operon was also present in other streptococci. 11 different versions of the riboflavin operon were identified among 13 non-pneumococcal *Streptococcus* spp. (Supplementary Table 2). The majority of these riboflavin operons were located between genes involved in arginine biosynthesis

(*argC*, *argJ*, *argB*, *argD*) and ribonucleotide reduction (*nrdF2*, *nrdE2*, *nrdH*) (Figure 5A). Despite identical gene synteny between different versions of the riboflavin operon (Figure 5A), they differed greatly in nucleotide sequence identity (Figure 5B).

The *S. pneumoniae* version of the riboflavin operon was found among all 16 genomes of *S. pseudopneumoniae* and two genomes each of *S. mitis* and *S. oralis* (Figure 5C-D). No riboflavin operon genes were identified among the remaining 48 *S. mitis* and 48 *S. oralis* genomes. *S. pneumoniae*, *S pseudopneumoniae*, *S. mitis* and *S. oralis* are all closely-related commensal streptococcal species that can exchange DNA with one another; therefore the limited number of riboflavin operons present in *S. mitis* and *S. oralis* suggest that the examples identified here were the result of horizontal genetic exchange (24). Some versions of riboflavin operons identified among other *Streptococcus* spp. were exclusively present in only one species, e.g. version 2 was identified in all 50 genomes of *S. agalactiae* but no other species; whereas in contrast, *S. equinus* contained three different versions of the riboflavin operon, one of which (version 5) was also found in *S. infantarius* (Figure 5D; Supplementary Table 2). Furthermore, over 2400 *S. pyogenes* genomes were investigated for the presence of Rib genes. The riboflavin operon could not be found in the genomes of *S. pyogenes* (data not shown).

Discussion

Our study is the first to demonstrate that MAIT cells can respond to and recognize *S. pneumoniae*. Given the urgent global need to tackle antimicrobial-resistant strains and pneumococci not covered by the currently available vaccines, understanding the mechanism by which MAIT cells are activated by *S. pneumoniae* provides a method of targeting a metabolic pathway that is highly conserved amongst *S. pneumoniae*.

We found that MAIT cell activation by *S. pneumoniae* through the MR1-restricted pathway was dependent on the type of antigen-presenting cell. There was a significant effect of blocking MR1 recognition by MAIT cells only when the *S. pneumoniae* was presented by macrophages, but not monocytes (either those amongst PBMCs or by the monocytic cell line THP1). Indeed, the lack of MAIT cell activation by *S. pneumoniae* in a previous paper (25) is in line with our finding that monocytes are not able to sufficiently activate MAIT cells solely through the presentation of ligands from *S. pneumoniae* by MR1. The reason for the difference in the ability of these antigen-presenting cells to activate MAIT cells through MR1 antigen presentation in the case of *S. pneumoniae* is not clear. It has been suggested that monocytes are poor antigen presenters (26) and instead function as antigen-transporting cells that deliver antigen to the draining lymph nodes (27). Although THP1 cells and monocytes have been used repeatedly to activate MAIT cells in an MR1-dependent manner with numerous other bacteria, the pathogenesis of *S. pneumoniae* is related to its many virulence factors, some of which are associated with minimizing phagocytosis and thereby antigen presentation (further

356 discussed below). Macrophages may be a more powerful antigen-presenting cell that
357 are able to efficiently phagocytose the bacteria, provide more co-stimulatory signals,
358 or resist interference with antigen presentation and loading of MR1. Indeed, mice
359 lacking alveolar macrophages, located at the interphase between air and lung tissue,
360 show a delayed and impaired bacterial clearance as compared with control mice
361 (21,22). Thus, alveolar macrophages provide the first line of cellular defense against
362 microbes and may be an important presenter of *S. pneumoniae* antigen to MAIT
363 cells.
364
365 Even in the presence of macrophages, the MAIT cell response to *S. pneumoniae*
366 was highly co-dependent on cytokines. This is most likely due to the weak TCR
367 signal induced by *S. pneumoniae*, as seen by the significant but weak activation of
368 MAIT-Jurkat cell lines even when using macrophages. MAIT cells are characterized
369 by downregulated TCR signaling components and a high dependency on co-
370 stimulation provided by cytokines for activation and expansion (28–30). The
371 particularly weak TCR signal provided by *S. pneumoniae* may be because *S.*
372 *pneumoniae* is a bacterium highly specialized to evade the host immune system by
373 circumventing phagocytosis and antigen presentation. The efficiency of phagocytosis
374 can be impeded by the presence of the thick polysaccharide capsule, while another
375 virulence-associated trait is the ability of this bacterium to undergo autolysis.
376 Autolysis of *S. pneumoniae* has been shown to specifically reduce the phagocytosis
377 of intact bacteria, even in the absence of a polysaccharide capsule, and reduce the
378 production of phagocyte-activating cytokines such as TNFα, IFNγ and IL-12 (31).

379 The uniquely high sensitivity of MAIT cells to cytokines such as IL-12 (32) allows
380 these cells to boost the immune response and provide early IFNγ production.
381
382 The biggest disease burden caused by *S. pneumoniae* by far is pneumonia, which is
383 the leading infectious cause of mortality in children under five years of age (1), as
384 well as accounting for a large part of community-acquired pneumonia in the elderly
385 (2). Given that we show the mechanism by which MAIT cells respond to *S.*
386 *pneumoniae,* and the critical role that pulmonary MAIT cells play *in vivo* against lung
387 infections (18,20), it would be reasonable to suggest that MAIT cells may play a role
388 in pneumococcal pneumonia. These cells may also be a critical factor in secondary
389 bacterial pneumonia that follow influenza infection, which are most commonly
390 caused by *S. pneumoniae* and associated with high mortality (33). MAIT cells are
391 reduced in frequency in the blood of patients with acute influenza infection,
392 particularly in those who succumbed to this disease (15,34). The loss of MAIT cells
393 due to primary influenza infection may leave the patient vulnerable to secondary
394 infection by *S. pneumoniae*, due to the lack of early control and immunomodulation
395 provided by these cells, as has been shown with γδ T cells (35). Whether the low
396 numbers of MAIT cells in neonates (36,37), or the decline in MAIT cell numbers in
397 the elderly (38) and in influenza (15) affects the susceptibility of these patients to
398 pneumococcal pneumonia will be important to investigate in *in vivo* models of
399 pneumococcal infection (39).
400
401 Interestingly, we found that temperature and riboflavin availability for *S. pneumoniae*
402 have the potential to affect riboflavin synthesis and MAIT cell activation. In our

models using macrophages, there was a difference in MR1-dependent MAIT cell activation induced by *S. pneumoniae* cultured in THB media compared to THB-Y media, the latter of which is supplemented with riboflavin-rich yeast extract. Although the THB media alone supports the growth of riboflavin-auxotrophs such as *Streptococcus pyogenes*, this has been shown to be due to the ability of *S. pyogenes* to utilize THB-derived components as a substitute for riboflavin, and mutant strains without this ability require the addition of riboflavin or riboflavin-rich yeast in order to grow in THB media (40). Thus, THB media on its own does not contain riboflavin. We found that *S. pneumoniae* grown in THB media consistently induced greater MAIT cell activation, which may be partly due to the increased production of riboflavin in bacteria grown in riboflavin-deficient THB media, compared to those grown in riboflavin-rich THB-Y media. We also found that riboflavin synthesis among *S. pneumoniae* was influenced by temperature: RNA-seq transcriptomic analyses of *S. pneumoniae* revealed that incubation at 40°C upregulated the riboflavin synthesis genes. A relationship between riboflavin production and temperature has been demonstrated before in *Lactococcus lactis*: it suffers from oxidative stress caused by riboflavin starvation at high temperatures and results in severe growth inhibition, which can be improved by the addition of riboflavin to the medium (41). Although no consistent increase in MR1-dependent MAIT cell activation could be detected in response to pneumococci grown at higher temperatures in our short-term activation models, it may be possible that MAIT cells are able to react to increased riboflavin production of *S. pneumoniae* at higher body temperatures during a fever. Thus, MAIT cells may be able to not only recognize the presence of *S. pneumoniae* but respond to the environment and condition in which

427 the bacteria is growing - an important indicator of the pathogenic potential of a
428 commensal pathogen such as *S. pneumoniae.*
429
430 We used a population genomics approach to assess the prevalence and diversity of
431 the riboflavin operon among a large and diverse collection of *S. pneumoniae*. This
432 revealed that the riboflavin genes are nearly ubiquitous and highly conserved at a
433 nucleotide level among *S. pneumoniae* that had been recovered over the past
434 century. This provides confidence that the findings presented here are generalizable
435 among *S. pneumoniae*, with rare exceptions.
436
437 We also assessed the presence of riboflavin synthesis genes among non-
438 pneumococcal *Streptococcus* spp. and found that a number of other species do
439 possess these genes, including other human-associated commensal streptococci
440 like *S. agalactiae*. Notably though, not all *Streptococcus* spp. possessed a riboflavin
441 operon (at least one that was detectable by our screening method) and there was
442 variation among the genomes that did, either in that more than one version of
443 riboflavin operon was detected within a bacterial species, or that the presence of a
444 riboflavin operon was not ubiquitous among genomes of a particular bacterial
445 species. Hence, caution must be exercised when extrapolating findings based on a
446 small number of bacterial strains to the population as a whole since they may not be
447 representative, and ideally a population-based approach is used whenever possible.
448 However, the fact that other *Streptococcus* spp. also possess a riboflavin operon
449 presents opportunities for future studies. For example, given that MAIT cells reside in
450 the female genital mucosa (42) it would be important to explore whether there is a

MAIT cell response in the context of vaginal colonization of *S. agalactiae* (group B streptococci) among pregnant women and invasive neonatal infections (43).

Overall these data show a robust response of MAIT cells to *S. pneumoniae*, and conservation of the relevant biosynthetic pathway in this organism and other closely-related *Streptococcus* spp. Given the low levels of MAIT cells in early life and their decline in old age – the highest risk populations for invasive pneumococcal disease – further understanding of the functional role of MAIT cells *in vivo* in host defense against this major pathogen should be of substantial interest.

Materials and Methods

Flow cytometry

For immunofluorescence staining, dead cells were excluded with the Live/Dead Fixable near-IR dead-cell stain (Invitrogen). For internal staining, cells were fixed with 1% formaldehyde (Sigma Aldrich) and permeabilised with permeabilisation buffer (eBioscience). Antibodies used were: CD3 Pacific Orange, Granzyme B APC (Life Technologies), CD8 PE-Cy7, TNFα PerCP-Cy5.5, Vα7.2 APC, FITC, or PE, CD107a PE-Cy7 (BioLegend), CD3 efluor 450, CD8 PerCP-Cy5.5, CD69 FITC or Pacific Blue, IFNγ PerCP-Cy5.5 (eBiosciences), CD161 PE, CD4 VioGreen, IFNγ FITC (Miltenyi Biotec).

Cells

Whole blood was obtained from leukocyte cones (NHS Blood and Transplant). PBMCs were isolated by standard density gradient centrifugation (Lymphoprep™ Axis Shield Diagnostics). All samples were collected with appropriate patient consent and local research ethics committee approval (COREC 04.OXA.010). Where indicated, CD8+ cells were positively enriched using CD8 Microbeads (Miltenyi Biotech) before *in vitro* stimulation. Monocyte-derived macrophages were generated by enriching for monocytes using CD14 Microbeads (Miltenyi Biotech) before culturing with 50ng/ml GM-CSF (Miltenyi Biotech) in RPMI media, penicillin/streptomycin, L-glutamine, and 10% human serum for 6-8 days.

The Jurkat-MAIT cell line was produced by cloning into a pHR-IRES vector the TRAV1-2–TRAJ33 (CDR3α: CAVMDSNYQLIW) MAIT cell α chain and then cloning in the TRBV6-1–TRBJ2-3 β chain (CDR3β-CASSETSGSPDTQYF). The vector was then co-transfected into 293T cells with the HIV gag-pol and VSV-G expression plasmids using X-tremeGENE™ 9 DNA Transfection Reagent (Sigma) according to the manufacturer's instructions. The supernatant from this culture containing the lentiviral particles was then used to transduce J.RT3-T3.5 (JRT3) cells (which lack an endogenous TCRβ chain). TCR expression and pairing with MAIT cell TRAV1-2 (Vα7.2) chain was confirmed by flow cytometry and cells sorted based on CD3 and Vα7.2 expression.

Pneumococcal reference strains used in experiments

Ten PMEN reference strains of *S. pneumoniae* were tested in this study (note: country of first detection^{serotype}; multilocus sequence type): PMEN2 (Spain^{6B}-2; ST90), PMEN3 (Spain^{9V}-3; ST156), PMEN9 (England^{14}-9; ST9), PMEN12 (Finland^{6B}-12; ST270), PMEN13 (S.Africa^{19A}-13; ST41), PMEN14 (Taiwan^{19F}-14; ST236), PMEN34 (Denmark^{12F}-34; ST218), PMEN35 (Netherlands^{14}-35; ST124), PMEN36 (Netherlands^{18C}-36; ST113), PMEN39 (Netherlands^{7F}-39; ST191)(44–46). Eight serotypes were represented, all of which are vaccine serotypes apart from 12F (PMEN34).

PMEN strains were cultured from freezer stocks to Columbia blood agar plates (Oxoid) and incubated overnight at 37˚C+5% CO_2, then transferred to THB media (Sigma Aldrich) with 0.5% yeast extract (THB-Y; Sigma Aldrich) and incubated at

522 37˚C + 5% CO_2 overnight, unless indicated otherwise. In specific experiments,
523 bacteria were grown in Riboflavin Assay Media (RAM; BD Difco). *E. coli* (DH5a,
524 Invitrogen) were cultured in Lysogeny Broth (LB) medium and incubated overnight in
525 a shaking incubator.
526
527 *S. pneumoniae* or *E. coli* were fixed in 1% PFA for 15 minutes, and washed
528 extensively in sterile, filtered phosphate-buffered saline (PBS). A negative control
529 was prepared in identical fashion and used as a sterility control. Following fixing,
530 bacteria were immediately used in *in vitro* stimulation assays. Alternatively, filtered
531 bacterial culture supernatant or unfixed live bacteria was added to assays where
532 indicated. Live pneumococci were added to PBMCs and antigen-presenting cells for
533 30 minutes at 37ºC before cells were extensively washed and cultured for the
534 remainder of the overnight assay in 100μg/ml gentamycin.
535
536 ***In vitro* stimulation of MAIT cells**
537 THP1 cells (ECACC, UK) were incubated overnight with PFA-fixed *S. pneumoniae* or
538 *E. coli* at a ratio of 30 bacteria per cell (BpC), or the sterility control (without bacteria,
539 see above). THP1 cells were washed and PBMCs or enriched CD8+ T cells were
540 added to THP1 cells overnight. Brefeldin A (eBioscience) was added for the final 4
541 hours of the stimulation. Alternatively, for the assessment of degranulation, anti-
542 CD107a PE-Cy7 (BioLegend) was added from the start of the stimulation. For
543 blocking experiments, anti-MR1 (BioLegend), anti-IL-12p40/70, anti-IL-18 antibody
544 (both BioLegend), or the appropriate isotype controls, were added for the duration of
545 the experiment.

546 **Data acquisition and statistical analysis**

547 Data on MAIT cell activation were collected on the MACSQuant Analyser (Miltenyi
548 Biotech) and were analysed using FlowJo v9.8 (TreeStar). All graphs and statistical
549 analyses were completed using GraphPad Prism software version 6. All data are
550 presented as means with standard error of the mean (S.E.M.) unless otherwise
551 indicated.

552

553 **Compilation of the genome datasets**

554 Two large genome datasets were compiled for this study and data were stored in a
555 BIGSdb database (47). The *S. pneumoniae* dataset consisted of 571 historical and
556 modern genomes isolated from 1916–2009 from people of all ages residing in 39
557 different countries. The *S. pneumoniae* isolates were recovered from both carriage
558 and disease, and 89 serotypes and 296 multilocus sequence types were represented
559 in this dataset (Supplementary Table 1). 486 *S. pneumoniae* genome sequences
560 were compiled from previously published studies or were downloaded from GenBank
561 (48). The remaining 85 *S. pneumoniae* genomes were recently sequenced. *S.*
562 *pneumoniae* cultures were prepared and incubated as described above, before DNA
563 was extracted using the Promega Maxwell® 16 Instrument and Buccal Swab LEV
564 DNA purification kits following the manufacturer's protocol. DNA extracts were sent
565 to the Oxford Genomics Centre where libraries were made and DNA was sequenced
566 on the Ilumina® platform. Velvet was used to make *de novo* genome assemblies,
567 which were further improved using SSPACE and GapFiller (49–51).

568

The non-pneumococcal *Streptococcus* spp. dataset contained 834 genomes of 69 different streptococcal species (Supplementary Table 2). 34 genomes were newly sequenced as above and the rest were downloaded from the ribosomal multilocus sequence typing (rMLST) database (52). For each species, the number of genomes included in this study dataset was capped at 50: if fewer than 50 genomes were available in the rMLST database for a given species, then all available genomes were included, but if more than 50 sequenced genomes were available then genomes were manually selected for inclusion. In these instances, the population structure of the species was depicted using PHYLOViZ and 50 genomes were selected with the aim of maximising the population-level diversity of that species from the available genomes (53).

Genomes were annotated using both RAST (54) and Prokka (55). Genome sequences can be accessed from the European Nucleotide Archive (56), PubMLST website(57), GenBank(58) and/or the rMLST database(59) (see Supplementary Tables 1 and 2).

RNA-sequencing experiments and analyses

S. pneumoniae isolate 2/2 was cultured in seven 10 ml tubes of brain-heart infusion broth and incubated at 40°C + 5% CO_2 for 6 hours to mimic heat shock. The experimental control was the same isolate 2/2, also cultured in seven 10 ml tubes of brain-heart infusion broth but incubated at standard conditions of 37°C + 5% CO_2. Broth cultures at five time points (2, 3, 4, 5 and 6 hours of incubation) were removed and 19 ml of RNAprotect Bacteria Reagent (Qiagen) was added to stabilise the RNA.

RNA was extracted from the samples using the Promega Maxwell® 16 Instrument and LEV simplyRNA Cells purification kit, following the manufacturer's protocol. Extracted RNA samples were sent to the Oxford Genomics Centre for processing. Library preps were made using RNA-Seq Ribozero kits (Illumina, Inc) and sequencing was performed on the MiSeq (Illumina, Inc).

The sequenced forward and reverse reads were paired and mapped to the *S. pneumoniae* strain 2/2 genome using Bowtie2 with the highest sensitivity option (60). Differential gene expression was analysed in Geneious version 9.1 (Biomatters Ltd) using the DESeq method (61). Genes with an adjusted p-value < 0.05 were considered to be differentially expressed. RNA sequencing data used in this paper have been deposited in the Gene Expression Omnibus (GEO) database with accession number GSMXXXXXXX (pending).

Genomic analyses of riboflavin operon genes

Genes involved in riboflavin metabolism were identified using the KEGG pathway database (KEGG entry number: snp00740) and previously published experimental work (12,62). Individual BLAST searches of *ribD*, *ribE*, *ribA* and *ribH* sequences among all 571 *S. pneumoniae* genomes were performed via the BIGSdb database. There were two instances each for *ribD* and *ribH* where the gene sequences were split over multiple assembly contigs and these sequences were excluded from further analyses. Individually, multiple nucleotide sequence alignments for *ribD*, *ribE*, *ribA* and *ribH* were performed in Geneious using the ClustalW algorithm with default parameters (Gap open cost=15, Gap extend cost=6.66) (63). To compute the dN/dS

ratio, the number of synonymous (dS) and non-synonymous (dN) substitutions per site was determined on codon-aligned sequences using a maximum likelihood method, conducted through HyPhy (64) on the Datamonkey server (65,66). For the pairwise positive selection analysis, amino acid sequences were aligned in Geneious using the ClustalW algorithm (Cost matrix: BLOSUM, Gap open cost=10, Gap extend cost=0.1).

Identification and categorisation of the riboflavin operon genes in non-pneumococcal *Streptococcus* spp.

The *ribD*, *ribE*, *ribA* and *ribH* sequences in *S. pneumoniae* strain 2/2 genome were used as the query to BLAST against the non-pneumococcal *Streptococcus* spp. genome dataset (parameters: word size 11; reward 5; penalty -4; gap open 8; gap extend 6). All BLAST hits were manually inspected to confirm the presence of the riboflavin genes. Protein domains in the identified genes were annotated using the Conserved Domain search feature at NCBI (67) to confirm the presence of riboflavin biosynthesis genes. To categorise the different versions of the riboflavin operon, all sequences were clustered using CD-Hit (68) at a ≥90% similarity threshold, and a representative sequence from each cluster was selected. One 'cluster' contained a single riboflavin operon sequence that was disrupted by a transposon and this single sequence was excluded from further analyses. Multiple nucleotide sequence alignments of different versions of the riboflavin operons was performed in Geneious using the ClustalW algorithm with default parameters (Gap open cost=15, Gap extend cost=6.66). The multiple nucleotide sequence alignment output was used within the Geneious environment to calculate the percentage identity matrix.

641 **Assessment of the relationships among *Streptococcus* spp.**

642 A phylogenetic tree was built using concatenated sequence data from the ribosomal
643 loci using the neighbour-joining method as implemented using the BIGSdb
644 PhyloTree plugin (52,69). The tree was annotated using iTOL (70) and Inkscape (71).

645

646 ## Acknowledgements

647 The authors gratefully acknowledge Professor Regine Hakenbeck at the University
648 of Kaiserslautern for the stock cultures of streptococci that were newly sequenced in
649 this study.

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

Figure legends

Figure 1. Genetic and transcriptomic data related to the riboflavin operon in *S. pneumoniae*. A) The riboflavin operon is depicted with riboflavin genes *ribD*, *ribE*, *ribA* and *ribH* (red) and flanking genes (grey). B) RNA expression data at five timepoints (2-6 hours after initial incubation) are illustrated for each riboflavin gene and the flanking genes. Genes marked with differential expression levels in green were upregulated and those in red were downregulated during 40°C incubation, as compared to normal 37°C incubation. An asterisk (*) to the left of a cell indicates a statistically significant result.

Figure 2. MAIT cells are activated by *S. pneumoniae*. 10 Pneumococcal Molecular Epidemiological Network (PMEN) reference strains were used to probe the activation of MAIT cells in the presence of the monocytic cell line, THP1. *Escherichia coli (E. coli)* were added as a positive control. (A-B) Frequency of cells expressing IFNγ in MAIT cells (A) or CD161- CD8+ T cells (B) are shown (n=9). (C-D) CD69 expression measured by geometric mean fluorescence intensity (gMFI) in MAIT cells (C) or CD161- CD8+ T cells (D) are shown (n=6). **P<0.01, *P<0.05 by repeated measures one-way ANOVA with Dunnett's multiple comparisons test, compared to the sterility control (SC). Numbers indicate the PMEN reference strains. SC= sterility control.

Figure 3. MAIT cell activation by *S. pneumoniae* in presence of monocytes is not MR1-dependent. (A) Paraformaldehyde (PFA)-fixed *S. pneumoniae* PMEN reference strains or *E. coli* were cultured with PBMCs and THP1 cells in the

presence of anti-MR1, anti-IL-12, and anti-IL-18 blocking antibodies. IFNγ expression from MAIT cells are shown. ****P<0.0001, ***P<0.001, **P<0.01 by repeated measures one-way ANOVA with Dunnett's multiple comparisons test, compared to the no antibody control (n=7). (B) Jurkat cells expressing the MAIT cell TCR (white bars) or control cells not expressing the MAIT cell TCR (black bars) were cultured with THP1 cells overnight with the indicated PMEN strains of *S. pneumoniae*, or *E. coli* as a positive control. Activation was measured as the geometric mean fluorescence intensity (gMFI) of CD69 expressed by Jurkat cells. SC=sterility control baseline activation, as indicated by the dotted line. Representative of three independent experiments. (C) The *S. pneumoniae* PMEN34 strain was grown overnight at 36°C or 40°C in THB-Y media and either cultured in THB-Y media for the last 4 hours or transferred to riboflavin-free media (RAM). The bacteria were then fixed and cultured with PBMCs and THP1 cells overnight. Frequency of IFNγ expressing MAIT cells is shown, in the presence or absence of anti-MR1 blocking antibody. ns=non-significant by two-way ANOVA with Sidak's multiple comparisons test (n=3). (D) The *S. pneumoniae* PMEN34 strain was grown overnight and the supernatant, fixed or live bacteria were added to THP1 cells overnight with either PBMCs (left) or enriched CD8+ T cells (right), in the presence or absence of anti-MR1 blocking antibody. Frequency of IFNγ expressing MAIT cells are shown. SC=sterility control, ns=non-significant by two-way ANOVA with Sidak's multiple comparisons test (n=3).

Figure 4. MR1-dependent activation of MAIT cells by *S. pneumoniae* in the presence of macrophages. A-D) The *S. pneumoniae* PMEN34 strain was grown

for 16 hours in THB media, with or without yeast extract (THB-Y and THB, respectively) at either 36˚C (36) or 40˚C (40 overnight). Half of the bacteria grown overnight at 36˚C were transferred to 40˚C for 4 hours (40 4 hr). The bacteria were fixed immediately and added to PBMCs and monocyte-derived macrophages overnight, in the presence or absence of anti-MR1 blocking antibody. Frequency of MAIT cells expressing IFNγ (A) and representative example of IFNγ expression from MAIT cells by FACS (B) are shown with isotype control or anti-MR1 blocking antibody. ****P<0.0001, ***P<0.001, and * P<0.05 by two-way ANOVA with Sidak's multiple comparisons test (n=6). Frequency of MAIT cells expressing CD107a (C) and representative example of CD107a expression from MAIT cells by FACS (D) are shown with isotype control or anti-MR1 blocking antibody. ***P<0.001, **P<0.01, or non-significant (ns) by two-way ANOVA with Sidak's multiple comparisons test (n=4). 36=36°C, 40=40°C. E-F) The *S. pneumoniae* PMEN34 strain was grown for 16 hours in THB-Y or THB media, at either 36˚C or 40˚C. Half of the bacteria grown overnight at 36˚C were transferred to 40˚C for 4 hours. The bacteria were fixed immediately and added to PBMCs and monocyte-derived macrophages overnight, in the presence or absence of indicated combinations of anti-MR1, anti-IL-12, and anti-IL-18 blocking antibodies. Frequency of MAIT cells expressing IFNγ (E) and CD107a (F) are shown. 36=36°C, 40=40°C. ****P<0.0001, ***P<0.001, **P<0.01, *P<0.05 or non-significant (ns) by two-way ANOVA with Dunnett's multiple comparisons test (n=3). (G) Jurkat cells expressing the MAIT cell TCR were cultured with monocyte-derived macrophages overnight with the PMEN34 strain of *S. pneumoniae*. Activation was measured as frequency of Jurkat cells expressing CD69. Dotted line indicates CD69 expression by Jurkats in the presence of sterility control (SC).

737 THB=Todd Hewitt Broth, THB-Y=THB with yeast extract. *P<0.05 by repeated
738 measures one-way ANOVA with Dunnett's multiple comparisons test. All
739 experiments were performed in duplicates, and representative of two independent
740 experiments.
741
742 **Figure 5. Evidence for different versions of riboflavin operons in other**
743 ***Streptococcus* spp.** A) The riboflavin operon found in *S. pneumoniae* (version 1)
744 and its flanking genes are depicted and compared to 10 additional representative
745 versions of the riboflavin operon found among other *Streptococcus* spp. B) Matrix of
746 pairwise comparisons of nucleotide similarity among the 11 different versions of the
747 riboflavin operon. C) Table summarizing the riboflavin operons found in 13 non-
748 pneumococcal *Streptococcus* spp. D) Phylogenetic tree constructed based upon the
749 concatenated sequences of 53 ribosomal MLST loci among 571 *S. pneumoniae* and
750 824 *Streptococcus* spp. genomes. Branches of the tree were coloured grey if no
751 riboflavin operon was identified within the genome, whereas other colours represent
752 genomes in bacterial species that did possess a version of a riboflavin operon. The
753 coloured outer ring indicates the version of riboflavin operon that was identified in
754 each genome or set of genomes. The rectangular box contains an expanded view of
755 the circled area of the phylogenetic tree.
756
757
758
759
760

References

1. Liu L, Oza S, Hogan D, Perin J, Rudan I, Lawn JE, et al. Global, regional, and national causes of child mortality in 2000–13, with projections to inform post-2015 priorities: an updated systematic analysis. Lancet. 2015;385(9966):430–40.
2. Drijkoningen JJC, Rohde GGU. Pneumococcal infection in adults: burden of disease. Clin Microbiol Infect. 2014;20:45–51.
3. Mehr S, Wood N. Streptococcus pneumoniae – a review of carriage, infection, serotype replacement and vaccination. Paediatr Respir Rev. 2012;13(4):258–64.
4. WHO | Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics. WHO. 2017;
5. Huang SS, Platt R, Rifas-Shiman SL, Pelton SI, Goldmann D, Finkelstein JA. Post-PCV7 Changes in Colonizing Pneumococcal Serotypes in 16 Massachusetts Communities, 2001 and 2004. Pediatrics. 2005;116(3):e408–13.
6. Singleton RJ, Hennessy TW, Bulkow LR, Hammitt LL, Zulz T, Hurlburt DA, et al. Invasive Pneumococcal Disease Caused by Nonvaccine Serotypes Among Alaska Native Children With High Levels of 7-Valent Pneumococcal Conjugate Vaccine Coverage. JAMA. 2007;297(16):1784.
7. Godfrey DI, Uldrich AP, McCluskey J, Rossjohn J, Moody DB. The burgeoning family of unconventional T cells. Nat Immunol. 2015;16(11):1114–23.
8. Treiner E, Duban L, Bahram S, Radosavljevic M, Wanner V, Tilloy F, et al. Selection of evolutionarily conserved mucosal-associated invariant T cells by MR1. Nature. 2003;422(6928):164–9.
9. Dusseaux M, Martin E, Serriari N, Péguillet I, Premel V, Louis D, et al. Human MAIT cells are xenobiotic-resistant, tissue-targeted, CD161hi IL-17-secreting T cells. Blood. 2011;117(4):1250–9.
10. Kjer-Nielsen L, Patel O, Corbett AJ, Le Nours J, Meehan B, Liu L, et al. MR1 presents microbial vitamin B metabolites to MAIT cells. Nature. 2012;491(7426):717–23.
11. Gherardin NA, Keller AN, Woolley RE, Le Nours J, Ritchie DS, Neeson PJ, et al. Diversity of T Cells Restricted by the MHC Class I-Related Molecule MR1 Facilitates Differential Antigen Recognition. Immunity. 2016;44(1):32–45.
12. Corbett AJ, Eckle SBG, Birkinshaw RW, Liu L, Patel O, Mahony J, et al. T-cell activation by transitory neo-antigens derived from distinct microbial pathways. Nature. 2014;509(7500):361–5.
13. Eckle SBG, Birkinshaw RW, Kostenko L, Corbett AJ, McWilliam HEG, Reantragoon R, et al. A molecular basis underpinning the T cell receptor heterogeneity of mucosal-associated invariant T cells. J Exp Med. 2014;211(8):1585–600.
14. Ussher JE, Bilton M, Attwod E, Shadwell J, Richardson R, Lara C De, et al. CD161++CD8+ T cells, including the MAIT cell subset, are specifically activated by IL-12+IL-18 in a TCR-independent manner. Eur J Immunol. 2014;44(1):195–203.
15. van Wilgenburg B, Scherwitzl I, Hutchinson EC, Leng T, Kurioka A, Kulicke C,

et al. MAIT cells are activated during human viral infections. Nat Commun. 2016;7:11653.

16. Le Bourhis L, Martin E, Péguillet I, Guihot A, Froux N, Coré M, et al. Antimicrobial activity of mucosal-associated invariant T cells. Nat Immunol. 2010;11(8):701–8.
17. Georgel P, Radosavljevic M, Macquin C, Bahram S. The non-conventional MHC class I MR1 molecule controls infection by Klebsiella pneumoniae in mice. Mol Immunol. 2011;48(5):769–75.
18. Meierovics A, Yankelevich WC, Cowley SC. MAIT cells are critical for optimal mucosal immune responses during in vivo pulmonary bacterial infection. Proc Natl Acad Sci U S A. 2013;110(33):E3119–28.
19. Chua W-J, Truscott SM, Eickhoff CS, Blazevic A, Hoft DF, Hansen TH. Polyclonal mucosa-associated invariant T cells have unique innate functions in bacterial infection. Infect Immun. 2012;80(9):3256–67.
20. Meierovics AI, Cowley SC. MAIT cells promote inflammatory monocyte differentiation into dendritic cells during pulmonary intracellular infection. J Exp Med. 2016;
21. Broug-Holub E, Toews GB, van Iwaarden JF, Strieter RM, Kunkel SL, Paine R, et al. Alveolar macrophages are required for protective pulmonary defenses in murine Klebsiella pneumonia: elimination of alveolar macrophages increases neutrophil recruitment but decreases bacterial clearance and survival. Infect Immun. 1997;65(4):1139–46.
22. Kooguchi K, Hashimoto S, Kobayashi A, Kitamura Y, Kudoh I, Wiener-Kronish J, et al. Role of alveolar macrophages in initiation and regulation of inflammation in Pseudomonas aeruginosa pneumonia. Infect Immun. 1998;66(7):3164–9.
23. Kurioka A, Ussher JE, Cosgrove C, Clough C, Fergusson JR, Smith K, et al. MAIT cells are licensed through granzyme exchange to kill bacterially sensitized targets. Mucosal Immunol. 2015;8(2):429–40.
24. Werno AM, Christner M, Anderson TP, Murdoch DR. Differentiation of Streptococcus pneumoniae from nonpneumococcal streptococci of the Streptococcus mitis group by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2012;50(9):2863–7.
25. Liuzzi AR, Kift-Morgan A, Lopez-Anton M, Friberg IM, Zhang J, Brook AC, et al. Unconventional Human T Cells Accumulate at the Site of Infection in Response to Microbial Ligands and Induce Local Tissue Remodeling. J Immunol. 2016;197(6):2195–207.
26. Kamphorst AO, Guermonprez P, Dudziak D, Nussenzweig MC. Route of Antigen Uptake Differentially Impacts Presentation by Dendritic Cells and Activated Monocytes. J Immunol. 2010;185(6):3426–35.
27. Samstein M, Schreiber HA, Leiner IM, Sušac B, Glickman MS, Pamer EG. Essential yet limited role for CCR2 $^{+}$ inflammatory monocytes during *Mycobacterium tuberculosis* -specific T cell priming. Elife. 2013;2:e01086.
28. Chen Z, Wang H, D'Souza C, Sun S, Kostenko L, Eckle SBG, et al. Mucosal-associated invariant T-cell activation and accumulation after in vivo infection depends on microbial riboflavin synthesis and co-stimulatory signals. Mucosal Immunol. 2016;doi: 10.1038/mi.2016.39.
29. Turtle CJ, Delrow J, Joslyn RC, Swanson HM, Basom R, Tabellini L, et al.

Innate signals overcome acquired TCR signaling pathway regulation and govern the fate of human CD161(hi) CD8α+ semi-invariant T cells. Blood. 2011;118(10):2752–62.

30. Slichter CK, McDavid A, Miller HW, Finak G, Seymour BJ, McNevin JP, et al. Distinct activation thresholds of human conventional and innate-like memory T cells. JCI insight. 2016;1(8).

31. Martner A, Skovbjerg S, Paton JC, Wold AE. Streptococcus pneumoniae autolysis prevents phagocytosis and production of phagocyte-activating cytokines. Infect Immun. 2009;77(9):3826–37.

32. Ussher JE, Bilton M, Attwod E, Shadwell J, Richardson R, Lara C De, et al. CD161 ++ CD8 + T cells , including the MAIT cell subset , are specifically activated by IL-12 + IL-18 in a TCR-independent manner. 2013;1–9.

33. Morens DM, Taubenberger JK, Fauci AS. Predominant Role of Bacterial Pneumonia as a Cause of Death in Pandemic Influenza: Implications for Pandemic Influenza Preparedness. J Infect Dis. 2008;198(7):962–70.

34. Loh L, Wang Z, Sant S, Koutsakos M, Jegaskanda S, Corbett AJ, et al. Human mucosal-associated invariant T cells contribute to antiviral influenza immunity via IL-18-dependent activation. Proc Natl Acad Sci U S A. 2016;113(36):10133–8.

35. Li W, Moltedo B, Moran TM. Type I interferon induction during influenza virus infection increases susceptibility to secondary Streptococcus pneumoniae infection by negative regulation of γδ T cells. J Virol. 2012;86(22):12304–12.

36. Walker LJ, Kang YH, Smith MO, Tharmalingham H, Ramamurthy N, Fleming VM, et al. Human MAIT and CD8αα cells develop from a pool of type-17 precommitted CD8 + T cells. Blood. 2012;119(2):422–33.

37. Koay H-F, Gherardin NA, Enders A, Loh L, Mackay LK, Almeida CF, et al. A three-stage intrathymic development pathway for the mucosal-associated invariant T cell lineage. Nat Immunol. 2016;17(11):1300–11.

38. Novak J, Dobrovolny J, Novakova L, Kozak T. The decrease in number and change in phenotype of mucosal-associated invariant T cells in the elderly and differences in men and women of reproductive age. Scand J Immunol. 2014;80(4):271–5.

39. Smith NM, Wasserman GA, Coleman FT, Hilliard KL, Yamamoto K, Lipsitz E, et al. Regionally compartmentalized resident memory T cells mediate naturally acquired protection against pneumococcal pneumonia. Mucosal Immunol. 2017;

40. Liu S, Sela S, Cohen G, Jadoun J, Cheung A, Ofek I. Insertional inactivation of Streptolysin S expression is associated with altered riboflavin metabolism inStreptococcus pyogenes. Microb Pathog. 1997;22(4):227–34.

41. Chen J, Shen J, Solem C, Jensen PR. Oxidative stress at high temperatures in Lactococcus lactis due to an insufficient supply of Riboflavin. Appl Environ Microbiol. 2013;79(19):6140–7.

42. Gibbs A, Leeansyah E, Introini A, Paquin-Proulx D, Hasselrot K, Andersson E, et al. MAIT cells reside in the female genital mucosa and are biased towards IL-17 and IL-22 production in response to bacterial stimulation. Mucosal Immunol. 2016;doi: 10.1038/mi.2016.30.

43. Martins ER, Pessanha MA, Ramirez M, Melo-Cristino J, Portuguese Group for the Study of Streptococcal Infections and the PG for the S of S. Analysis of

group B streptococcal isolates from infants and pregnant women in Portugal revealing two lineages with enhanced invasiveness. J Clin Microbiol. 2007;45(10):3224–9.

44. McGee L, McDougal L, Zhou J, Spratt BG, Tenover FC, George R, et al. Nomenclature of major antimicrobial-resistant clones of Streptococcus pneumoniae defined by the pneumococcal molecular epidemiology network. J Clin Microbiol. 2001;39(7):2565–71.
45. PMEN :: Pneumococcal Molecular Epidemiology Network [Internet]. Available from: http://www.pneumogen.net/pmen/
46. MLST allelic profiles, pbp gene profiles and macrolide resistance determinants for clones [Internet]. Available from: http://web1.sph.emory.edu/PMEN/pmen_table2.html
47. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics. 2010;11(1):595.
48. Brueggemann AB, Harrold CL, Rezaei Javan R, van Tonder AJ, McDonnell AJ, Edwards BA. Pneumococcal prophages are diverse, but not without structure or history. Sci Rep. 2017;7:42976.
49. Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18(5):821–9.
50. Boetzer M, Pirovano W. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. BMC Bioinformatics. 2014;15(1):211.
51. Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. Genome Biol. 2012;13(6):R56.
52. Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, et al. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. Microbiology. 2012;158(Pt 4):1005–15.
53. Francisco AP, Vaz C, Monteiro PT, Melo-Cristino J, Ramirez M, Carrio JA. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. BMC Bioinformatics. 2012;13(1):87.
54. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 2014;42:D206-14.
55. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068–9.
56. European Nucleotide Archive [Internet]. Available from: http://www.ebi.ac.uk/ena
57. Streptococcus pneumoniae MLST website [Internet]. Available from: http://pubmlst.org/spneumoniae/
58. GenBank [Internet]. Available from: https://www.ncbi.nlm.nih.gov/genbank/
59. Keith Jolley. Ribosomal Multilocus Sequence Typing (rMLST) [Internet]. Available from: https://pubmlst.org/rmlst/
60. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.
61. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11(10):R106.
62. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 1999;27(1):29–34.

952 63. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al.
953 Geneious Basic: An integrated and extendable desktop software platform for
954 the organization and analysis of sequence data. Bioinformatics.
955 2012;28(12):1647–9.
956 64. Pond SLK, Frost SDW, Muse S V. HyPhy: hypothesis testing using
957 phylogenies. Bioinformatics. 2005;21(5):676–9.
958 65. Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL. Datamonkey 2010: a
959 suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics.
960 2010;26(19):2455–7.
961 66. Pond SLK, Frost SDW. Datamonkey: rapid detection of selective pressure on
962 individual sites of codon alignments. Bioinformatics. 2005;21(10):2531–3.
963 67. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-
964 Scott C, et al. CDD: a Conserved Domain Database for the functional
965 annotation of proteins. Nucleic Acids Res. 2011;39(Database):D225–9.
966 68. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets
967 of protein or nucleotide sequences. Bioinformatics. 2006;22(13):1658–9.
968 69. Arnaoudova E, Jaromczyk JW, Moore N, Schardl CL, Yoshida R. Phylotree - a
969 toolkit for computing experiments with distance-based methods for genome
970 coevolution. BMC Bioinformatics. 2010;11(Suppl 4):P6.
971 70. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display
972 and annotation of phylogenetic and other trees. Nucleic Acids Res.
973 2016;44(W1):W242–5.
974 71. Inkscape [Internet]. Available from: https://inkscape.org/en/
975
976

977 ## Supporting information captions

978 Table S1. Descriptive data for the 571 *Streptococcus pneumoniae* genomes included

979 in this study.

980 Table S2. Descriptive data for the 824 non-pneumococcal *Streptococcus* species

981 genomes included in this study.

982

Table 1. Description of the four riboflavin operon gene sequences within 571 pneumococcal genomes.

Gene	Present (%)[a]	Nucleotide % pairwise identity	Amino acid % pairwise identity	Mean dN/dS	Gene annotation
ribD	559 (98.2%)	99.6%	99.8%	0.36	Diaminohydroxyphosphoribosylamino-pyrimidine deaminase
ribE	561 (98.2%)	99.6%	99.9%	0.28	riboflavin synthase
ribA	561 (98.2%)	99.7%	99.9%	0.42	3,4-dihydroxy-2-butanone-4-phosphate synthase
ribH	559 (98.2%)	99.5%	99.1%	0.12	6,7-dimethyl-8-ribityllumazine synthase

a. Two genomes possessed sequence assembly gaps in *ribD* and *ribH,* and 10 genomes were missing all four of the riboflavin genes (see Results).

Figure 1

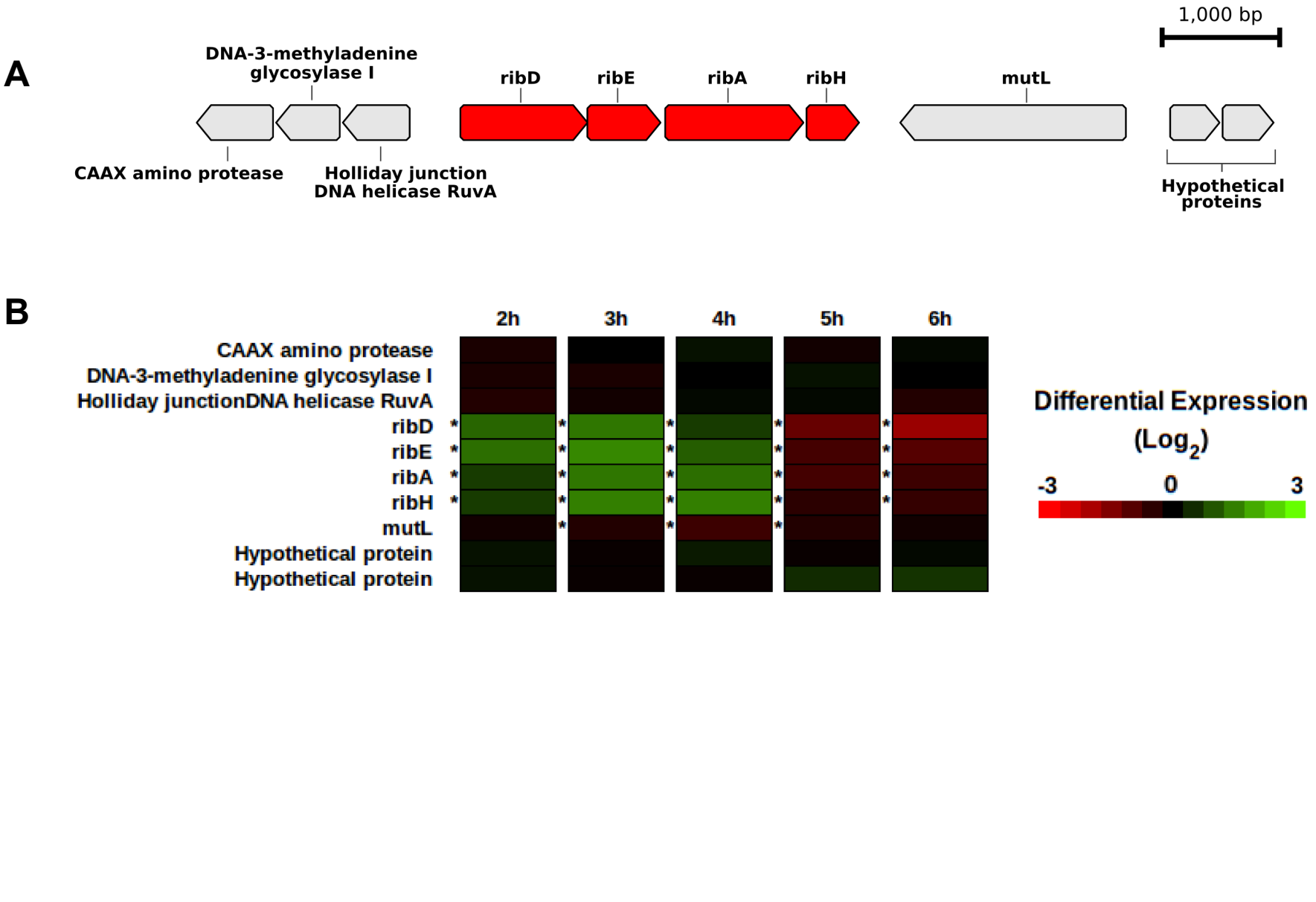

A
1,000 bp
DNA-3-methyladenine glycosylase I
ribD
ribE
ribA
ribH
mutL
CAAX amino protease
Holliday junction DNA helicase RuvA
Hypothetical proteins
B
2h
3h
4h
5h
6h
CAAX amino protease
DNA-3-methyladenine glycosylase I
Holliday junctionDNA helicase RuvA
ribD
ribE
ribA
ribH
mutL
Hypothetical protein
Hypothetical protein
Differential Expression
(Log2)
-3
0
3

Figure 2

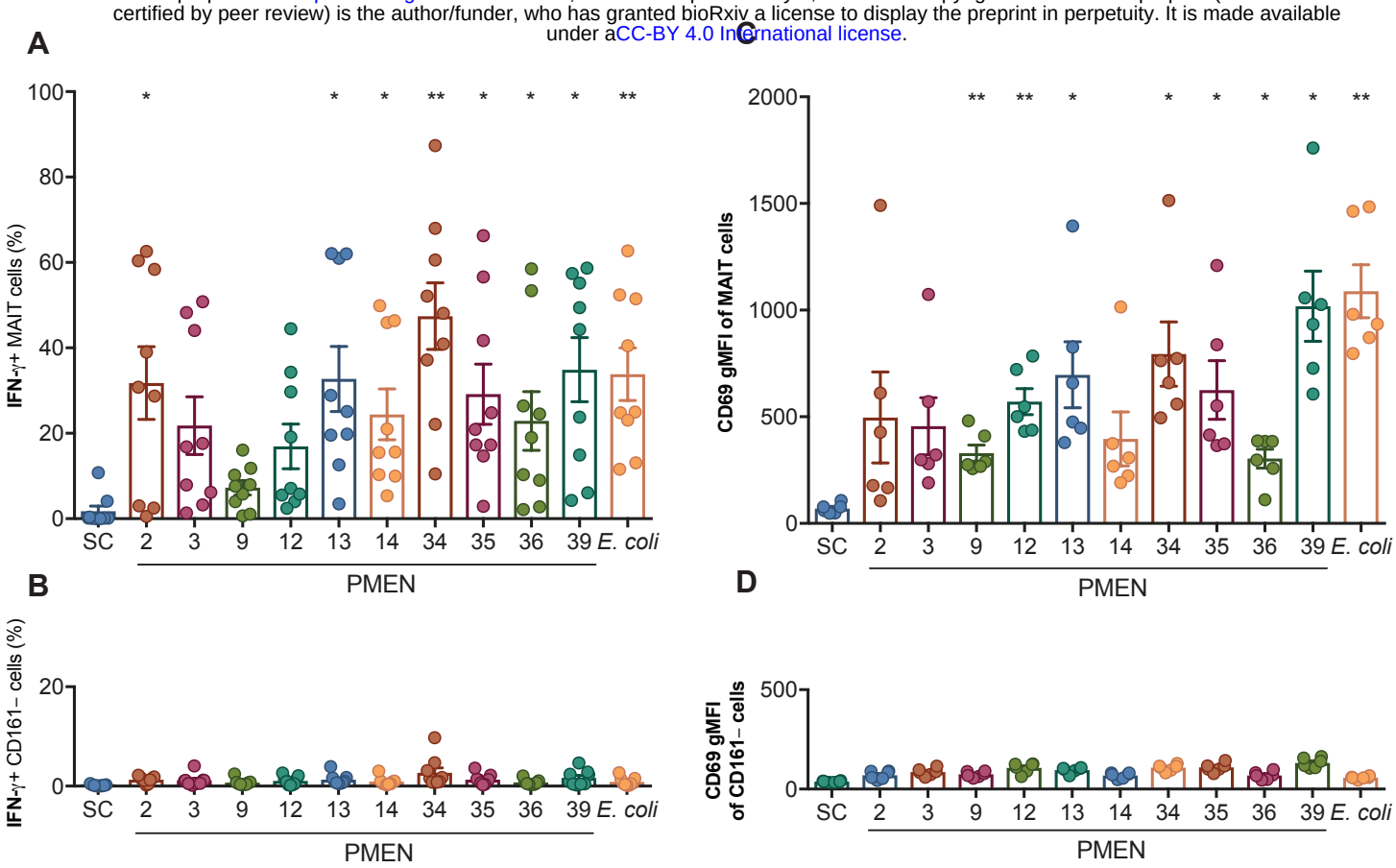


Figure 3

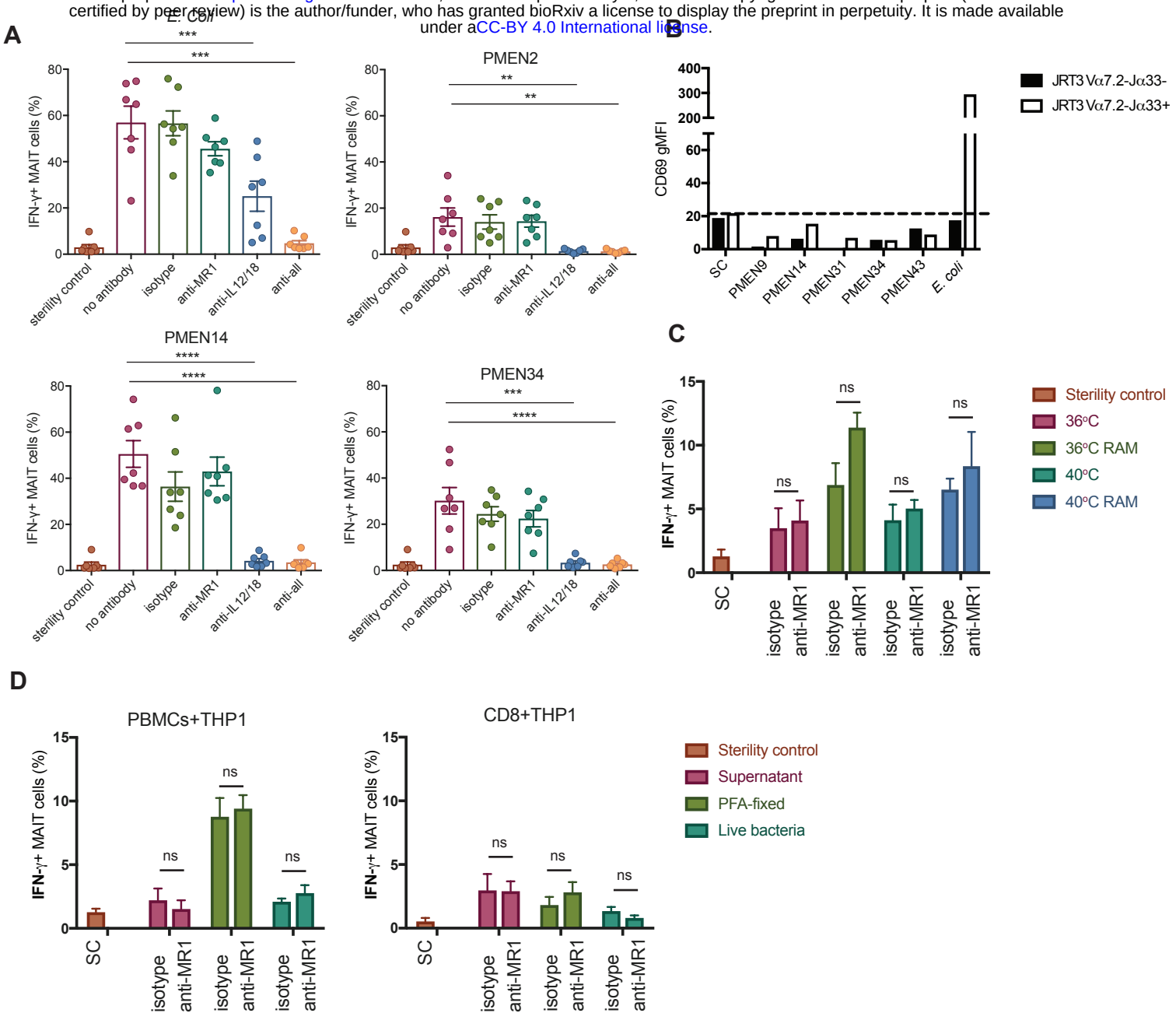
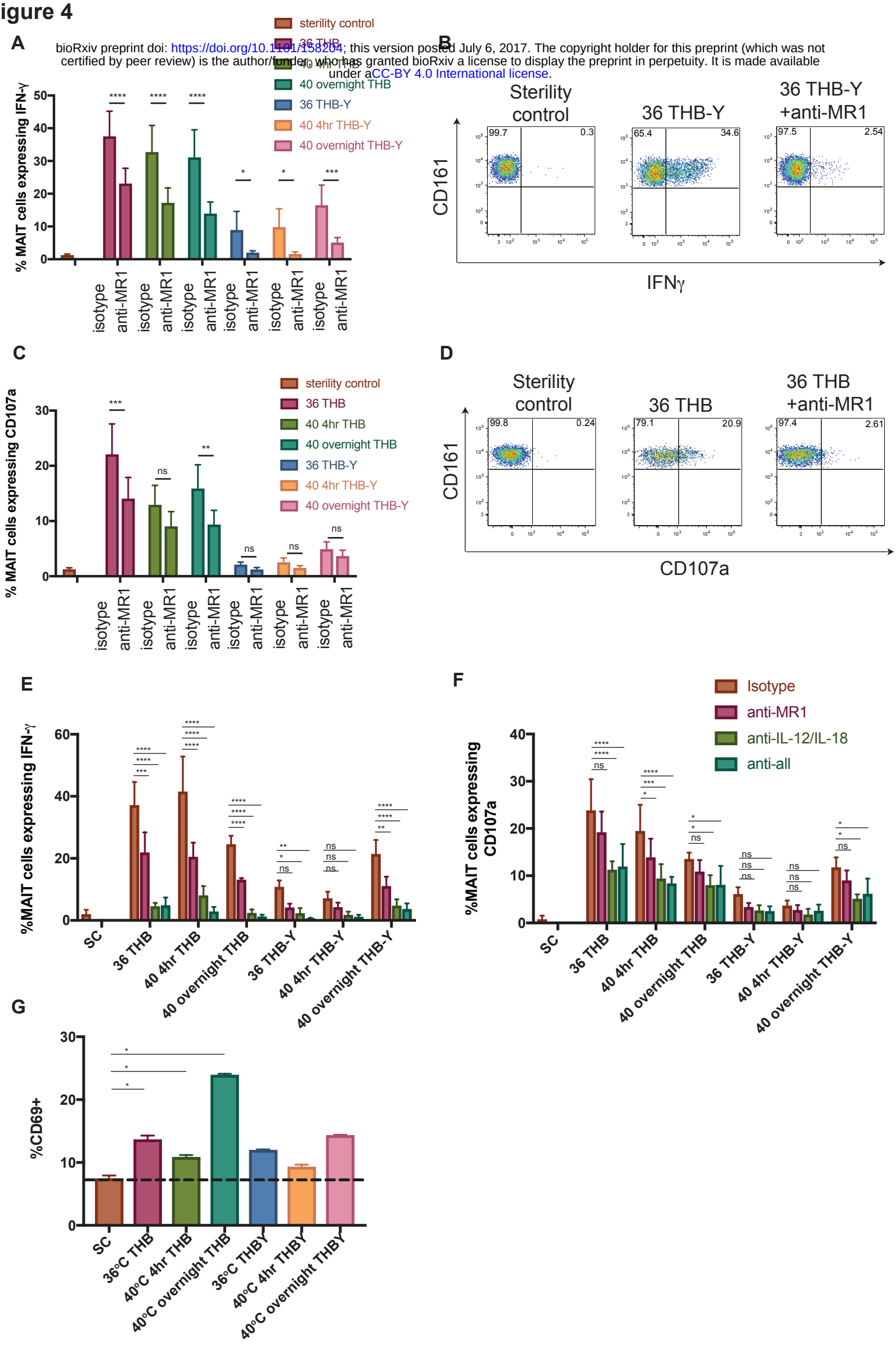
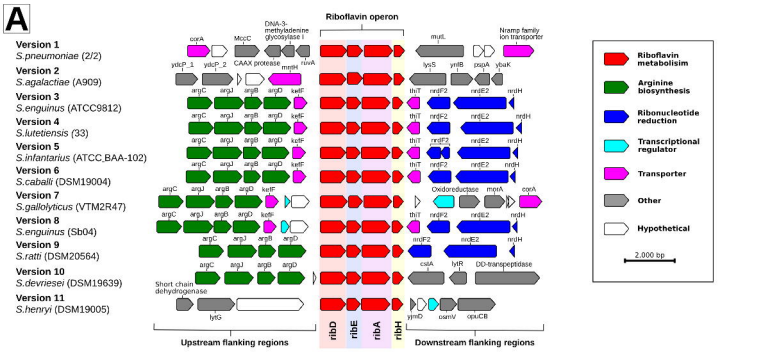


Figure 4





B

	Version 1	Version 2	Version 3	Version 4	Version 5	Version 6	Version 7	Version 8	Version 9	Version 10	Version 11
Version 1		60%	52%	52%	53%	53%	53%	53%	51%	49%	52%
Version 2	60%		53%	53%	53%	52%	53%	53%	52%	50%	53%
Version 3	52%	53%		88%	81%	80%	80%	80%	63%	61%	65%
Version 4	52%	53%	88%		88%	81%	80%	81%	63%	61%	65%
Version 5	53%	53%	81%	88%		81%	81%	83%	64%	61%	65%
Version 6	53%	52%	80%	81%	81%		86%	86%	63%	61%	65%
Version 7	53%	53%	80%	80%	81%	88%		83%	63%	61%	65%
Version 8	53%	53%	80%	81%	83%	86%	83%		64%	62%	66%
Version 9	51%	52%	63%	63%	64%	63%	63%	64%		72%	62%
Version 10	49%	50%	61%	61%	61%	61%	61%	62%	72%		59%
Version 11	52%	53%	65%	65%	65%	65%	65%	66%	62%	59%	

C

Species	Number of genomes in the dataset	Number of genomes with a riboflavin operon	Percentage of genomes with a riboflavin operon
S.pseudopneumoniae	16	16	100%
S.mitis	50	2	4%
S.oralis	50	2	4%
S.agalactiae	50	50	100%
S.equinus	27	27	100%
S.lutetiensis	2	2	100%
S.infantarius	2	2	100%
S.gallolyticus	17	17	100%
S.pasteurianus	5	5	100%
S.henryi	2	2	100%
S.caballi	1	1	100%
S.ratti	1	1	100%
S.devriesei	1	1	100%