Year : 2021

Harnessing CRISPR-Cas9 for use in the opportunistic human pathogen Streptococcus pneumoniae

Synefiaridou Dimitra

Synefiaridou Dimitra, 2021, Harnessing CRISPR-Cas9 for use in the opportunistic human pathogen Streptococcus pneumoniae

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive http://serval.unil.ch
Document URN : urn:nbn:ch:serval-BIB_B23C3169CE8F2

UNIL | Université de Lausanne
Faculté de biologie et de médecine

Department of Fundamental Microbiology

Harnessing CRISPR-Cas9 for use in the opportunistic human pathogen *Streptococcus pneumoniae*

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Dimitra SYNEFIARIDOU

Biologiste diplômé ou Master de l'Université de Groningen

Jury

Prof. Yves Poirier Yves Poirier, Président
Prof. Jan-Willem Veening, Directeur de thèse
Prof. Tim van Opijnen, Expert
Prof. Yolanda Schärli, Experte

Lausanne
(2021)

UNIL | Université de Lausanne
Faculté de biologie et de médecine

Ecole Doctorale
Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président·e	Monsieur	Prof.	Yves	**Poirier**
Directeur·trice de thèse	Monsieur	Prof.	Jan-Willem	**Veening**
Expert·e·s	Madame	Prof.	Yolanda	**Schaerli**
	Monsieur	Prof.	Tim	**van Opijnen**

le Conseil de Faculté autorise l'impression de la thèse de

Madame Dimitra Synefiaridou

Master in science, University of Groningen, Pays-Bas

intitulée

Harnessing CRISPR-Cas9 for use in the opportunistic human pathogen *Streptococcus pneumoniae*

Date de l'examen : 19 mars 2021

Date d'émission de l'imprimatur : Lausanne, le 13 avril 2021

pour le Doyen
de la Faculté de biologie et de médecine

Prof. Niko GELDNER
Directeur de l'Ecole Doctorale

Table of Contents

Academic Summary

CRISPR–Cas systems are part of a diverse group of RNA-guided nucleases that naturally defend prokaryotes against viral invaders. Currently, editing tools based on programmable CRISPR-associated nucleases, have been widely adapted for an increasing number of organisms. *Streptococcus pneumoniae,* an important human pathogen and a valuable model to study the cell biology of ovoid-shaped bacteria, is one of the first organisms in which RNA-guided editing by CRISPR-Cas9 system was used to introduce precise mutations in its genome. Additionally, CRISPRi studies, using nuclease-inactive versions of Cas9, provided high-throughput interrogation of genome-wide gene functions and genetic interactions, facilitating the study of pneumococcal essential genes and pathogenicity. The research presented in this thesis aims at paving the way for CRISPR-Cas applications in *S. pneumoniae.*

In Chapter 2, we demonstrate a newly established CRISPR-Cas9-sgRNA system, integrated in the pneumococcal genome, which can successfully introduce double-stranded breaks (DSBs). Our first approach was to establish the Non-Homologous End Joining (NHEJ) system in pneumococcus and exploit it, alongside with CRISPR-Cas9, to remove genes or large genomic regions. Two sgRNAs, targeting on both sides of the deletion target, would guide Cas9 to introduce the DSBs. Subsequently, the NHEJ mechanism would facilitate the repair by re-ligating the two ends, leaving out the deletion target. This way we could have markerless deletions with only requirement, every time, two specific sgRNAs. Most prokaryotic cells, including *Streptococcus pneumoniae*, lack a NHEJ system. Therefore, we transplanted four different NHEJ machineries from *Bacillus subtilis, Mycobacterium tuberculosis, M. smegmatis,* and *Streptomyces ambofaciens* to *S. pneumoniae*. Although we nearly exhausted the possible approaches, our efforts were not fruitful. Our results show that none of these NHEJ systems can facilitate repair from the DSB induced either by Cas9, or by the meganuclease I-SceI. Similarly, engineered ligases p50 and sso7D do not repair the DSB by Cas9. Additionally, we failed to facilitate a repair from a non-enzymatic DSB, using Mitomycin C (MMC). Finally, we discuss potential strategies to tackle this issue and achieve the setup of a successful NHEJ system.

After excluding the NHEJ as a viable option for genome engineering in the pneumococcus, we explored the idea of using CRISPR-Cas9 editing with HR (homologous recombination) templates to repair the DSB. In Chapter 3, we transferred the CRISPR-Cas9-sgRNA system to a not integrated pneumococcal plasmid. The plasmid that we designed has the temperature sensitive origin of replication *pG$^+$host*, facilitating the elimination of the plasmid from the strains, upon the desired deletion. This feature allowed us to successfully perform consecutive deletions, since we are ultimately interested to remove multiple genes and chromosomal

regions from the genome. Additionally, we demonstrated the advantages by using our system, compared to just performing natural transformation without counterselection, which would be an alternative for clean deletions. Finally, using our plasmid, we created a new a strain with three important virulence factors removed, proving that our CRISPR-Ca9 system has shown robustness and flexibility in genetic manipulation of the bacterial genome. This final strain was confirmed with Whole Genome Sequencing, eliminating also a major concern of the CRISPR-Cas9 system, the off target DSB. Finally, our plasmid is also promising to be applicable to other Gram-positive bacteria as well, with minor modifications, like exchanging the specific to pneumococci zinc-inducible promoter of Cas9, by an alternative inducible promoter for the species of interest.

In Chapter 4, we used dCas9, this time not for genome editing, but to study the cellular processes of competence for transformation and fratricide, demonstrating at the same time the versatility of the CRISPR-Cas technologies. Performing a genome-wide CRISPRi-seq targeting 1499 TSS, we evaluated the hits that were underrepresented when both the library and competence were induced, expecting that the bacteria undergoing transcriptional downregulation of genes important for immunity against fratricins would be outcompeted or lysed by fratricins produced by bacteria in which the sgRNA targets a neutral gene. After rejecting false positives, or confirming the hits, we started investigating in detail these genes, with the majority of them being involved in TA synthesis pathways. Specifically, we demonstrate that the levels, or relative amounts, of LTA and WTA are altered when competence is triggered, proposing a model in which competence activation during infection is crucial to transfer important virulence factors to the outside of the pneumococcal cell to ensure better adherence or immune evasion. More functional assays to support our findings are currently undergoing. This chapter showed the possibilities of CRISPR interference as otherwise essential genes (such as many TA genes) can be probed for additional roles (in this case their role in protecting from autolysins). We anticipate that CRISPR-cas9 based genome-wide screens will be more and more applied in microbial research, similarly to the way transposon-insertion sequencing studies have transformed genetic screening.

Résumé Académique

Les systèmes CRISPR-Cas font partie d'un groupe varié de nucléases guidées par ARNs et sont naturellement utilisés par les procaryotes pour se défendre contre les envahisseurs viraux. Ces dernières années, des outils d'édition basés sur des nucléases associées à CRISPR ont été développés pour un nombre croissant d'organismes. *Streptococcus pneumoniae* est un important pathogène pour l'Homme et un modèle de choix pour étudier la biologie cellulaire des bactéries de forme ovoïde. Il a, de plus, été l'un des premiers organismes pour lequel l'édition guidée par ARN utilisant le système CRISPR-Cas9 a été utilisée dans le but d'introduire des mutations précises dans son génome. De plus, des études utilisant des versions de Cas9 inactivées pour leur activité de nucléase ont participé à élucider la fonction de certains gènes à l'échelle du génome et à établir des interaction génétiques, facilitant l'étude de gènes essentiels chez le pneumocoque et sa pathogénicité. La recherche présentée dans cette thèse vise à ouvrir la voie à de nouvelles applications de CRISPR-Cas chez *S. pneumoniae*.

Dans le chapitre 2, nous présentons un nouveau système CRISPR-Cas9-sgRNA intégré dans le génome du pneumocoque, capable d'introduire des cassures double brin (DSB) dans l'ADN. Notre première approche a été d'établir le système de jonction d'extrémités non homologues (NHEJ) dans le pneumocoque et de l'utiliser avec CRISPR-Cas9 pour éliminer des gènes ou de grandes régions génomiques. Pour ce faire, deux sgRNAs, ciblant les deux côtés d'une zone à supprimer, peuvent être utilisés pour guider Cas9 et ainsi introduire les cassures double brin. Par la suite, le mécanisme NHEJ peut faciliter la réparation en reliant les deux extrémités, excluant de cette manière la zone d'intérêt. Il est ainsi possible d'obtenir des délétions sans marqueur de résistance en utilisant seulement à chaque fois deux sgRNAs spécifiques. La plupart des procaryotes, y compris *Streptococcus pneumoniae*, n'ont pas de système NHEJ. Par conséquent, nous avons transplanté à *S. pneumoniae* quatre différentes machineries NHEJ provenant de *Bacillus subtilis, Mycobacterium tuberculosis, M. smegmatis* et *Streptomyces ambofaciens*. Bien que nous ayons essayé presque toutes les approches possibles, nos efforts n'ont pas été fructueux. Nos résultats indiquent qu'aucun des systèmes NHEJ ne peut faciliter la réparation du DSB induit par Cas9 ou par la méganucléase I-SceI. De manière similaire, les ligases p50 et sso7D n'étaient pas capable de réparer les DSB par Cas9. De plus, nous n'avons pas réussi à faciliter la réparation de DSB non enzymatique tel que lors de l'utilisation de mitomycine C (MMC). Enfin, nous envisageons les stratégies potentielles qui peuvent aider à résoudre ce problème et à mettre en place un système NHEJ fonctionnel.

Après avoir exclu le NHEJ comme option envisageable pour manipuler le génome du pneumocoque, nous avons exploré l'idée d'utiliser le CRISPR-Cas9 avec des modèles de

recombinaison homologue (HR) pour réparer les DSB. Au chapitre 3, nous avons transféré le système CRISPR-Cas9-sgRNA à un plasmide du pneumocoque incapable de s'intégrer dans le génome. Le plasmide que nous avons conçu possède une origine de réplication *pG*$^{+}$*host* sensible à la température, ce qui facilite l'élimination du plasmide lors des délétions souhaitées. Cette fonctionnalité nous a permis d'effectuer avec succès des délétions consécutives, notre objectif étant de supprimer plusieurs gènes et régions chromosomiques du génome. De plus, nous avons démontré les avantages de l'utilisation de notre système par rapport à une transformation naturelle sans contre-sélection, qui pourrait être une alternative intéressante aux délétions classiques. Enfin, en utilisant notre plasmide, nous avons pu créer une souche dans laquelle trois des importants facteurs de virulence ont été supprimés, indiquant que notre système CRISPR-Ca9 peut être robuste et flexible pour manipuler génétiquement les génomes bactériens. Cette souche dernièrement créée a été confirmée par séquençage complet de son génome, ce qui a également permis d'exclure la présence de DSB aspécifique lors de l'utilisation de système CRISPR-Cas9. Enfin, notre plasmide pourrait potentiellement être applicable chez d'autres bactéries à Gram-positif, en apportant quelques modifications mineures, comme échanger le promoteur de Cas9 inductible au zinc, qui est spécifique au pneumocoque, par un promoteur inductible alternatif en fonction de l'espèce d'intérêt.

Au chapitre 4, nous avons utilisé dCas9, cette fois ci non pas éditer le génome, mais pour étudier les processus cellulaires de compétence et de fratricide, démontrant en même temps la polyvalence de la technologie CRISPR-Cas. En effectuant une interférence par CRISPR couplée à un séquençage haut débit (CRISPRi-seq) ciblant 1499 sites d'initiation de la transcription (TSS) lorsque la compétence pour transformation était induite, nous avons pu évaluer les gènes qui étaient sous-représentés. Nous avons étudié en détails les gènes importants pour l'immunité contre les fratricines et espéré que ceux présentant une baisse de régulation transcriptionnelle impacteraient la capacité des cellules à résister aux fratricines produites par d'autres bactéries pour lesquelles le sgRNA ciblerait un gène neutre. Ces cellules seraient donc lysées ou surpassées. Après avoir rejeté les faux positifs et confirmé les résultats, nous avons trouvé que la plupart de ces gènes étaient impliqués dans les voies de biosynthèse des acides téichoïques (TA). Plus précisément, nous avons démontré que les niveaux, ou quantités relatives, de LTA et WTA étaient modifiés lorsque la compétence était induite, en proposant un modèle où l'activation de la compétence lors d'infection est cruciale pour transférer des facteurs de virulence à l'extérieur de la cellule dans le but d'assurer une meilleure adhérence ou évasion immunitaire. De plus amples tests sont nécessaires pour étayer nos découvertes et sont actuellement en cours. Ce chapitre permet de démontrer les possibilités d'interférence par CRISPR pour des gènes autrement essentiels (tels que de

nombreux gènes TA) qui peuvent être étudier pour des rôles supplémentaires (dans ce cas, leur rôle dans la protection contre les autolysines). Nous anticipons que les criblages à l'échelle du génome basés sur CRISPR-cas9 seront de plus en plus appliqués à la recherche microbienne, de la même manière que les études de séquençage par insertion de transposons (Tn-seq) ont transformé le dépistage génétique.

Non-Academic Summary

The ability to make virtually any targeted change in the genome of any living cell or organism is a longstanding aspiration of the life sciences. Precise nucleic acid editing technologies are valuable for studying cell biology and potential novel therapeutics. In 2012, the proposition of programmable gene editing using CRISPR-Cas9, made this genome engineering ambition a reality. This not only caused enabled scientists to engineer higher eukaryotes, but CRISPR-Cas systems also enabled genetic research in non-model microorganisms such as the here studied bacterium *Streptococcus pneumoniae* (pneumococcus), which is an important human pathogen.

The CRISPR-Cas system is a tool for cutting DNA at a specifically targeted location. It was first discovered in a bacterial immune system, and, later, adapted into a powerful tool for genomic research. The system consists of two components, the DNA cutting protein, named Cas9, which acts as molecular scissors and introduces a double-stranded break (DSB) into the DNA at a specific location, defined by an RNA molecule, known as the guide RNA. After the DSB has been introduced, the cells have two options, in order to repair it and survive. They can either use a pathway that will join together the two broken ends, called Non-Homologous End Joining (NHEJ) pathway, or they can be provided with a new DNA fragment, identical (homologous) to the one that has the DSB, and use it as a patch to repair it, called Homologous Recombination (HR) pathway.

In this thesis we are interested in using the CRISPR-Cas9 technology at pneumococcus and apply it to perform genome editing, with the ultimate goal of genome minimization and uncover novel pneumococcal biology. To achieve this, we first established a CRISPR-Cas9 system, which performed with very high efficiency DSB and we explored the two options to harness it after the DSB; NHEJ and HR. Our first approach was to establish the NHEJ system in the pneumococcus and exploit it, alongside with CRISPR-Cas9, to remove genes or large genomic regions. Most prokaryotic cells, including *S. pneumoniae*, lack a NHEJ system. Therefore, we transplanted four different NHEJ machineries from bacteria, with such a system. Despite the fact that we nearly exhausted the possible approaches, our efforts were not fruitful. Next, we transferred the CRISPR-Cas9-sgRNA system to a pneumococcal plasmid and exploited the homology-repair (HR) pathway, that the pneumococcus naturally carries. Specifically, we demonstrate an approach for making targeted gene knockouts and large genome deletions. Finally, using this system, we engineered a strain with three major virulence factors deleted.

Additionally, modified versions of CRISPR-Cas9 have been used to explore fundamental cell biology questions. Performing a screen using a catalytically inactive Cas9, we identified

genes that become essential during competence, which is an important state during which cells can take up foreign (extracellular) DNA. We anticipate that CRISPR-cas9 based genome-wide screens will be more and more applied in microbial research.

Résumé Non Académique

La capacité d'effectuer un changement ciblé dans le génome de n'importe quelle cellule ou organisme vivant, est depuis longtemps un objectif important des sciences de la vie. Les technologies d'édition précises d'acide nucléique sont précieuses pour étudier la biologie cellulaire et ainsi que pour le développement de nouvelles thérapies potentielles. En 2012, la proposition d'édition de gène programmable à l'aide de CRISPR-Cas9, a fait de cette ambition d'ingénierie du génome une réalité. Cela a non seulement permis aux scientifiques de modifier des eucaryotes pluricellulaires, mais les systèmes CRISPR-Cas ont également permis la recherche génétique sur des micro-organismes non modèles tels que la bactérie Streptococcus pneumoniae (pneumocoque), faisant l'objet de cette étude, qui est un pathogène humain important.

Le système CRISPR-Cas est un outil pour couper l'ADN à un endroit spécifiquement ciblé. Il a premièrement été découvert dans un système immunitaire bactérien, puis adapté en un outil puissant pour la recherche génomique. Le système utilise deux composants, la protéine qui coupe l'ADN, appelée Cas9, agit comme des ciseaux moléculaires et introduit une cassure double brin (DSB) de l'ADN à un emplacement spécifique, défini par une molécule d'ARN, connue sous le nom d'ARN guide (sgRNA). Une fois le DSB introduit, les cellules ont deux options pour réparer cette cassure et ainsi survivre. Ils peuvent soit utiliser un mécanisme qui reliera les deux extrémités cassées, appelée voie de jonction d'extrémité non homologue (NHEJ), soit ils peuvent être fournis d'un nouveau fragment d'ADN, identique (homologue) au fragment possédant le DSB, et l'utiliser comme modèle pour réparer la cassure, effectuer ainsi une recombinaison homologue (HR).

Dans cette thèse, nous nous intéressons à l'utilisation de la technologie CRISPR-Cas9 dans le pneumocoque et à son application pour effectuer l'édition du génome, dans le but ultime de minimiser le génome et de découvrir une nouvelle biologie du pneumocoque. Pour y parvenir, nous avons d'abord mis en place un système CRISPR-Cas9, qui effectuait un DSB à très haut rendement et avons exploré les deux options après le DSB ; NHEJ et HR. Notre première approche a été d'établir le système NHEJ dans le pneumocoque et de l'exploiter, aux côtés de CRISPR-Cas9, pour éliminer des gènes ou de grandes régions génomiques. La plupart des cellules procaryotes, y compris *S. pneumoniae*, n'ont pas de système NHEJ. Par conséquent, nous avons essayé d'implémenter dans le pneumocoque quatre machineries NHEJ différentes issues d'autres bactéries possédant un tel système. Malgré nos efforts d'implémenter ces machineries dans le pneumocoque en utilisant différentes approches possibles, cela n'a pas été fructueux. Nous avons ensuite transféré le système CRISPR-Cas9-sgRNA dans un plasmide pneumococcique et exploité le mécanisme de recombinaison homologue (HR), que le pneumocoque possède naturellement. Plus précisément, nous

démontrons une approche pour faire des délétions de gènes ciblés et de grandes délétions dans le génome. Enfin, en utilisant ce système, nous avons conçu une souche avec trois facteurs de virulence majeurs supprimés.

De plus, des versions modifiées de CRISPR-Cas9 ont été utilisées pour explorer les questions fondamentales de biologie cellulaire. En effectuant un criblage génétique à l'aide d'une protéine Cas9 catalytiquement inactif, nous avons identifié des gènes qui deviennent essentiels au cours de la compétence, ce qui est un état important au cours duquel les cellules peuvent absorber de l'ADN étranger (extracellulaire). Nous prévoyons que les criblages à l'échelle du génome basés sur l'utilisation de CRISPR-Cas9 seront de plus en plus appliqués à la recherche microbienne.

Chapter 1: Thesis Introduction

(D.S. wrote this chapter)

A brief history of the pneumococcus

Streptococcus pneumoniae, also known as the pneumococcus, is an organism associated with studies that have been central to some of the most profoundly influential biological findings of the past 150 years. Isolated independently in 1881 by Sternberg in the United States (Sternberg, 1881) and Pasteur in France (Pasteur, 1881), it was described as lancet-shaped pairs of coccoid bacteria in human saliva, with pathogenic potential, and it was initially named *Diplococcus pneumoniae* in 1920, by the Society of American Microbiologists (Winslow et al., 1920). Finally, in 1974 it was classified under the genus *Streptococcus* and was given its present name, *Streptococcus pneumoniae* (Deibel and Seeley Jr, 1974).

Since its discovery, the pneumococcus has assisted in advancing biological research on various occasions. Among others, it is associated with the development of Gram's stain, a bacteriologic tool that is still in everyday use almost two centuries after its original description (Austrian, 1960), the Quellung reaction, providing a simple method of serotyping (Neufeld and Etinger-Tulczynska, 1933), and the identification of the pathogenic role of the pneumococcal capsule and its polysaccharide nature, the birth of quantitative immunology (Heidelberger, Aisenberg and Hassid, 1954; Austrian, 1999).

In 1928, the British bacteriologist Frederick Griffith, while trying to develop a vaccine against pneumonia, conducted a series of experiments using the bacterium, discovering natural transformation (Griffith, 1928). Following up Griffith's experiment, in 1944, Avery, MacLeod and McCartey, in their quest to identify this "transforming principle", demonstrated that DNA is the genetic material (Avery et al., 1944), a finding which is considered a milestone of Molecular Biology.

The pneumococcus

S. pneumoniae is a Gram-positive, rugby-ball shaped, typically considered extracellular, opportunistic pathogen. It has also been reported that it has an intracellular phase inside splenic macrophages that serves as a reservoir for septicaemia (Ercoli et al., 2018). In humans, it asymptomatically colonizes the mucosal surfaces of the upper respiratory tract (URT) (Kadioglu et al., 2008). However, in susceptible groups like children, the elderly and the immunocompromised, it can occasionally become pathogenic causing diseases which range from mild upper respiratory tract infections, such as acute otitis media and sinusitis, to severe and potentially life-threatening conditions such as pneumonia, bacteremia and meningitis

(Simell et al., 2012). It is responsible for more than one million deaths annually (O'Brien et al., 2009) and in 2017, the World Health Organization (WHO) classified *S. pneumoniae* as one of 12 priority pathogens for which new antibiotics are urgently needed.

The pneumococcus produces a range of colonization and virulence factors involved in the disease process. The molecular activity of some of these virulence factors has been elucidated. For instance, the capsule, probably the most important pneumococcal virulence factor, is expressed in a form of a biochemically distinct polysaccharide that is in most cases covalently attached to the cell wall (Weinberger et al., 2009). Its role is to prevent entrapment in the nasal mucus, thereby allowing access to epithelial surfaces and to inhibit effective opsonophagocytosis (Kadioglu et al., 2008). Another major virulence factor is *pneumolysin (Ply),* the single pneumococcal toxin which forms pores after membrane insertion. It is also responsible for reduced URT inflammation, shedding and the ability to transmit to littermates (Weiser et al. 2018).

Aim of the thesis

Although several key virulence factors are well studied, it still remains unknown how the majority of pneumococcal genes are involved with and contribute to virulence. The main goal of this thesis work is to establish tools for the first step towards a minimal pneumococcal genome**,** which could aid in assessing the minimal set of genes required to become pathogenic and to help identify crucial virulence factors and their mechanism of action. These tools also pave the path for a *S. pneumoniae* strain with all virulence factors removed, to finally determine the minimal requirements for the pneumococcus to become virulent and cause diseases. *S. pneumoniae* might represent an easier starting point for genome minimization compared to *E. coli* or *B. subtilis* as it has a much smaller genome size of ~2.1 Mbp (Lanie et al., 2007), is naturally competent for DNA uptake and integration, and many genetic tools have become available over the years.

The first aim of this thesis was to develop CRISPR-Cas9 based genetic engineering for *S. pneumoniae* (Chapters 2 and 3). CRISPR-Cas systems are facilitating genome editing in a whole range of organisms and such a system was not yet available for the pneumococcus. It was also not clear if non-homologous end joining could be engineered and harnessed for the pneumococcus (see below).

A second aim of this thesis was to better understand how competence is connected to virulence using a CRISPR-dCas9 based whole genome screening (Chapter 4). Virulence appears to be connected to competence; however, the nature of this connection is not fully elucidated. In the state of competence, we observe a phenomenon called 'pneumococcal fratricide' (Håvarstein et al., 2006), where cells synthesize killing/lytic factors, that are directed

against non-competent siblings, and release chromosomal DNA and the cytoplasmic virulence factor pneumolysin (Ply) (Guiral et al., 2006). How competence and its associated fratricide are controlled, and their role in virulence, remains unknown. This thesis aims to fill that knowledge gap.

To address these aims and questions, we need new, scalable tools. Here, we develop a simple tool to perform genome editing in *S. pneumoniae* utilizing CRISPR-Cas9 (Chapter 3), and demonstrate how using a CRISPRi-seq screen can reveal new insights in fratricide and virulence (Chapter 4). This thesis work now opens up new areas of research and the approaches documented here will aid other researchers in their projects.

CRISPR-Cas

A close relative of the pneumococcus, *Streptococcus pyogenes,* is the organism used to develop CRISPR-Cas9 genome editing technology (Jinek et al., 2012), which turned out to be another milestone of Biological Sciences, resulting in the Nobel of Chemistry in 2020. CRISPR-Cas stands for Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-Associated proteins (Bhaya et al., 2011). A timeline of key events of CRISPR research is displayed in Figure 1.

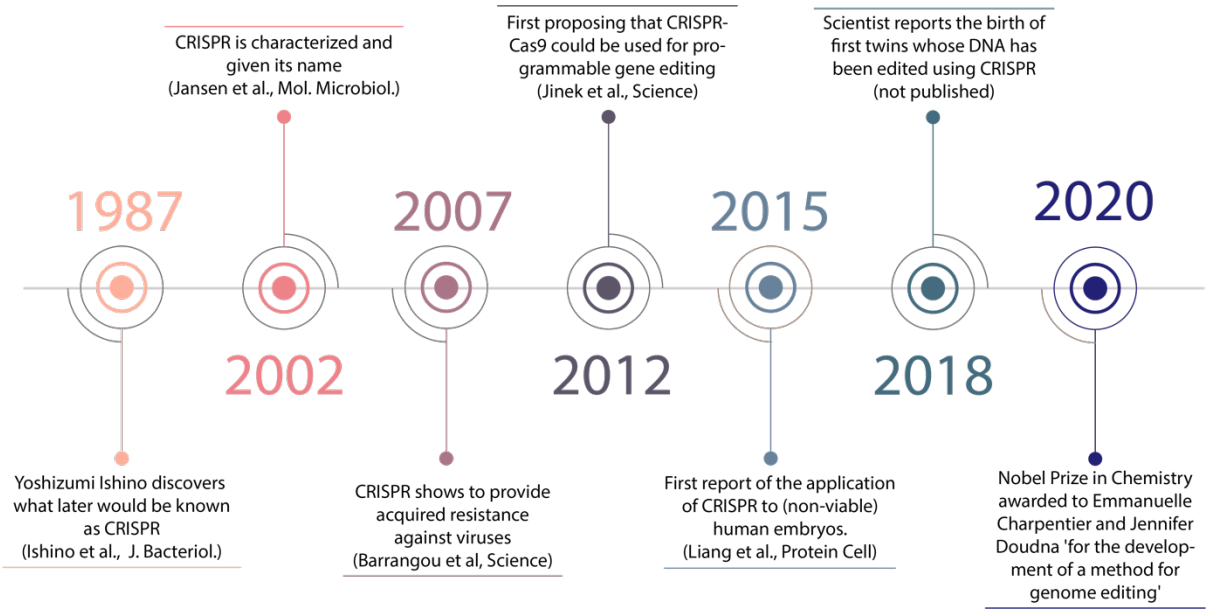


Figure 1: A timeline of milestones of CRISPR-related discoveries and events.

CRISPR-Cas systems are present in many bacteria and most archaea (Jansen et al., 2002). Naturally, the CRISPR-Cas systems provide resistance against foreign genetic elements (virus or plasmid) via small noncoding RNAs that are derived from CRISPR loci. CRISPR–Cas immunity involves three distinct mechanistic stages: adaptation, expression and

interference. These defense systems rely on specifically incorporating short sequences from invading genetic material into a region of its genome that is distinguished by CRISPRs. Then, after these sequences are transcribed and precisely processed into small RNAs, they guide a multifunctional protein complex to recognize and cleave incoming foreign genetic material, resulting in both the destruction of the invader's genome and immunity (Morimoto et al. 2009; Bhaya et al. 2011).

A robust classification of CRISPR–Cas systems based on their evolutionary relationships is essential for the progress of CRISPR research. Although this classification has been changing (Makarova *et al.*, 2011, 2015, 2020), the most abundant types and subtypes of CRISPR–Cas systems are now known, and the overall structure of the current classification is likely to stand the test of time. CRISPR-Cas systems are broadly divided into two classes, six types and 33 subtypes, according to the Cas proteins they carry and depending on the architecture of the interference module. Class 1 CRISPR–Cas systems, include types I, III, IV, and 17 subtypes and are defined by the presence of a multisubunit CRISPR RNAs (crRNA)–effector complex, while in class 2 CRISPR–Cas systems, which include types types II, V, VI and 16 subtypes, all functions of the effector complex are carried out by a single protein. This class includes the commonly used Cas9 system from *S. pyogenes* (Makarova et al. 2015; Koonin, et. al 2017; Shmakov et al. 2017; Makarova et al. 2020).

In class 2, type II CRISPR systems, the mature crRNA that is base-paired to a trans-activating crRNA (tracrRNA) forms a two-RNA structure that directs the CRISPR-associated proteins (e.g., Cas9) to introduce a double-stranded break (DSB) into the target DNA locus. Site-specific cleavage occurs at locations determined by both base-pairing complementarity between the crRNA and the target protospacer DNA, and a short protospacer adjacent motif (PAM) (Jinek et al., 2012). It has been demonstrated that the endonuclease can be programmed by engineering the mature dual-tracrRNA: crRNA as a single RNA chimera (sgRNA for single guide RNA), to cleave specific DNA sites. Thereby, modified versions of the system can be exploited as a biotechnological tool for precise, RNA-programmable genome targeting and editing (Jinek et al., 2012).

The development of new CRISPR–Cas genome editing tools continues to drive major advances in the life sciences. Four classes of CRISPR–Cas-derived genome editing agents; nucleases, base editors, transposases/recombinases and prime editors, are currently available for modifying genomes in experimental systems (Anzalone et al., 2020).

In principle, most applications of the CRISPR–Cas system can be divided into three categories; editing, knockout and recruitment experiments. The feature that all of these experimental interventions have in common is that the Cas protein is directed to a specific

genomic destination by the guide RNA (Hanna and Doench, 2020). In editing experiments, specific base changes are created, either by an exogenous template DNA delivered in parallel (Liang et al., 2017) or by modified Cas proteins that directly alter the endogenous DNA (Anzalone et al., 2019; Gaudelli et al., 2017; Komor et al., 2016). In knockout experiments, insertions or deletions (indels) are introduced from mismatch repair of double-strand DNA breaks, which either generate knockout alleles of protein-coding genes or disrupt transcription factor binding sites. Alternatively, Cas9-induced DNA breaks promote interchromosomal translocation for larger deletions or chromosomal rearrangements (Choi and Meyerson, 2014). Finally, in recruitment experiments, a Cas protein causes gene activation, transcriptional interference or chromatin modifications, by inactivating its nuclease domains and either directly or indirectly recruiting another protein to the specific genomic locus that it binds (Gilbert et al., 2014).

None of the available pneumococcal genomes harbor CRISPR loci and *S. pneumoniae* does not contain an endogenous CRISPR-Cas system (Grissa et al., 2007; Makarova et al., 2011a, 2011b). In *S. pneumoniae,* to date, CRISPR tools have been established either for editing experiments (Jiang et al. 2013; Synefiaridou and Veening 2021, Chapter 3) or for recruitment experiments (Bikard et al., 2012; Kjos, 2019; Liu et al., 2017). In editing experiments, the CRISPR-Cas9 system is introducing a DSB and is utilized for guided genome editing of the pneumococcal genome. In recruitment experiments, a nuclease inactive Cas9 (dCas9) is repurposed for CRISPR interference (CRISPRi). After binding in a sequence-specific target, it causes a steric block that halts transcript elongation by RNA polymerase, silencing the gene expression (Larson et al., 2013; Qi et al., 2013). Such studies have showed the power of CRISPRi for functional gene analysis. Additionally, pooled CRISPRi libraries can be used for genome-wide functional analysis (Cui et al., 2018; Jiang et al., 2015; Lee et al., 2019; Liu et al., 2021; Wang et al., 2016; de Wet et al., 2018). For instance, recently a pooled library targeting nearly all operons of the prototypic *S. pneumoniae* strain D39V has been created and used in combination with NGS (called CRISPRi-seq) to identify genes important for virulence in vivo (Liu et al., 2021). In Chapter 4 of this thesis, we have used CRISPRi-seq to identify the mechanism responsible for immunity to fratricins and found a direct link to TA biosynthesis.

DNA repair mechanisms

When the CRISPR-Cas9 system is being expressed, Cas9 is introducing a DSB at the specific DNA locus. After the DSB has been introduced by the Cas nuclease, the cell can utilize two major pathways in order to repair the break and survive -homologous recombination (HR) or non-homologous end-joining (NHEJ)- that differ in several fundamental manners. In HR, a second intact copy of the broken chromosome segment, homologous to the DSB site, serves as a template for DNA synthesis across the break. In this mechanism, the crucial process of locating and recombining the homologous sequence is performed by RecA (Shuman and Glickman, 2007).

NHEJ does not rely on a homologous DNA template, as the two DNA ends are rejoined directly together. In its simplest form it consists of a two-component system. Ku and a multifunctional ATP-dependent DNA ligase (LigD). Together, Ku and LigD possess all the break-recognition, end-processing, and ligation activities that repair DSBs. Ku recognizes the DSB and binds to the 3'-protruding ends and recruits LigD. Next, LigD recognizes the 5'phosphate (P) and directly mediates via a specific loop the synapsis event required for end-joining (Pitcher et al. 2007).

Most prokaryotic cells do not have the NHEJ system (Shuman and Glickman, 2007). In *S. pneumoniae,* NHEJ homologues have not been identified and such a system has not been described. In Chapter 2, we attempted to setup a NHEJ system in the pneumococcus and utilize it to repair the DSB created by Cas9, with a final scope to achieve targeted deletions and speed up genome minimization efforts.

Competence-Fratricide

The pneumococcus possesses the ability to become naturally competent for genetic transformation. Natural bacterial transformation is considered a parasexual process that involves internalization of exogenous DNA and integration into the recipient genome by homologous recombination. It enables bacteria to repair DNA (Chapter 3, Synefiaridou and Veening, 2021) and acquire new genetic traits, an important mechanism for genome plasticity. Consequently, it allows them to adapt to changing environmental conditions, enabling the pathogen to cope with host defenses and to promote resistance to antibiotics and evasion of vaccines (Andam and Hanage, 2015; Johnston et al., 2014). Requirement for transformation is that the cells enter into a transient state, termed competence. Natural competence for transformation is a widespread phenomenon among bacteria and its distribution and regulation has been covered by several original publications and reviews (Chen and Dubnau, 2004; Johnston et al., 2014; Lorenz and Wackernagel, 1994). The pneumococcal competent state has been reported to comprise, besides transformation as a means of genetic plasticity,

additional cellular mechanisms, including bacteriocin production, DNA damage, nutrient acquisition and stress response (Dagkessamanskaia et al., 2004; Guiral et al., 2007; Kjos et al., 2016; Prudhomme et al., 2006).

A plethora of genes that are involved in competence regulation and the DNA-uptake process have been identified (Figure 2). In brief, competence is controlled by a regulatory cascade, which is initiated by the binding of a small extracellular peptide (CSP). The pre-CSP, encoded by *comC* is exported and cleaved during export by its dedicated secretion apparatus, a proteolytic ABC-transporter, ComAB. After CSP accumulates above a certain threshold, it stimulates the autophosphorylation of its receptor, a membrane-bound histidine kinase, ComD, which subsequently activates its cognate response regulator ComE. While the two-component system ComDE recognizes CSP and positively regulates the early com genes, the alternative σ^x factor, ComX, activates late com gene expression (Guiral et al., 2007).

Although the entire com regulon comprises 141 CSP-responsive genes (Slager et al., 2019), only 22 are necessary for transformation, and at least 70 are individually dispensable, suggesting that this system serves purposes other than genetic transformation (Guiral et al., 2005). Such genes are *cbpD, lytA, cibA and cibB*, which, although they are non-essential for natural transformation, they have been implicated in competence-induced cell lysis, a phenomenon termed fratricide (Håvarstein et al., 2006). Fratricide takes place in mixed populations of competent and non-competent pneumococci. Competent pneumococci secrete toxins that lyse their non-competent cells from the same species, releasing DNA and other cell constituents. It has been proposed that the role of fratricide is to increase the efficiency of gene exchange between pneumococci and other closely related streptococcal species under natural conditions (Claverys and Håvarstein, 2007). Additionally, it mediates the release of the important virulence factor pneumolysin in agar plates (Guiral et al., 2005), indicating that fratricide also contributes to virulence during human infections. To protect themselves against their own fratricins, these species produce an immunity protein named ComM, which is encoded by an early competence gene (Håvarstein et al., 2006). What the mechanisms are behind ComM-dependent immunity against fratricide are unknown and is the topic of Chapter 4.

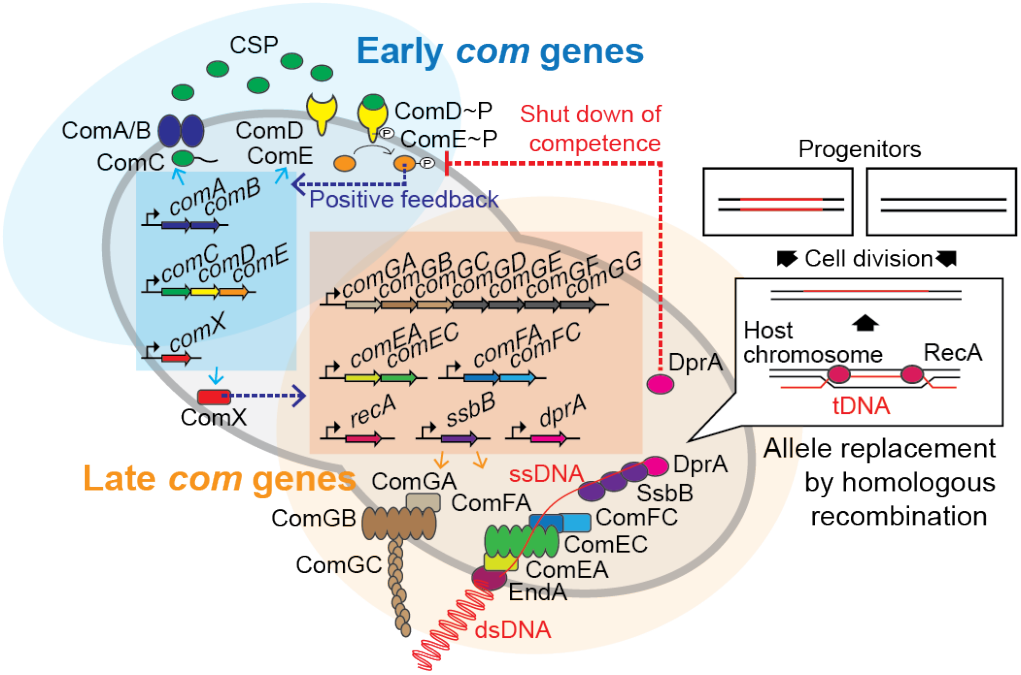


Figure 2: Regulation of pneumococcal competence and transformation. Adapted from Kurushima et al., 2020.

Cell wall-Teichoic acids

The bacterial cell wall is responsible for a number of processes, from cell shape to the mechanical strength required to withstand the internal turgor pressure. The pneumococcal cell wall is involved in uptake of substrates, cell growth and division, binding to external macromolecules and adhesion to surfaces, interactions with the human host organism, signaling in quorum sensing, uptake of DNA during transformation, autolysis, and adsorption of phages. The main constituents of the cell wall are peptidoglycans (PG) and teichoic acids (TAs) , which contain as structural components phosphoryl choline residues (Brundish and Baddiley, 1968a; Tomasz, 1967), capsular polysaccharides, and cell surface proteins (Vollmer, 2007).

The net-like peptidoglycan is formed by glycan chains that are crosslinked by short peptides. They form a network of glycan chains of alternating N-acetylmuramic acid (MurNAc) and N-acetylglucosamine (GlcNAc) residues that are connected by short peptides containing L- and D-amino acids and unusual amide bonds (Vollmer et al., 2008). PG also serves as a scaffold for anchoring other cell envelope components such as proteins (Dramsi et al., 2008), the capsule and wall teichoic acids (Neuhaus and Baddiley, 2003).

Teichoic Acids are covalently attached to the cell wall of most Gram-positive bacteria. In most bacteria, they are polymers of glycerol-phosphate or ribitol-phosphate to which are attached glycosyl and d-alanyl ester residues (Neuhaus and Baddiley, 2003). In *S.*

pneumoniae, though, TA contain phosphorylcholine substituents instead of D-alanyl esters (Fischer, 1997).

There are two major classes of TAs: lipoteichoic acids (LTAs), which are anchored to a lipid embedded in the cell membrane, and wall teichoic acids (WTAs), which are covalently attached to PG. WTA is a major cell wall polymer and may contribute to up to 50% of the cell wall's dry weight, whereas LTA is usually less abundant (Neuhaus and Baddiley, 2003). In most Gram-positive bacteria, LTAs are synthesized by completely different biosynthetic pathways from WTAs, except in the case of *S. pneumoniae* where the repeating units are structurally identical and are thought to be assembled using the same enzymes (Denapaite et al. 2012; Fischer et al. 1993). The biosynthetic pathway of TA synthesis has been well studied in many Gram-positive bacteria, like *B. subtills* and *S. aureus*, in which most of the enzymes have been identified and characterized. In the pneumococcus, TA biosynthesis involves at least 16 genes. However, there are still TA enzymes of which their role remains elusive (Figure 3, Denapaite et al., 2012; Rajagopal and Walker, 2017). Recently, it has also been shown that TacL is responsible for producing LTA from the precursor teichoic acid (Flores-Kim et al., 2019; Heß et al., 2017).

It is estimated that WTAs and other polyanionic polymers comprise up to 60 % of the cell wall mass. Along with LTAs, these polymers play central roles in numerous cellular processes (Rajagopal and Walker, 2017). A unique characteristic of *S. pneumoniae* is its auxotrophic requirement for choline (Rane and Subbarow, 1940). Choline is imported from the growth medium and incorporated into the TAs (Brundish and Baddiley, 1968; Fischer, 1997). Choline-binding proteins, which include the highly studied virulence protein PspA, have been implicated in numerous functions from adhesion to virulence, and cell wall hydrolysis (Rosenow et al., 1997). Additionally, alterations in surface TAs play a key role in antibiotic-induced lysis of the pneumococcus (Flores-Kim et al., 2019; Kovács et al., 2006). What the exact roles of TAs during competence development and fratricide are (see below) is unknown and is the topic of Chapter 4.

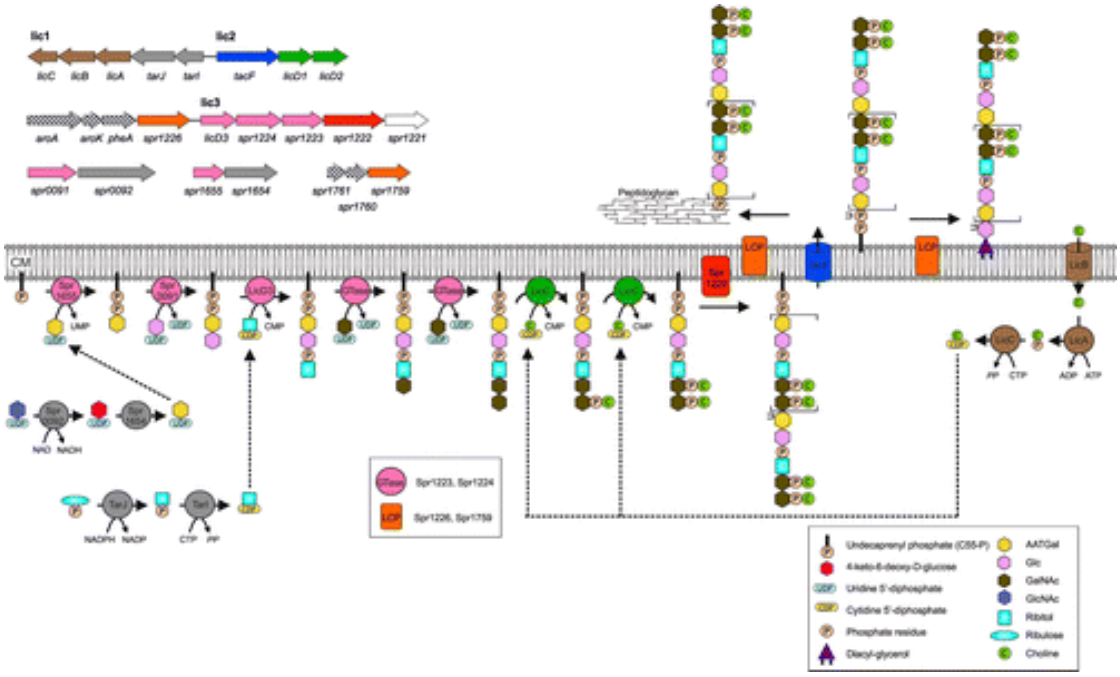


Figure 3: Proposed pathway for the biosynthesis of TAs in *S. pneumoniae* and the genetic organization of the corresponding genes. The gene/protein names and numbers are from S. pneumoniae R6. Adapted from Denapaite et al., 2012

Thesis outline

Ever since the CRISPR–Cas9 system was first adapted for use in targeted genome editing in 2012, the scientific community has made remarkable progress in improving the efficiency and accuracy of editing, expanding the list of applications and targetable species. The main goals of this thesis work are to establish tools for pneumococcal research using the CRISPR-Cas technology, to apply them to perform genome editing and genome minimization, and to better understand competence-dependent fratricide. To reach these goals, we were first interested in establishing a framework for rational engineering of the CRISPR toolbox in *S. pneumoniae*. An approach to achieve this requires the prior setup of a NHEJ system in the pneumococcus (Chapter 2). Alternatively, we can use the HR pathway. Specifically, we demonstrate a genetic system for making targeted, marker-less gene knockouts and large genome deletions (Chapter 3).

Additionally, modified versions of CRISPR-Cas9 have been used to explore fundamental cell biology questions. Performing a CRISPRi-seq screen using a catalytically inactive Cas9, dCas9, we identified genes that become essential when competence is triggered. Our results revealed that most were related to TA synthesis pathways, suggesting their important role in protecting cells from competence-related autolysis, leading us to investigate the molecular mechanism underlying immunity against fratricide. Furthermore, we propose an updated biosynthetic pathway for TA (Chapter 4).

References

Andam, C.P., and Hanage, W.P. (2015). Mechanisms of genome evolution of *Streptococcus*. Infect. Genet. Evol. *33*, 334–342.

Anzalone, A. V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. Nature *576*, 149–157.

Anzalone, A. V., Koblan, L.W., and Liu, D.R. (2020). Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. Nat. Biotechnol. *38*, 824–844.

Austrian, R. (1999). Pneumococcus and the Brooklyn connection. Am. J. Med. *107*, 2–6.

Austrian, R. (1960). The Gram stain and the etiology of lobar pneumonia, an historical note. Bacteriol. Rev. *24*, 261–265.

Avery, O.T., Macleod, C.M., and Mccarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus Type III. J. Exp. Med. *79*, 137–158.

Bhaya, D., Davison, M., and Barrangou, R. (2011). CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation. Annu. Rev. Genet. *45*, 273–297.

Bikard, D., Hatoum-Aslan, A., Mucida, D., and Marraffini, L. a. (2012). CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. Cell Host Microbe *12*, 177–186.

Brundish, D.E., and Baddiley, J. (1968a). Pneumococcal C-substance, a ribitol teichoic acid containing choline phosphate. Biochem. J. *110*, 573–582.

Brundish, D.E., and Baddiley, J. (1968b). Pneumococcal C-Substance, a Ribitol Teichoic Acid Containing Choline Phosphate.

Chen, I., and Dubnau, D. (2004). DNA uptake during bacterial transformation. Nat. Rev. Microbiol. *2*, 241–249.

Choi, P.S., and Meyerson, M. (2014). Targeted genomic rearrangements using CRISPR/Cas technology. Nat. Commun. *5*, 1–6.

Claverys, J.P., and Håvarstein, L.S. (2007). Cannibalism and fratricide: Mechanisms and raisons d'être. Nat. Rev. Microbiol. *5*, 219–229.

Cui, L., Vigouroux, A., Rousset, F., Varet, H., Khanna, V., and Bikard, D. (2018). A CRISPRi screen in *E. coli* reveals sequence-specific toxicity of dCas9. Nat. Commun. *9*, 1–10.

Dagkessamanskaia, A., Moscoso, M., Hénard, V., Guiral, S., Overweg, K., Reuter, M., Martin, B., Wells, J., and Claverys, J.P. (2004). Interconnection of competence, stress and

CiaR regulons in *Streptococcus pneumoniae*: Competence triggers stationary phase autolysis of ciaR mutant cells. Mol. Microbiol. *51*, 1071–1086.

Deibel, R.H., and Seeley Jr, H.W. (1974). Family II. Streptococcaceae fam. nov. 7. *Streptococcus pneumoniae*. Bergey's Man. Determ. Bacteriol. Williams Wilkins Co., Balt. 499–500.

Denapaite, D., Brückner, R., Hakenbeck, R., and Vollmer, W. (2012). Biosynthesis of teichoic acids in *Streptococcus pneumoniae* and closely related species: Lessons from genomes. Microb. Drug Resist. *18*, 344–358.

Dramsi, S., Magnet, S., Davison, S., and Arthur, M. (2008). Covalent attachment of proteins to peptidoglycan. FEMS Microbiol. Rev. *32*, 307–320.

Ercoli, G., Fernandes, V.E., Chung, W.Y., Wanford, J.J., Thomson, S., Bayliss, C.D., Straatman, K., Crocker, P.R., Dennison, A., Martinez-Pomares, L., et al. (2018). Intracellular replication of *Streptococcus pneumoniae* inside splenic macrophages serves as a reservoir for septicaemia. Nat. Microbiol. *3*, 600–610.

Fischer, W. (1997). Pneumococcal lipoteichoic and teichoic acid. Microb. Drug Resist. *3*, 309–325.

Flores-Kim, J., Dobihal, G.S., Fenton, A., Rudner, D.Z., and Bernhardt, T.G. (2019). A switch in surface polymer biogenesis triggers growth-phase-dependent and antibiotic-induced bacteriolysis. Elife *8*.

Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., and Liu, D.R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature *551*, 464–471.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647–661.

Griffith, F. (1928). The Significance of Pneumococcal Types. J. Hyg. (Lond). *27*, 333–339.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics *8*, 1–10.

Guiral, S., Mitchell, T.J., Martin, B., and Claverys, J.P. (2005). Competence-programmed predation of noncompetent cells in the human pathogen *Streptococcus pneumoniae*: Genetic requirements. Proc. Natl. Acad. Sci. U. S. A. *102*, 8710–8715.

Guiral, S., Hénard, V., Granadel, C., Martin, B., and Claverys, J.P. (2006). Inhibition of competence development in *Streptococcus pneumoniae* by increased basal-level expression of the ComDE two-component regulatory system. Microbiology *152*, 323–331.

Guiral, S., Moscoso, M., Dagkessamanskaia, A., and Claverys, J.P. (2007). Molecular Biology of Streptococci (Horizon Bioscience).

Hanna, R.E., and Doench, J.G. (2020). Design and analysis of CRISPR–Cas experiments. Nat. Biotechnol. *38*, 813–823.

Håvarstein, L.S., Martin, B., Johnsborg, O., Granadel, C., and Claverys, J.P. (2006). New insights into the pneumococcal fratricide: Relationship to clumping and identification of a novel immunity Factor. Mol. Microbiol. *59*, 1297–1037.

Heidelberger, M., Aisenberg, A.C., And Hassid, W.Z. (1954). Glycogen, an immunologically specific polysaccharide. J. Exp. Med. *99*, 343–353.

Heß, N., Waldow, F., Kohler, T.P., Rohde, M., Kreikemeyer, B., Gómez-Mejia, A., Hain, T., Schwudke, D., Vollmer, W., Hammerschmidt, S., et al. (2017). Lipoteichoic acid deficiency permits normal growth but impairs virulence of *Streptococcus pneumoniae*. Nat. Commun. *8*.

Jansen, R., Embden, J.D.A. van, Gaastra, W., and Schouls, L.M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. Mol. Microbiol. *43*, 1565–1575.

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat. Biotechnol. *31*, 233–239.

Jiang, Y., Chen, B., Duan, C., Sun, B., Yang, J., and Yang, S. (2015). Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 system. Appl. Environ. Microbiol. *81*, 2506–2514.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816–821.

Johnston, C., Martin, B., Fichant, G., Polard, P., and Claverys, J.-P. (2014). Bacterial transformation: distribution, shared mechanisms and divergent control. Nat. Rev. Microbiol. *12*, 181–196.

Kadioglu, A., Weiser, J.N., Paton, J.C., and Andrew, P.W. (2008). The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. Nat. Rev. Microbiol. *6*, 288–301.

Kjos, M. (2019). Transcriptional Knockdown in Pneumococci Using CRISPR Interference. In Methods in Molecular Biology, (Humana Press Inc.), pp. 89–98.

Kjos, M., Miller, E., Slager, J., Lake, F.B., Gericke, O., Roberts, I.S., Rozen, D.E., and Veening, J.-W. (2016). Expression of *Streptococcus pneumoniae* Bacteriocins Is Induced by Antibiotics via Regulatory Interplay with the Competence System. PLOS Pathog. *12*, e1005422.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature *533*, 420–424.

Koonin, E. V, Makarova, K.S., and Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. Curr. Opin. Microbiol. *37*, 67–78.

Kovács, M., Halfmann, A., Fedtke, I., Heintz, M., Peschel, A., Vollmer, W., Hakenbeck, R., and Brückner, R. (2006). A functional dlt operon, encoding proteins required for incorporation of D-alanine in teichoic acids in gram-positive bacteria, confers resistance to cationic antimicrobial peptides *in Streptococcus pneumoniae*. J. Bacteriol. *188*, 5797–5805.

Kurushima, J., Campo, N., van Raaphorst, R., Cerckel, G., Polard, P., and Veening, J.-W. (2020). Unbiased homeologous recombination during pneumococcal transformation allows for multiple chromosomal integration events. Elife *9*.

Lanie, J.A., Ng, W.-L., Kazmierczak, K.M., Andrzejewski, T.M., Davidsen, T.M., Wayne, K.J., Tettelin, H., Glass, J.I., and Winkler, M.E. (2007). Genome sequence of Avery's virulent serotype 2 strain D39 of *Streptococcus pneumoniae* and comparison with that of unencapsulated laboratory strain R6. J. Bacteriol. *189*, 38–51.

Larson, M.H., Gilbert, L.A., Wang, X., Lim, W.A., Weissman, J.S., and Qi, L.S. (2013). CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat. Protoc. *8*, 2180–2196.

Lee, H.H., Ostrov, N., Wong, B.G., Gold, M.A., Khalil, A.S., and Church, G.M. (2019). Functional genomics of the rapidly replicating bacterium *Vibrio natriegens* by CRISPRi. Nat. Microbiol. *4*, 1105–1113.

Liang, X., Potter, J., Kumar, S., Ravinder, N., and Chesnut, J.D. (2017). Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. J. Biotechnol. *241*, 136–146.

Liu, X., Gallay, C., Kjos, M., Domenech, A., Slager, J., Kessel, S.P., Knoops, K., Sorg, R.A., Zhang, J., and Veening, J. (2017). High-throughput CRISPRi phenotyping identifies new essential genes in *Streptococcus pneumoniae* . Mol. Syst. Biol. *13*, 931.

Liu, X., Kimmey, J.M., Matarazzo, L., de Bakker, V., Van Maele, L., Sirard, J.C., Nizet, V., and Veening, J.W. (2021). Exploration of Bacterial Bottlenecks and *Streptococcus pneumoniae* Pathogenesis by CRISPRi-Seq. Cell Host Microbe *29*, 107-120.e6.

Lorenz, M.G., and Wackernagel, W. (1994). Bacterial gene transfer by natural genetic transformation in the environment. Microbiol. Mol. Biol. Rev. *58*.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J.M., Wolf, Y.I., Yakunin, A.F., et al. (2011a). Evolution and classification of the CRISPR-Cas systems. Nat. Rev. Microbiol. *9*, 467–477.

Makarova, K.S., Aravind, L., Wolf, Y.I., and Koonin, E. V. (2011b). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. Biol. Direct *6*, 1–27.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J.J., Charpentier, E., Haft, D.H., et al. (2015). An updated evolutionary classification of CRISPR–Cas systems. Nat. Rev. Microbiol. *13*, 722–736.

Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., et al. (2020). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nat. Rev. Microbiol. *18*, 67–83.

Morimoto, T., Ara, K., Ozaki, K., and Ogasawara, N. (2009). A new simple method to introduce marker-free deletions in the *Bacillus subtilis* genome. Genes Genet. Syst. *84*, 315–318.

Neufeld, F., and Etinger-Tulczynska, R. (1933). Schnelldiagnose der Pneumokokkentypen aus dem Auswurf. Zeitschrift Für Hyg. Und Infekt. *115*, 431–435.

Neuhaus, F.C., and Baddiley, J. (2003). A Continuum of Anionic Charge: Structures and Functions of d-Alanyl-Teichoic Acids in Gram-Positive Bacteria. Microbiol. Mol. Biol. Rev. *67*, 686–723.

O'Brien, K.L., Wolfson, L.J., Watt, J.P., Henkle, E., Deloria-Knoll, M., McCall, N., Lee, E., Mulholland, K., Levine, O.S., and Cherian, T. (2009). Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. Lancet *374*, 893–902.

Pasteur, L. (1881). Sur une maladie nouvelle provoquée par la salive d'un enfant mort de rage. In Bulletin de l'Academie de Medicine (Paris), pp. 94–108.

Pitcher, R.S., Brissett, N.C., and Doherty, A.J. (2007). Nonhomologous end-joining in bacteria: a microbial perspective. Annu. Rev. Microbiol. *61*, 259–282.

Prudhomme, M., Attaiech, L., Sanchez, G., Martin, B., and Claverys, J.P. (2006). Antibiotic stress induces genetic transformability in the human pathogen streptoccus pneumoniae. Science (80-.). *313*, 89–92.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. Cell *152*, 1173–1183.

Rajagopal, M., and Walker, S. (2017). Envelope structures of gram-positive bacteria. In Current Topics in Microbiology and Immunology, (Springer Verlag), pp. 1–44.

Rane, L., and Subbarow, Y. (1940). Nutritional Requirements of the Pneumococcus: I. Growth Factors for Types I, II, V, VII, VIII. J. Bacteriol. *40*, 695–704.

Rosenow, C., Ryan, P., Weiser, J.N., Johnson, S., Fontan, P., Ortqvist, A., and Masure, H.R. (1997). Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of *Streptococcus pneumoniae*. Mol. Microbiol. *25*, 819–829.

Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O.O., Gootenberg, J.S., Makarova, K.S., Wolf, Y.I., et al. (2017). Diversity and evolution of class 2 CRISPR–Cas systems. Nat. Rev. Microbiol. *15*, 169–182.

Shuman, S., and Glickman, M.S. (2007). Bacterial DNA repair by non-homologous end

joining. Nat. Rev. Microbiol. *5*, 852–861.

Simell, B., Auranen, K., Käyhty, H., Goldblatt, D., Dagan, R., and O'Brien, K.L. (2012). The fundamental link between pneumococcal carriage and disease. Expert Rev. Vaccines *11*, 841–855.

Slager, J., Aprianto, R., and Veening, J.W. (2019). Refining the pneumococcal competence regulon by RNA sequencing. J. Bacteriol. *201*.

Sternberg, G. (1881). A fatal form of septicaemia in the rabbit produced by the subcutaneous injection of human saliva : an experimental research (Baltimore: Printed by John Murphy & Co.).

Synefiaridou, D., and Veening, J.-W. (2021). Harnessing CRISPR-Cas9 for genome editing in *Streptococcus pneumoniae* D39V . Appl. Environ. Microbiol.

Tomasz, A. (1967). Choline in the cell wall of a bacterium: Novel type of polymer-linked choline in pneumococcus. Science (80-.). *157*, 694–697.

Vollmer, W. (2007). Structure and Biosynthesis of the Pneumococcal Cell Wall (Ovid Technologies (Wolters Kluwer Health)).

Vollmer, W., Blanot, D., and De Pedro, M.A. (2008). Peptidoglycan structure and architecture. FEMS Microbiol. Rev. *32*, 149–167.

Wang, Y., Zhang, Z.-T., Seo, S.-O., Lynn, P., Lu, T., Jin, Y.-S., and Blaschek, H.P. (2016). Bacterial Genome Editing with CRISPR-Cas9: Deletion, Integration, Single Nucleotide Modification, and Desirable "Clean" Mutant Selection in *Clostridium beijerinckii* as an Example. ACS Synth. Biol. *5*, 721–732.

Weinberger, D.M., Trzciński, K., Lu, Y.-J., Bogaert, D., Brandes, A., Galagan, J., Anderson, P.W., Malley, R., and Lipsitch, M. (2009). Pneumococcal Capsular Polysaccharide Structure Predicts Serotype Prevalence. PLoS Pathog. *5*, e1000476.

Weiser, J.N., Ferreira, D.M., and Paton, J.C. (2018). *Streptococcus pneumoniae:* transmission, colonization and invasion. Nat. Rev. Microbiol. *16*, 355–367.

de Wet, T.J., Gobe, I., Mhlanga, M.M., and Warner, D.F. (2018). CRISPRi-Seq for the Identification and Characterisation of Essential Mycobacterial Genes and Transcriptional Units. BioRxiv 358275.

Winslow, C.-E.A., Broadhurst, J., Buchanan, R.E., Krumwiede, C., Rogers, L.A., and Smith, G.H. (1920). The Families and Genera of the Bacteria: Final Report of the Committee of the Society of American Bacteriologists on Characterization and Classification of Bacterial Types. J. Bacteriol. *5*, 191–229.

Chapter 2: Transplantation of heterologous non-homologous end-joining systems and engineered DNA ligases in *Streptococcus pneumoniae*

Authors: Dimitra Synefiaridou and Jan-Willem Veening

(D.S. designed research, performed experiments, analyzed data, and wrote this chapter;
unpublished)

Abstract

Genome editing typically relies on introducing a double stranded break (DSB) at the target locus followed by DNA repair. In case of CRISPR-Cas9-based genome editing, after introduction of a DSB into the DNA, the cell can utilize the non-homologous end-joining (NHEJ) system to repair the break. NHEJ typically will result in gene inactivation by producing indels of variable lengths. A minimal two-component bacterial NHEJ repair apparatus composes of Ku, which binds to 3'-protruding ends and further recruits a multifunctional DNA ligase (LigD), which ligates the nicks and repairs the break. Although most prokaryotic cells, including *Streptococcus pneumoniae*, lack a NHEJ system, it has been identified in some prokaryotes. Here, we transplanted the NHEJ machinery from *Bacillus subtilis, Mycobacterium tuberculosis, M. smegmatis* and *Streptomyces ambofaciens* to pneumococcus, with the overall goal to establish NHEJ and exploit it for advanced genome engineering approaches. Our results show that none of these NHEJ systems can facilitate repair from the DSB induced either by Cas9, or by the meganuclease I-SceI. Similarly, engineered ligases p50 and sso7D do not repair the DSB by Cas9. Additionally, we failed to facilitate a repair from a non-enzymatic DSB, using Mitomycin C (MMC). Our findings suggest an incompatibility between the CRISPR-Cas9 and the NHEJ system. Finally, we discuss potential strategies to tackle this issue and achieve the setup of a successful NHEJ system.

Introduction

After a double-stranded break (DSB) has been introduced by Cas9, the cell can utilize two major pathways in order to repair the break and survive; homologous recombination (HR) or non-homologous end-joining (NHEJ). HR and NHEJ differ in several fundamental manners. In HR, a second, intact copy of the broken chromosome segment, homologous to the DSB site, serves as a template for DNA synthesis across the break. In this mechanism, the crucial process of locating and recombining the homologous sequence is performed by RecA (Shuman and Glickman, 2007).

NHEJ does not rely on a homologous DNA template, as the two DNA ends are rejoined directly together. It is a potentially less acurate, error-prone pathway, in which insertion or deletion mutations (indels) of various lengths are introduced into the target locus (Weterings and Chen, 2008). These indels usually alter the gene's open reading frame and consequently lead to the generation of gene knockouts (Jiang and Marraffini, 2015). Although the molecular mechanism of the NHEJ process looks deceptively simple, in reality, it is rather complex and requires a tightly coordinated interplay between the key enzymes to timely initiate each subset of the entire process. A set of enzymes mediate (1) the capture of both ends of the broken DNA molecule, (2) the formation of a molecular bridge that brings the two DNA ends back together; and (3) the subsequent re-ligation of the broken DNA molecule (Weterings and Chen, 2008).

Historically, all of the major DNA-repair pathways were first described in prokaryotes and, subsequently, equivalent eukaryotic counterparts were characterized. However, NHEJ was initially discovered in the eukarya, before a homologous NHEJ apparatus was identified in prokaryotes (Aravind and Koonin, 2001; Weller *et al.*, 2002). NHEJ is a major pathway for repairing DNA double-strand breaks (DSBs) in mammalian cells and it may occur by several distinct mechanisms (Krejci *et al.*, 2003).

The NHEJ pathway in eukaryotes is best defined by its minimal core elements: the Ku70/80 heterodimer (Yku70/Yku80 in yeast) and the ligase IV/XRCC4/XLF (LXX) complex (Dnl4/Lif1/Nej1 in yeast) (Pitcher, Brissett and Doherty, 2007). The primary DNA end-binding component of NHEJ, Ku70/80, forms an open ringlike structure through which a variety of DNA end structures can be threaded, potentially aligning the termini and protecting the ends from unspecific nucleolytic degradation (Walker, Corpina and Goldberg, 2001). DNA-bound Ku helps to recruit the LXX complex, thereby enhancing its ligation activity. In higher eukaryotes, Ku also recruits the DNA-dependent protein kinase catalytic subunit (DNA-PKcs), thereby activating its kinase activity, which is required for DSB rejoining (Lees-Miller and Meek, 2003). Additional factors, such as polymerases (μ, λ, and Pol4), nucleases (Artemis

and Fen1), and other additional activities (polynucleotide kinase and phosphatase), are required to process DNA ends prior to ligation (Krejci *et al.*, 2003; Daley *et al.*, 2005).

A functional NHEJ repair apparatus, composed of Ku and a multifunctional DNA ligase (LigD), has also been identified several prokaryotes. NHEJ offers protection to species that live in stressful environments when only a single copy of the genome is available, such as after sporulation or during stationary phase (Moeller *et al.*, 2007; Pitcher *et al.*, 2007). In the minimal two-component bacterial NHEJ repair complex, initially, Ku binds to 3'-protruding ends and further recruits LigD. Specifically, it recognizes the 5' phosphate (P) and directly mediates, via a specific loop, the synapsis event required for end-joining. After microhomology pairing, non-extendable 3' termini can be cleaved by the nuclease domain (NucDom) of LigD. Once the required nucleolysis has occurred, resynthesis by the polymerase domain (PolIDom) and ligation of nicks complete break repair (Pitcher, Brissett and Doherty, 2007).

Among the bacteria for which a NHEJ system has been identified are *Bacillus subtilis, Mycobacterium tuberculosis, M. smegmatis* and *Streptomyces ambofaciens. B. subtilis* belongs to the class of Bacilli, similarly to *S. pneumoniae.* The *B. subtilis* NHEJ pathway consists of two genes in the same operon; *ykoV,* with a significant homology to the eukaryotic Ku proteins (BsuKu*)* and *ykoU*, which is encoding for a putative bimodular LigD (BsuLigD) with a C-terminal DNA ligase catalytic domain (Aravind and Koonin, 2001; de Vega, 2013). In *M. tuberculosis,* Mt-Ku and Mt-Lig proteins constitute a two-component NHEJ repair machine. Their enzyme activities allow these two proteins to join incompatible DSB ends *in vitro*, as well as to reconstitute NHEJ *in vivo* in mutant yeast (Della *et al.*, 2004). Similarly, *M. smegmatis,* has a robust NHEJ pathway that requires Ku and a specialized polyfunctional ATP-dependent DNA ligase (LigD). (Gong *et al.*, 2005). *S. ambofaciens* has "core" NHEJ gene set genes constituted of conserved loci. It is hypothesized that the NHEJ ligase could be LigC rather than LigD, since both carry the ligase domain. Both ligases are devoid of the PolDom domain known to be involved in its recruitment by Ku. Therefore, LigD and LigC may ensure similar functional roles in Streptomyces and provide an equivalent contribution to the bacterium, therefore potentially, LigC, could replace LigD in the "core" NHEJ set (Hoff *et al.*, 2016).

Our aim is to utilize NHEJ for genome minimization in the pneumococcus. Our planned approach is as follows: once a deletion target is selected, two sgRNAs, targeting on both sides of the deletion target will be constructed and transformed into a strain with the CRISPR-Cas9 and NHEJ systems. Next, the Cas9 will be induced leaving two DSB on both sides of the deletion target. Subsequently, the NHEJ mechanism will be induced and recruited at the break sites, where it facilitates the repair by re-ligating the two ends, leaving out the deletion target. With this process, the final product is a new strain, with a markerless deletion, achieved by a

precise, scalable and multiplex approach, with the only requirement to perform any new deletion, two specific sgRNAs (Figure S1).

In this study, we transplanted the NHEJ systems of the aforementioned four different species, to *S. pneumoniae* with the overall goal to establish NHEJ and exploit it for advanced genome engineering approaches.

Results

IPTG-inducible Cas9 successfully performs DSB in pneumococcus

First, *S. pneumoniae* strain VL324 was constructed, which contained all the necessary components of the CRISPR-Cas9 system to introduce a DSB. In this strain, the *S. pyogenes cas9* gene was placed under control of a LacI-based isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible promoter, named *Plac* (Sorg et al PNAS 2020), and was integrated into the chromosome via double crossover at the non-essential *bgaA* locus. The gene coding for the *E. coli* LacI repressor was integrated at the non-essential *prs1* site (Liu *et al.*, 2017). To confirm the reliability of the system, the firefly luciferase (*luc*) reporter gene was used. Finally, a sgRNA targeting *luc* was placed under the constitutive P3 promoter and integrated at the non-essential locus CEP (Figure 1a).

In the absence of IPTG, expression of *cas9* is tightly repressed and cells grow similar to wild type. In addition, transcription of the target gene (*luc*) can proceed, which can be followed by luciferase activity. Induction of Cas9 with 1 mM IPTG resulted in strong growth retardation (Figure 1b) and absence of luciferase activity (Figure 1c), suggesting that the DSB was performed efficiently leading to cell death. Note that when a catalytically inactive Cas9, dCas9, was used to target *luc*, transcription of *luc* was inhibited but growth was unperturbed (Liu et al. MSB 2017), suggesting that the observed growth defect upon induction of wild type Cas9 is caused by introduction of a DSB and not by a transcriptional roadblock (Figure 1b).

Next, the minimal concentration of IPTG to induce Cas9 and perform the DSB was determined by addition of several concentrations of IPTG. It was observed that the level of growth defect is tunable by using different concentrations of IPTG; the higher the concentration of IPTG, the more severe the growth defect (Figure 1a). However, the system was saturated at 20 μM or higher of IPTG, which was determined as the minimal concentration required to kill the cells. These findings also corresponded to luciferase expression levels (Figure 1b). Note that this is the sufficient concentration required to induce a DSB using the sgRNA targeting luciferin gene and higher concentrations of IPTG might be necessary with less efficient sgRNAs. In any case, these results shows that induction of Cas9 is titratable via the *Plac* promoter.

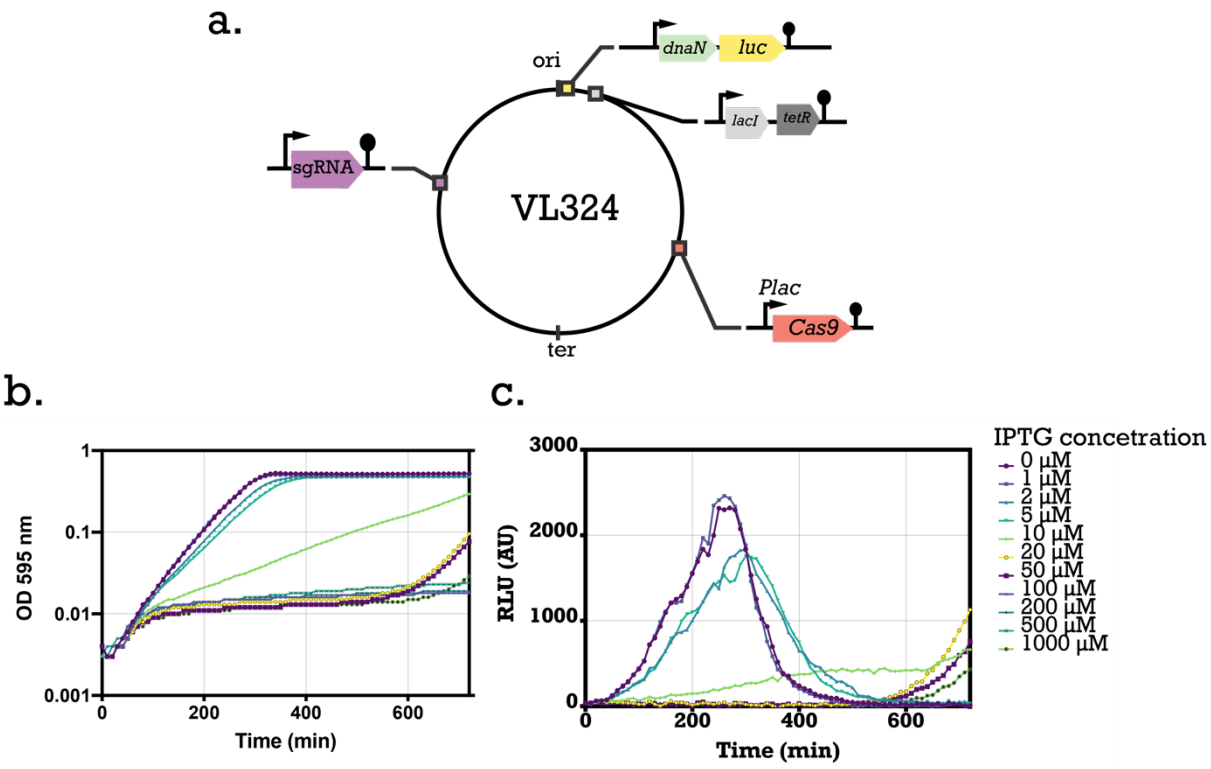


Figure 1: An IPTG-inducible CRISPR-cas9 system in S. pneumoniae (a) Genetic map of CRISPR-Cas9 luc reporter strain VL324. Expression of Cas9 was induced by addition of different concentrations of IPTG. **(b)** Cell density (OD595) and **(c)** luciferase activity (shown as RLU) of the bacterial cultures were measured every 10 min. The values represent averages of three replicates. Experiments were repeated at least 3 times. A representative growth curve is shown.

Nonhomologous End Joining Repair Machinery transplantation in *S. pneumoniae* does not facilitate repair from DSB by Cas9

After we established the CRISPR-Cas9 system in *S. pneumoniae* resulting in strain VL324, we transplanted the NHEJ system into this strain. For this, the two (or three) genes of the minimal NHEJ system were placed under control of a tetracycline-inducible promoter, named *Ptet* (Sorg *et al.*, 2020) (Figure 2a). If the NHEJ system would work, it would facilitate the repair of the DSB, by formation of indels, that would block further cleavages and restore growth, but disturb the activity of the luciferin gene, so that he repaired strain would not emit light.

To test this, we performed growth assays. Cells were pre-grown in C+Y with aTC and were induced either with IPTG to demonstrate that they are successfully killed, or with the minimal amount of IPTG and additionally the maximum amount of aTC, to activate the repair system, and their growth as well as the luciferase activity was measured with a plate reader. In all cases, we did not observe restored growth upon induction of NHEJ systems. Cell density and

luciferase activity do not recover in all the NHEJ systems, indicating that they are not expressed appropriately or functional in *S. pneumoniae* (Figure 2b).

For a final confirmation, we repeated the aforementioned growth experiment, in liquid media, but this time at 30°C, a temperature that pneumococcus can grow fine, but slower. The reasoning behind this temperature switch was that the slower growth could facilitate the function of the NHEJ mechanisms. However, this was not the case, since growth was not restored in any of the cases (Figure S2).

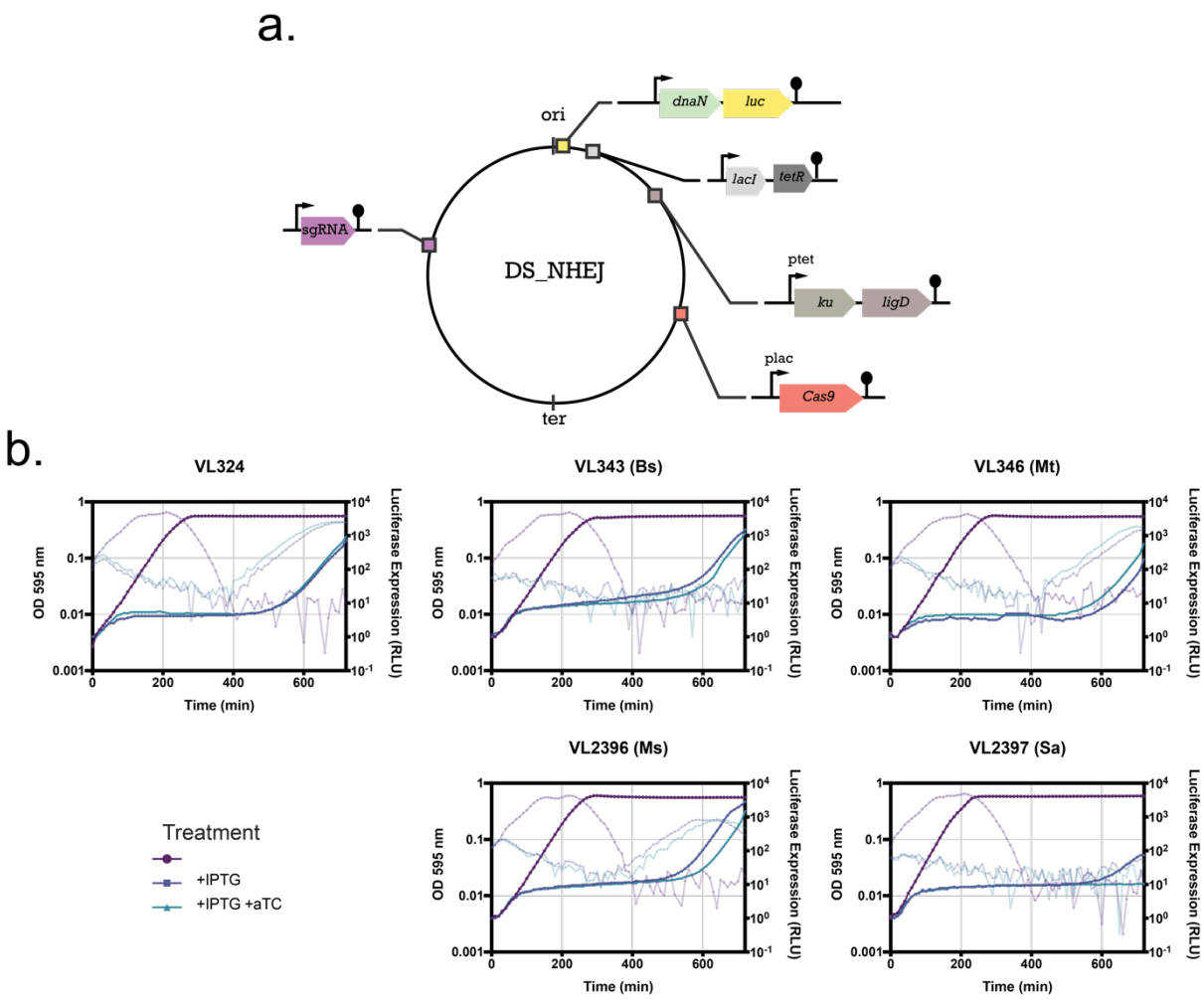


Figure 2: An IPTG-inducible CRISPR-cas9 system and an aTC inducible NHEJ system in S. pneumoniae. (a) Genetic map of CRISPR-Cas9 luc reporter strain with NHEJ system. **(b)** Cell density (OD595) (left y axis) and luciferase activity (shown as RLU) (right y axis) of the bacterial cultures were measured every 10 min. The values represent averages of three replicates. Experiments were repeated at least 3 times. A representative growth curve is shown. Control strain with no NHEJ system (strain VL324), NJEJ system of *B. subtilis* (strain VL343), NHEJ system of *M. tuberculosis* (strain VL344), NHEJ system of *M. smegmatis* (strain VL2396), and NHEJ system of *S. ambofaciens* (strain VL2397) were tested in three conditions. No induction, induction with 20 μM of IPTG and induction with 20 μM IPTG+500 ng/ml aTC. All strains were grown at 37°C in C+Y medium.

Nonhomologous End Joining Repair Machinery transplantation in *S. pneumoniae* does not facilitate repair from DSB by meganuclease I-SceI

To investigate whether there was an incompatibility between the NHEJ and the Cas9, but the NHEJ apparatus could be functional in the case that the DSB was performed by another enzyme, we used a strain that was producing, under the control of the Plac promoter, the meganuclease I-SceI, an intron-encoded homing endonuclease isolated from the yeast Saccharomyces cerevisiae (Jacquier and Dujon, 1985) and additionally carried its 18 bp recognition site, unique in the pneumococcal genome (VL347). In this strain, we transplanted the NHEJ systems from the four different species and we performed the same growth assays. During growth in C+Y at 37°C, unlike Cas9, when we induce the expression of I-SceI, we do not observe any growth defect, suggesting that that a successful DSB does not occur (Figure S3). To resolve this issue, we performed the same experiments at 30°C, expecting that the slower growth could facilitate the function of the enzyme. Indeed, at this temperature we observed some growth defect (Figure 3), suggesting partial DSB, but this growth defect was less severe compared to the one caused by Cas9-induced DSBs. Nevertheless, we did not observe growth restoration upon induction of the NHEJ components.

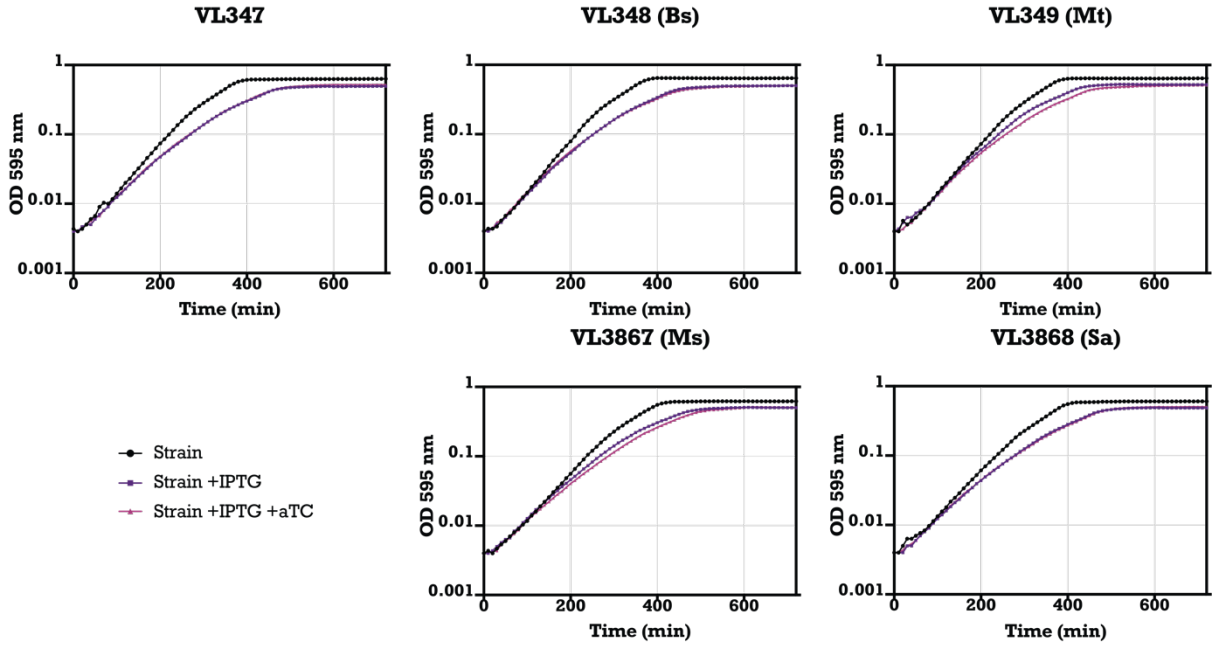


Figure 3: **An IPTG-inducible I-SceI and an aTC-inducible NHEJ system in *S. pneumoniae*.** Cell density (OD595) of the bacterial cultures were measured every 10 min. The values represent averages of three replicates. Experiments were repeated at least 3 times. A representative growth curve is shown. Control strain with no NHEJ system (strain VL347), NJEJ system of *B. subtilis* (strain VL348), NHEJ system of *M. tuberculosis* (strain VL349), NHEJ system of *M. smegmatis* (strain VL3867), and NHEJ system of *S. ambofaciens* (strain VL3868) were tested in three conditions. No induction, induction with 20 µM of IPTG and induction with 20 µM IPTG+500 ng/ml aTC. All strains were grown at 30°C.

Ligases p50 and sso7D in *S. pneumoniae* do not facilitate repair from DSB by Cas9

Since we were not successful in setting up a functional NHEJ system, we used another approach to repair the DSB and transplanted two engineered versions of the DNA ligase from bacteriophage T4. T4 ligase is one of the most widely used enzymes in molecular biology. Although it has evolved to seal single-stranded nicks in double-stranded DNA, and not to join double-stranded fragments with cohesive or blunt ends, engineered active variants p50-ligase and Sso7d-ligase (i.e. NF-kB p50 and Sso7d fused to T4 DNA ligase) showed noticeable abilities to join blunt-ended DNA (Wilson *et al.*, 2013).

In the strain VL324, p50-ligase or Sso7d-ligase were placed under control of *Ptet.* Next, we followed the growth in C+Y liquid medium (Figure 4). We did not observe rescue of the growth defect caused by Cas9. Whether this is due to lack of repair by the ligases or whether repair results in new template for Cas9, needs to be determined.

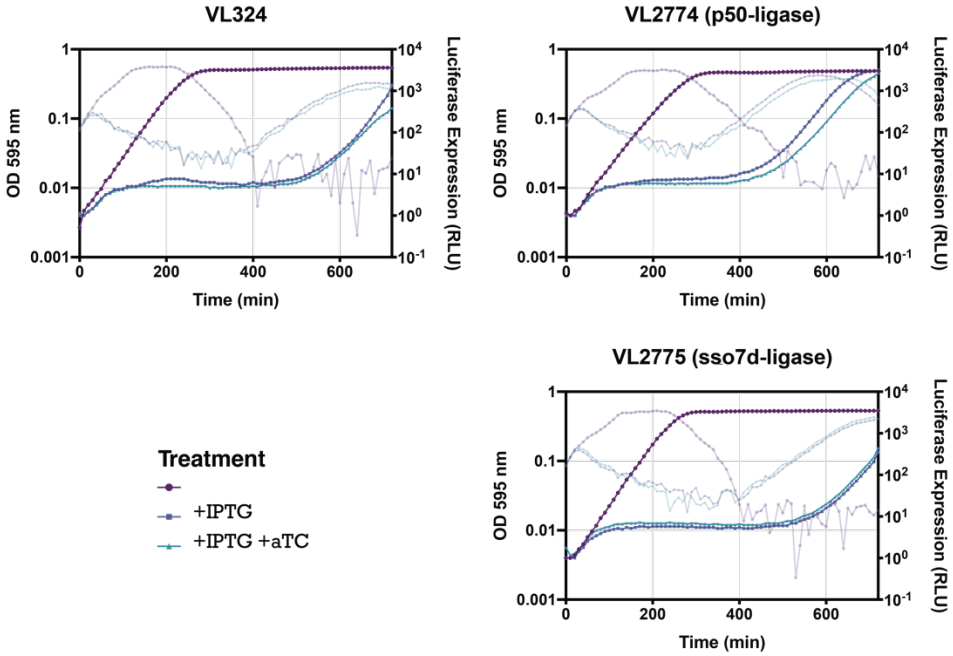


Figure *4*: An IPTG-inducible CRISPR-cas9 system and an aTC inducible ligase in *S. pneumoniae*. Cell density (OD595) (left y axis) and luciferase activity (shown as RLU) (right y axis) of the bacterial cultures were measured every 10 min. The values represent averages of three replicates. Experiments were repeated at least 3 times. A representative growth curve is shown. Control strain with no NHEJ system (strain VL324), p50-ligase (strain VL2774), and sso7d-ligase (strain V:2775) were tested in three conditions. No induction, induction with 20 µM of IPTG and induction with 20 µM IPTG+500 ng/ml aTC. All strains were grown at 37°C.

Nonhomologous End Joining Repair Machinery *transplantation in S. pneumoniae does not facilitate repair from DSB by Mitomycin C*

Although, it was not possible to setup a functional repair from one of the tested NHEJ system when the DSB is introduced by an enzyme (Cas9 and I-SceI), we were interested in testing whether it could facilitate a repair from a non-enzymatic DNA lesion. For this assay, we used Mitomycin C (MMC), a naturally produced antibiotic, originally isolated from *Streptomyces lavendulae*, which reacts specifically with guanine residues in DNA and results in three principal modifications; it either forms a monoadduct by reacting with a single guanine, an intrastrand cross-link on adjacent guanines on the same strand, or an interstrand cross-link wherein the two guanines on opposite strands of CpG sequences are covalently linked. In bacteria, MMC adducts and intrastrand cross-links are repaired by nucleotide excision repair and interstrand cross-links are repaired by a combination of nucleotide excision repair and homologous recombination (Burby and Simmons, 2019).

First, the minimal concentration MMC to perform the DNA lesion was determined by addition of several concentrations of MMC. It was observed that the level of growth defect is tunable by using different concentrations of MMC and the higher the concentration, the more severe the growth defect (Figure 5a). The MMC concentration of 5000 ng/ml was selected, since no growth was observed using it.

Bacteria were pre-grown with aTC to induce the NHEJ system, before exposing them with 5000 ng/ml of MMC for 20 min. Growth analysis did not observe any restoration of growth upon induction of any of the NHEJ systems (Figure 5b).

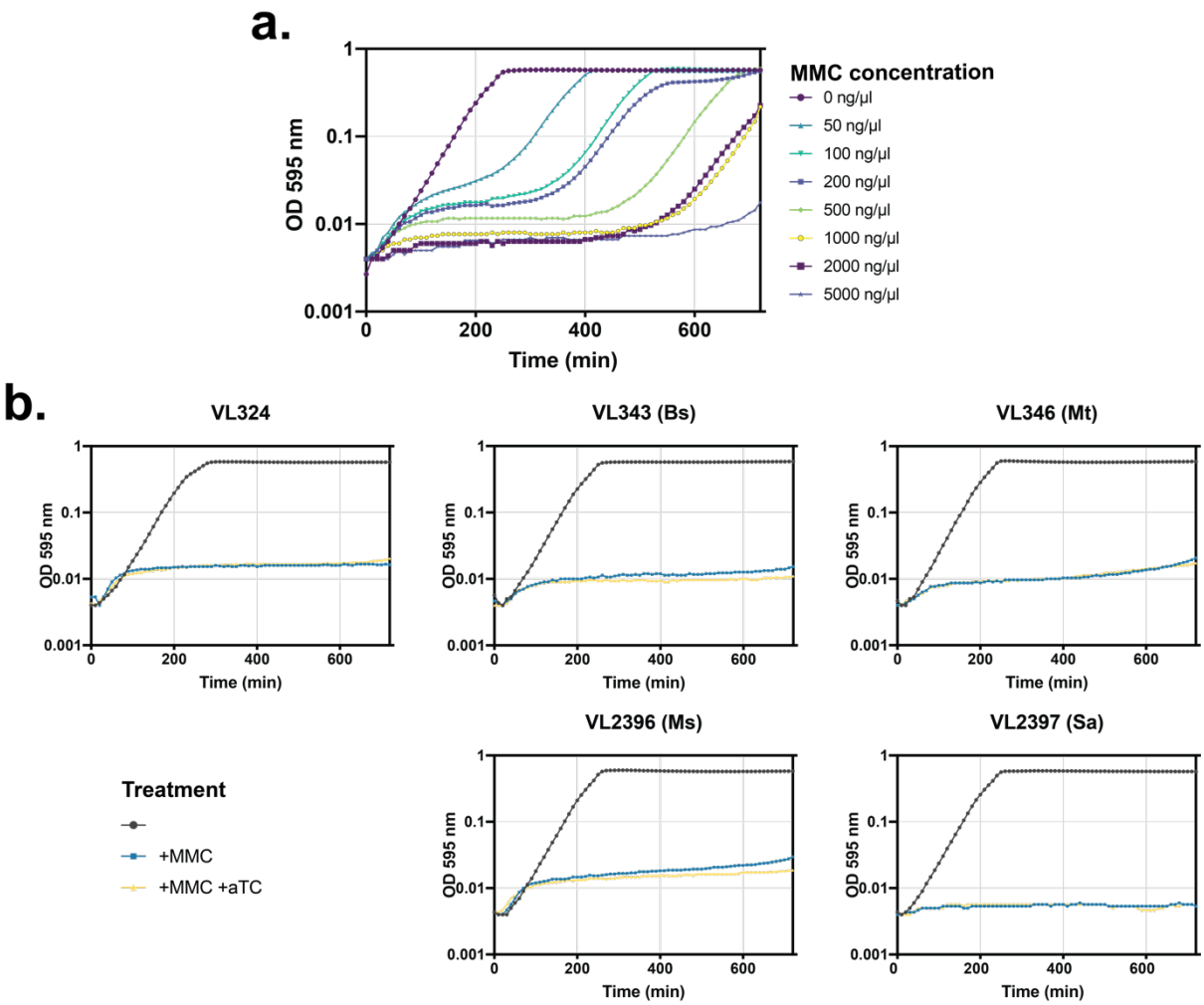


Figure 5: *S. pneumoniae* incubated with Mitomycin C (MMC). Cell density (OD595) of the bacterial cultures were measured every 10 min. The values represent averages of three replicates. Experiments were repeated at least 3 times. A representative growth curve is shown. (a) Strain VL324 (dnaA-dnaN-luc (ery); CEP::P3_sgRNAluc1 (spec); prs1::lacI-tetR (Gm); bgaA::Plac_cas9 (tet)) incubated with different concentrations of MMC. (b) . Control strain with no NHEJ system (strain VL324), NJEJ system of B. subtilis (strain VL343), NHEJ system of M. tuberculosis (strain VL344), NHEJ system of M. smegmatis (strain VL2396), and NHEJ system of S. ambofaciens (strain VL2397) were tested in three conditions. No MMC, induction with 5000 ng/ml of MMC and induction with 5000 ng/ml of MMC +500 ng/ml aTC. All strains were grown at 37°C.

Discussion

The main advantage a combined CRISPR-Cas9/NHEJ system would have over a CRISPR-Cas9/HR system, is that the user does not need to provide a HR template to repair the DSB. Thus, simply by a single transformation with a (dual) sgRNA construct, one could introduce indels or large genomic deletions, in contrast to HR-dependent DSB repair where one would need to first introduce a sgRNA followed by introduction of the HR template in the second step (Figure S1). To setup such a revolutionary system for bacteria, we transplanted four different NHEJ systems from other bacteria to *S. pneumoniae.* The NJEJ system of *Bacillus subtilis* (BsNHEJ), since the organisms are both gram-positive and belong to the same class of bacilli, we considered it a good candidate. However, it did not seem to be functional in pneumococcus. Next, we tested the NHEJ system of *Mycobacterium tuberculosis* (MtNHEJ), which also seemed to be a good candidate since a previous study has shown it can be functional when transferred in another organism. In particular, its ability to join incompatible DSB ends *in vitro*, as well as to reconstitute NHEJ *in vivo* in mutant yeast (Della *et al.*, 2004). Yet, it was not functional either. Following, we tested the NHEJ system of *Mycobacterium smegmatis* (MsNHEJ). Previous studies in *E. coli* supported that that MsNHEJ is more efficient in repair of DSBs than BsNHEJ and MtNHEJ (Zheng *et al.*, 2017), however MsNHEJ was also not functional in our assays. Last, we tested the NHEJ system of *Streptomyces ambofacience* (SaNHEJ) (Hoff *et al.*, 2016), but without any success. It is noteworthy, that to exclude the hypothesis that the NHEJ was seamlessly repairing the DSB, therefore the Cas9 was continuously cutting the DNA, we also tested our setup with two sgRNAs that were binding in close proximity. This way we would eliminate the two sites of the DSB. However preliminary results (not shown) show that DSB was not repaired either.

Our inability to set up a functional NHEJ system in *S. pneumoniae* comes in agreement with a previous study which reveals that there is a negative association between NHEJ and CRISPR in prokaryotes and there is only one single case of co- occurrence of the NHEJ and CRISPR systems among 5563 bacterial genomes. Specifically, it has been shown that there is a possible incompatibility between NHEJ and CRISPR systems, because the Cas9 protein inhibits NHEJ repair (Bernheim *et al.*, 2017). Additionally, Cas9 irreversibly binds to DNA sites (Farasat and Salis, 2016), potentially inhibiting the NHEJ apparatus to bind to the site and ligate the break and effectively acts as a roadblock to bacterial Ku and LigD. To test this, we could construct a less stable Cas9 that contains an ssrA-tag at its C-terminus which will recruit it to the ClpP proteolytic system. Nevertheless, NHEJ work well in CRISPR-Cas9 edited eukaryotic cells (Ran *et al.*, 2013).

To examine whether another enzymatic DSB could be repaired by any of the transplanted NHEJ machineries, we introduced the breaks with the meganuclease I-SceI.

Although only a partial DSB was observed, we did not observe any improved survival when any of the NHEJ systems were induced.

Another approach we tried, was to perform the repair with engineered T4 ligases instead of the NHEJ proteins, without any success. Probably, the recruitment by the Ku proteins is necessary, in order to stimulate ligase activity. However, combining the T4 ligases with Ku is not promising, since previous studies have shown that Ku inhibits T4 DNA ligase activity (Weller *et al.*, 2002). It could also be that we need to combine the DNA ligases with a double sgRNA construct to ensure the sgRNA target sites are removed by the resulting double DSBs.

To this point, we could not successfully setup a NHEJ repair, by an enzymatic break, so we tested if a non-enzymatic break could be repaired, instead, but this approach was not successful either. This could be explained by the observation that the response of mycobacteria, which actually have a NHEJ system, to mitomycin C preferentially involves a RecA-dependent pathway (Korycka-Machala *et al.*, 2006). UV radiation could be an alternative for a non-enzymatic DSB, but it has been shown that the loss of the NHEJ components Ku and LigD had no effect on sensitivity to UV radiation (Stephanou *et al.*, 2007).

While the transplanted NHEJ systems were correctly integrated and showed the correct sequences in the pneumococcal genome, we did not test whether the proteins are actually produced upon induction. Future work should perform Western blotting to verify that the NHEJ proteins are well expressed in pneumococcus. However, since some of the systems (MsNHEJ and SaNHEJ) were codon optimized for expression in *S. pneumoniae* (see Methods) it is unlikely all four systems are not expressed properly upon induction. A more likely hypothesis is that for NHEJ to work, it requires other factors apart from the proteins Ku and LigD. It is possible that interactions with the native DNA replication and DNA repair machinery are required. To identify if this is the case, we could test the BsNHEJ system in *Bacillus* itself. We could construct a *Bacillus* strain with the CRISPR system and also express constitutively the NHEJ system, since naturally it gets expressed only during sporulation (Moeller *et al.*, 2007). This way we can see if we could achieve a DSB repair. Alternatively, we can try inducing competence at the same time as the NHEJ system, as several DNA repair enzymes are included in the competence regulon (Slager, Aprianto and Veening, 2019).

To sum up, after many different approaches, we were still not able to successfully perform a NHEJ repair, therefore we needed to utilize the other mechanism available to repair the break; the homologous recombination (HR) (Chapter 3 of this thesis).

Materials and Methods

Bacterial strains, transformations and growth conditions

All pneumococcal strains used in this study are derivatives of the serotype 2 *S. pneumoniae* strain D39V (Avery, Macleod and Mccarty, 1944; Slager, Aprianto and Veening, 2018). All the strains are shown in Table 1.

Table 1: Strain list

***S. pneumoniae* Strains**	**Relevant Genotype**	**Reference**
D39V	Serotype 2 strain, wild type	Avery et al 1944, Slager et al 2018, Public Health England, NCTC14078
VL323	*dnaA-dnaN-luc (ery);* CEP*::P3_sgRNAluc1 (spec); prs1::lacI-tetR (Gm)*	Lab collection
VL324	*dnaA-dnaN-luc* (*ery)*; CEP::*P3_sgRNAluc1* (*spec*); *prs1*::*lacI-tetR* (*Gm*); bgaA::Plac*_cas9* (*tet*)	This study
VL343	*dnaA-dnaN-luc* (*ery)*; CEP::*P3_sgRNAluc1* (*spec*); *prs1*::*lacI-tetR* (*Gm*); bgaA::Plac*_cas9* (*tet*); CIL::*Ptet_ykoVU* (*kan*)	This study
VL346	*dnaA-dnaN-luc* (*ery)*; CEP::*P3_sgRNAluc1* (*spec*); *prs1*::*lacI-tetR* (*Gm*); bgaA::Plac*_cas9* (*tet*); CIL::*Ptet_mtku-mtligD* (*kan*)	This study
VL2396	*dnaA-dnaN-luc* (*ery)*; CEP::*P3_sgRNAluc1* (*spec*); *prs1*::*lacI-tetR* (*Gm*); bgaA::Plac*_cas9* (*tet*); CIL::*Ptet_msku-msligD* (*kan*)	This study
VL2397	*dnaA-dnaN-luc* (*ery)*; CEP::*P3_sgRNAluc1* (*spec*); *prs1*::*lacI-tetR* (*Gm*); bgaA::Plac*_cas9* (*tet*); CIL::*Ptet_SakuA-SaligC-SaligD* (*kan*)	This study
VL347	*prs1*::*lacI-tetR* (*Gm*); *blpSRHC*::I-*SceI*-site-*ery*; CEP::*lacI – Plac*_2-I-*sceI* (*spec*)	Lab collection
VL348	*prs1*::*lacI-tetR* (*Gm*); *blpSRHC*::I-*SceI*-site-*ery*; CEP::*lacI – Plac*_2-I-*sceI* (*spec*); CIL::*Ptet_ykoVU* (*kan)*	This study
VL349	*prs1*::*lacI-tetR* (*Gm*); *blpSRHC*::I-*SceI*-site-*ery*; CEP::*lacI – Plac*_2-I-*sceI* (*spec*); CIL::*Ptet_mtku-mtligD* (*kan*)	This study
VL3867	*prs1*::*lacI-tetR* (*Gm*); *blpSRHC*::I-*SceI*-site-*ery*; CEP::*lacI – Plac*_2-I-*sceI* (*spec*); CIL::*Ptet_msku-msligD* (*kan*)	This study
VL3868	*prs1*::*lacI-tetR* (*Gm*); *blpSRHC*::I-*SceI*-site-*ery*; CEP::*lacI – Plac*_2-I-*sceI* (*spec*); CIL::*Ptet_SakuA-SaligC-SaligD* (*kan*)	This study

VL2774	*dnaA-dnaN-luc* (*ery)*; CEP::*P3_sgRNAluc1* (*spec*); *prs1*::*lacI-tetR* (*Gm*); bgaA::Plac_*cas9* (*tet*); CIL::*Ptet_p50-ligase* (*kan*)	This study
VL2775	*dnaA-dnaN-luc* (*ery)*; CEP::*P3_sgRNAluc1* (*spec*); *prs1*::*lacI-tetR* (*Gm*); bgaA::Plac_*cas9* (*tet*); CIL::*Ptet_sso7d-ligase* (*kan*)	This study
Plasmids		
	pJWV102-Plac_dCas9	Liu *et al.*, 2017
	pJWV102-PZn_dCas9	Lab collection
	pPEPY-Ptet	Lab collection

S. pneumoniae was grown at 37°C without shaking in liquid C+Y medium adapted from Adams and Roe (Martin *et al.*, 1995) and contained the following compounds: adenosine (68.2 µM), uridine (74.6 µM), L-asparagine (302 µM), L-cysteine (84.6 µM), L-glutamine (137 µM), L-tryptophan (26.8 µM), casein hydrolysate (4.56 g L-1), BSA (729 mg L-1), biotin (2.24 µM), nicotinic acid (4.44 µM), pyridoxine (3.10 µM), calcium pantothenate (4.59 µM), thiamin (1.73 µM), riboflavin (0.678 µM), choline (43.7 µM), CaCl2 (103 µM), K2HPO4 (44.5 mM), MgCl2 (2.24 mM), FeSO4 (1.64 µM), CuSO4 (1.82 µM), ZnSO4 (1.58 µM), MnCl2 (1.29 µM), glucose (10.1 mM), sodium pyruvate (2.48 mM), saccharose (861 µM), sodium acetate (22.2 mM) and yeast extract (2.28 g L-1).

Transformation

To transform *S. pneumoniae*, cells were grown in C+Y medium (pH 6.8) at 37 °C to an OD_{595} of 0.1. Then, cells were treated for 12 min at 37°C with synthetic CSP-1 (100 ng mL^{-1}) and incubated for 20 min at 30°C with the plasmid. After incubation, cells were grown in C+Y medium at the permissive temperature of 30°C for 120 min. *S. pneumoniae* transformants were selected by plating inside Columbia agar supplemented with 3% of defibrinated sheep blood (Thermo Scientific). Depending on the construct created, the following concentrations of antibiotics were used for antibiotic selection: tetracycline (tet) at 1 µg/ml, spectinomycin (spec) at 100 µg/ml, kanamycin (kan) at 250 µg/ml, erythromycin (ery) at 1 µg/ml, gentamycin (gm) at 80 µg/ml and chloramphenicol (cam) at 4 µg/ml. Plates were incubated at 37°C.

Correct transformation was verified by PCR and sequencing. Working stocks of cells were prepared by growing cells in C+Y (pH 6.8), until an OD_{595} of 0.4. Cells were collected by centrifugation (1595 × *g* for 10 min) and resuspended in fresh C+Y medium with 15% glycerol and stored at −80°C.

Recombinant DNA techniques

Oligonucleotides were ordered from Sigma and are listed in Table 2. Phanta Max Super-Fidelity DNA Polymerase (Vazyme) was used in PCR amplifications, restriction enzymes (New England Biolabs) were used for digestions and T4 DNA Ligase (Vazyme) was used for ligations.

Table 2: Oligonucleotides used in this study

Name	Sequence (5′–3′); restriction enzyme recognition site (underlined)
pLac-F-AflII	TGGACACTTAAGCTCTAGAC
pLac-R-NotI	TTTTGAATTCGCGGCCGCC
EcorI+BglII+rbs-ykoV-F	GGCCGAATTC AGATCTAGGAGGTACACATGAATCGTACTCCTTCTC
ykoU-R+XhoI	GGCCCTCGAGAGTCAGCTCTTTTTCTTCAACTGATG
PEPY_homoregion-F	CCTTACTAAAGTATATAATTTAGGCAGTCAATTCTGTATG
PEPY_homoregion-R	CATTCATCATAACCCCCAGAACTTAAATAAC
t5-R+rbs+SapI	GTCGAAGGCTCTTCGTTTTTTTTCCTCCTAGATCTACTCTATC
mku-F+SapI	GTCGAAGGCTCTTCCAAATGCGAGCCATTTGGACGGGTTCG
mku-R+SapI	GTCGAAGGCTCTTCCCCCATTTTTTTTCCTCCTTCAACTGGCTGTC ACGGAGGCGTTG
ligD-F+SapI	GTCGAAGGCTCTTCGGGGTTCGGCGTCGGAGCAACGGGTG
ligD-R+SapI	GTCGAAGGCTCTTCGTCATTCGCGCACCACCTCACTGGG
T_B1006-F+SapI	GTCGAAGGCTCTTCGTGATCTCGAGAAAAAAAAACCGCGCCCC
OVL118_(msku)-t5-R	AGTATGGCGTACCGCACGGTTCATTTTTTTCCTCCTAGAT CTACTCTATC
OVL119_(t5)-msku-F	ATAGAGTAGATCTAGGAGGAAAAAAATGAACCGTGCGGT ACGCCATACTG
OVL120_(msligD)-msku-R	CATAGCGCTCCATTTTTTTCCTCCTCCCTGAGTTACGACG ATCTACGAC
OVL121_(msku)-msligD-F	TCGTAACTCAGGGAGGAGGAAAAAAATGGAGCGCTATGA GCGGGTTCGCC
OVL122_(ter)-msligD-R	GGTTTTTTTTTCTCGAGAGAGATCTCTATTCCCACACAAC CTCATCGGGT

OVL123_(msligD)-ter-F	ACCCGATGAGGTTGTGTGGGAATAGAGATCTCTCTCGAGAAAAAAAAACC
OVL1604_pPEPY-PtetF	GATCTGCACCTGCCTGTATCTAAATAAATAAGGAGGAAAATTAATGC
OVL1605_pPEPY-PtetR	GATCTGCACCTGCCTGTAAACTCTATCAATGATAGAGTTATTATACTC
OVL1606_Sa_ligC F	GATCTGCACCTGCCGAAAGGAGGC
OVL1607_Sa_ligD R	GATCTGCACCTGCTGGAAGATCTTTATTAGACC
OVL1608_Sa_kuA F	GATCTGCACCTGCTCCAGTTTAAGGAG
OVL1609_Sa_kuA R	GATCTGCACCTGCTCCATCCTTAATTATTAAGC
OVL2780_82A-F	CTCTTCCGTCTCGAGTGTAAGCAATTCTGCAGTACTAG
OVL2781_82A-R	CTCTTCCGTCTCGCTTTTATTTATTTAGATCTACTCTATCAATGATAGAG
OVL2782_82B-F	CTCTTCCGTCTCCAAAGAGGAGAAATTAACTATGAGAG
OVL2783_82B-R	CTCTTCCGTCTCCCACTTTTTATAGACCAGTTACCTCATGAAAATC
OVL2784_83A-F	GAACCTCGTCTCGAGTGTAAGCAATTCTGCAGTACTAG
OVL2785_83A-R	GAACCTCGTCTCGCTTTTATTTATTTAGATCTACTCTATCAATGATAGA
OVL2786_83B-F	GAACCTCGTCTCCAAAGAGGAGAAATTAACTATGAGAG
OVL2787_83B-R	GAACCTCGTCTCCCACTTTTTATAGACCAGTTACCTCATGAAAA

Strain construction

VL324 (*dnaA-dnaN-luc* (*ery)*; CEP::*P3_sgRNAluc1* (*spec*); *prs1*::*lacI-tetR* (*Gm*); bgaA::Plac*_cas9* (*tet*)). Plasmid *pJWV102-PZn_dCas9* was digested with restriction enzymes *NotI* and *AflII* and the backbone without the *PZn* promoter was isolated from gel (Fragment 1). The *Plac* promoter was amplified from plasmid *pJWV102-Plac_dCas9* with primers pLac-F-AflII and pLac-R-NotI and the PCR product was digested with restriction enzymes *NotI* and *AflII* (Fragment 2). The two fragments were ligated together, and the ligation product transformed into VL323 and transformants were selected were selected inside Columbia blood agar with tetracycline. Correct assembly was confirmed by PCR and sequencing.

VL343 (*dnaA-dnaN-luc (ery);* CEP*::P3_sgRNAluc1 (spec); prs1::lacI-tetR (Gm); bgaA::Plac_cas9 (tet);* CIL*::Ptet_ykoVU (kan)*). Plasmid p*PEPY-Ptet* was digested with restriction enzymes *BglII* and *XhoI* and the backbone was isolated from gel (Fragment 1). The *ykoVU* operon was amplified from genomic DNA of *Bacillus subtillis* of the lab collection, with primers EcorI+BglII+rbs-ykoV-F and ykoU-R+XhoI and the PCR product was digested with restriction enzymes *BglII* and *XhoI* (Fragment 2). The two fragments were ligated together, and the ligation product transformed into VL324 and transformants were selected were selected inside Columbia blood agar with kanamycin. Correct assembly was confirmed by PCR and sequencing.

VL346 (*dnaA-dnaN-luc* (*ery)*; CEP::*P3_sgRNAluc1* (*spec*); *prs1*::*lacI-tetR* (*Gm*); bgaA::Plac*_cas9* (*tet*); CIL::*Ptet_mtku-mtligD* (*kan*)). The downstream region of the CIL locus was amplified from VL343 with primers PEPY_homoregion-R and t5-R+rbs+SapI (Fragment 1). The *mtku* gene was amplified from genomic DNA of *Mycobacterium tuberculosis* of the lab collection, with primers mku-F+SapI and mku-R+SapI (Fragment 2). The *mtligD* gene was amplified from genomic DNA of *Mycobacterium tuberculosis* of the lab collection with primers ligD-F+SapI and ligD-R+SapI (Fragment 3). The upstream region of the CIL locus and the kanamycin resistance gene were amplified from VL343 with primers PEPY_homoregion-F and T_B1006-F+SapI (Fragment 4). The four fragments were digested all together with restriction enzyme *SapI* and ligated. The ligation product was transformed into VL324 and transformants were selected inside Columbia blood agar with kanamycin. Correct assembly was confirmed by PCR and sequencing.

VL2396 (*dnaA-dnaN-luc* (*ery)*; CEP::*P3_sgRNAluc1* (*spec*); *prs1*::*lacI-tetR* (*Gm*); bgaA::Plac*_cas9* (*tet*); CIL::*Ptet_msku-msligD* (*kan*)). The downstream region of the CIL locus was amplified from VL343 with primers PEPY_homoregion-R and OVL118_(msku)-t5-R (Fragment 1). The *msku* gene was amplified from a gBlock, with primers OVL121_(msku)-msligD-F and OVL120_(msligD)-msku-R (Fragment 2). The *mtligD* gene was amplified from

a gBlock with primers OVL121_(msku)-msligD-F and OVL122_(ter)-msligD-R (Fragment 3). The upstream region of the CIL locus and the kanamycin resistance gene were amplified from VL343 with primers PEPY_homoregion-F and OVL123_(msligD)-ter-F (Fragment 4). The four fragments were assembled together with Gibson Assembly. The ligation product was transformed into VL324 and transformants were selected inside Columbia blood agar with kanamycin. Correct assembly was confirmed by PCR and sequencing.

VL2397 (*dnaA-dnaN-luc* (*ery)*; CEP::*P3_sgRNAluc1* (*spec*); *prs1*::*lacI-tetR* (*Gm*); bgaA::Plac_*cas9* (*tet*); CIL::*Ptet_Saku-SaligC-SaligD* (*kan*)). The backbone of the plasmid *pPEPY-Ptet* was amplified with primers OVL1604_pPEPY-PtetF and OVL1605_pPEPY-PtetR (Fragment 1). The *Saku* gene was amplified from a gBlock, with primers OVL1608_Sa_kuA F and OVL1609_Sa_kuA R (Fragment 2). The genes *SaligC* and *SaligD* were amplified from a gBlock with primers OVL1606_Sa_ligC F and OVL1607_Sa_ligD R (Fragment 3). The three fragments were assembled together with restriction enzyme *AarI* and ligated. The ligation product was transformed into VL324 and transformants were selected inside Columbia blood agar with kanamycin. Correct assembly was confirmed by PCR and sequencing

VL348 (*prs1*::*lacI-tetR* (*Gm*); *blpSRHC*::I-*SceI*-site-*ery*; CEP::*lacI – Plac*_2-I-*sceI* (*spec*); CIL::*Ptet_ykoVU* (*kan)).* CIL locus from VL343 was amplified with PEPY_homoregion-F and PEPY_homoregion-R and transformed into VL347. Transformants were selected inside Columbia blood agar with kanamycin. Correct assembly was confirmed by PCR and sequencing

VL349 (*prs1*::*lacI-tetR* (*Gm*); *blpSRHC*::I-*SceI*-site-*ery*; CEP::*lacI – Plac*_2-I-*sceI* (*spec*); CIL::*Ptet_mtku-mtligD* (*kan)).* CIL locus from VL346 was amplified with PEPY_homoregion-F and PEPY_homoregion-R and transformed into VL347. Transformants were selected inside Columbia blood agar with kanamycin. Correct assembly was confirmed by PCR and sequencing

VL3867 (*prs1*::*lacI-tetR* (*Gm*); *blpSRHC*::I-*SceI*-site-*ery*; CEP::*lacI – Plac*_2-I-*sceI* (*spec*); CIL::*Ptet_msku-msligD* (*kan)).* CIL locus from VL2396 was amplified with PEPY_homoregion-F and PEPY_homoregion-R and transformed into VL347. Transformants were selected inside Columbia blood agar with kanamycin. Correct assembly was confirmed by PCR and sequencing

VL3868 (*prs1*::*lacI-tetR* (*Gm*); *blpSRHC*::I-*SceI*-site-*ery*; CEP::*lacI – Plac*_2-I-*sceI* (*spec*); CIL::*Ptet_Saku-SaligC-SaligD* (*kan)).* CIL locus from VL2397 was amplified with PEPY_homoregion-F and PEPY_homoregion-R and transformed into VL347. Transformants were selected inside Columbia blood agar with kanamycin. Correct assembly was confirmed by PCR and sequencing

VL2774 (*dnaA-dnaN-luc* (*ery)*; CEP::*P3_sgRNAluc1* (*spec*); *prs1*::*lacI-tetR* (*Gm*); bgaA::Plac_*cas9* (*tet*); CIL::*Ptet_p50-ligase* (*kan*)) The backbone of the plasmid *pPEPY-Ptet* was amplified with primers OVL2780_82A-F and OVL2781_82A-R (Fragment 1). p50-ligase was amplified from pCA24N-p50-ligase (Addgene #87742) with primers OVL2782_82B-F and OVL2783_82B-R (Fragment 2). The two fragments were digested together with restriction enzyme *BsmBI* and ligated. The ligation product was transformed into VL324 and transformants were selected inside Columbia blood agar with kanamycin. Correct assembly was confirmed by PCR and sequencing.

VL2775 (*dnaA-dnaN-luc* (*ery)*; CEP::*P3_sgRNAluc1* (*spec*); *prs1*::*lacI-tetR* (*Gm*); bgaA::Plac_*cas9* (*tet*); CIL::*Ptet_sso7d-ligase* (*kan*)). The backbone of the plasmid *pPEPY-Ptet* was amplified with primers OVL2784_83A-F and OVL2785_83A-R (Fragment 1). sso7d -ligase was amplified from pCA24N-sso7d-ligase (Addgene #87745) with primers OVL2786_83B-F and OVL2787_83B-R (Fragment 2). The two fragments were digested together with restriction enzyme *BsmBI* and ligated. The ligation product was transformed into VL324 and transformants were selected inside Columbia blood agar with kanamycin. Correct assembly was confirmed by PCR and sequencing.

Microtiter Plate Reader Assays

Cells were pre-cultured in acid C+Y (pH 6.8) at 37 °C to an $OD_{595\,nm}$ of 0.4. Right before inoculation, they were collected by centrifugation (8000 rpm for 3 min) and resuspended in fresh C+Y. Unless indicated otherwise all experiments were started with an inoculation density of OD_{595} 0.004 with cells from a preculture. Luciferase assays were performed in Costar 96-well plates (white, clear bottom) with a total assay volume of 300 µl per well with a Tecan Infinite 200 PRO luminometer at 37 °C. D-Luciferin Firefly (Biosynth) was added at a concentration of 0.1 mg mL^{-1} to monitor the integrity of the gene by means of luciferase activity. Optical density ($OD_{595\,nm}$) and luminescence (relative luminescence units (RLU)) were measured every 10 min with 25 flashes, luminescence measurement with an integration time of 1 s.

References

Aravind, L. and Koonin, E. V. (2001) 'Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system', *Genome Research*. Cold Spring Harbor Laboratory Press, 11(8), pp. 1365–1374. doi: 10.1101/gr.181001.

Avery, O. T., Macleod, C. M. and Mccarty, M. (1944) 'Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus Type III', *The Journal of experimental medicine*, 79(2), pp. 137–58. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2135445&tool=pmcentrez&rendertype=abstract (Accessed: 16 October 2014).

Bernheim, A. *et al.* (2017) 'Inhibition of NHEJ repair by type II-A CRISPR-Cas systems in bacteria', *Nature Communications*. Nature Publishing Group, 8(1), p. 2094. doi: 10.1038/s41467-017-02350-1.

Burby, P. E. and Simmons, L. A. (2019) 'A bacterial DNA repair pathway specific to a natural antibiotic', *Molecular Microbiology*. Blackwell Publishing Ltd, 111(2), pp. 338–353. doi: 10.1111/mmi.14158.

Daley, J. M. *et al.* (2005) 'Nonhomologous end joining in yeast.', *Annu Rev Genet*, 39, pp. 431–451. doi: 10.1146/annurev.genet.39.073003.113340.

Della, M. *et al.* (2004) 'Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine.', *Science (New York, NY)*, 306(5696), pp. 683–685. doi: 10.1126/science.1099824.

Farasat, I. and Salis, H. M. (2016) 'A Biophysical Model of CRISPR/Cas9 Activity for Rational Design of Genome Editing and Gene Regulation', *PLoS Computational Biology*. Public Library of Science, 12(1). doi: 10.1371/journal.pcbi.1004724.

Gong, C. *et al.* (2005) 'Mechanism of nonhomologous end-joining in mycobacteria: A low-fidelity repair system driven by Ku, ligase D and ligase C', *Nature Structural and Molecular Biology*. Nat Struct Mol Biol, 12(4), pp. 304–312. doi: 10.1038/nsmb915.

Hoff, G. *et al.* (2016) 'Multiple and Variable NHEJ-Like Genes Are Involved in Resistance to DNA Damage in *Streptomyces ambofaciens*', *Frontiers in Microbiology*. Frontiers, 7, p. 1901. doi: 10.3389/fmicb.2016.01901.

Jacquier, A. and Dujon, B. (1985) 'An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene', *Cell*. Cell Press, 41(2), pp. 383–394. doi: 10.1016/S0092-8674(85)80011-8.

Jiang, W. and Marraffini, L. a. (2015) 'CRISPR-Cas: New Tools for Genetic Manipulations from Bacterial Immunity Systems', *Annual Review of Microbiology*, 69(1), p.

150724172101001. doi: 10.1146/annurev-micro-091014-104441.

Korycka-Machala, M. *et al.* (2006) 'Distinct DNA repair pathways involving RecA and nonhomologous end joining in *Mycobacterium smegmatis*', *FEMS Microbiology Letters*. Oxford Academic, 258(1), pp. 83–91. doi: 10.1111/j.1574-6968.2006.00199.x.

Krejci, L. *et al.* (2003) 'Mending the Break: Two DNA Double-Strand Break Repair Machines in Eukaryotes', *Progress in Nucleic Acid Research and Molecular Biology*. Academic Press, 74, pp. 159–201. doi: 10.1016/S0079-6603(03)01013-4.

Lees-Miller, S. P. and Meek, K. (2003) 'Repair of DNA double strand breaks by non-homologous end joining', *Biochimie*. Elsevier, pp. 1161–1173. doi: 10.1016/j.biochi.2003.10.011.

Liu, X. *et al.* (2017) ' High-throughput CRISPRi phenotyping identifies new essential genes in *Streptococcus pneumoniae* ', *Molecular Systems Biology*. EMBO, 13(5), p. 931. doi: 10.15252/msb.20167449.

Martin, B. *et al.* (1995) 'The recA gene of *Streptococcus pneumoniae* is part of a competence-induced operon and controls lysogenic induction.', *Molecular microbiology*, 15(2), pp. 367–79. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7538190 (Accessed: 27 October 2014).

Moeller, R. *et al.* (2007) 'Role of DNA repair by nonhomologous-end joining in *Bacillus subtilis* spore resistance to extreme dryness, mono- and polychromatic UV, and ionizing radiation.', *Journal of bacteriology*, 189(8), pp. 3306–11. doi: 10.1128/JB.00018-07.

Pitcher, R. S. *et al.* (2007) 'NHEJ protects mycobacteria in stationary phase against the harmful effects of desiccation', *DNA Repair*. Elsevier, 6(9), pp. 1271–1276. doi: 10.1016/j.dnarep.2007.02.009.

Pitcher, R. S., Brissett, N. C. and Doherty, A. J. (2007) 'Nonhomologous end-joining in bacteria: a microbial perspective.', *Annual review of microbiology*, 61, pp. 259–282. doi: 10.1146/annurev.micro.61.080706.093354.

Ran, F. A. *et al.* (2013) 'Genome engineering using the CRISPR-Cas9 system', *Nature Protocols*. Nature Publishing Group, 8(11), pp. 2281–2308. doi: 10.1038/nprot.2013.143.

Shuman, S. and Glickman, M. S. (2007) 'Bacterial DNA repair by non-homologous end joining', *Nature Reviews Microbiology*, 5(11), pp. 852–861. doi: 10.1038/nrmicro1768.

Slager, J., Aprianto, R. and Veening, J. W. (2018) 'Deep genome annotation of the opportunistic human pathogen *Streptococcus pneumoniae* D39', *Nucleic Acids Research*. Oxford University Press, 46(19), pp. 9971–9989. doi: 10.1093/nar/gky725.

Slager, J., Aprianto, R. and Veening, J. W. (2019) 'Refining the pneumococcal competence regulon by RNA sequencing', *Journal of Bacteriology*. American Society for Microbiology, 201(13). doi: 10.1128/JB.00780-18.

Sorg, R. A. *et al.* (2020) 'Synthetic gene-regulatory networks in the opportunistic human

pathogen *Streptococcus pneumoniae*', *Proceedings of the National Academy of Sciences*. National Academy of Sciences, p. 201920015. doi: 10.1073/pnas.1920015117.

Stephanou, N. C. *et al.* (2007) 'Mycobacterial nonhomologous end joining mediates mutagenic repair of chromosomal double-strand DNA breaks.', *Journal of bacteriology*. American Society for Microbiology Journals, 189(14), pp. 5237–46. doi: 10.1128/JB.00332-07.

de Vega, M. (2013) 'The Minimal *Bacillus subtilis* Nonhomologous End Joining Repair Machinery', *PLoS ONE*. Public Library of Science, 8(5), p. e64232. doi: 10.1371/journal.pone.0064232.

Walker, J. R., Corpina, R. A. and Goldberg, J. (2001) 'Structure of the Ku heterodimer bound to dna and its implications for double-strand break repair', *Nature*. Nature, 412(6847), pp. 607–614. doi: 10.1038/35088000.

Weller, G. R. *et al.* (2002) 'Identification of a DNA nonhomologous end-joining complex in bacteria', *Science*. Science, 297(5587), pp. 1686–1689. doi: 10.1126/science.1074584.

Weterings, E. and Chen, D. J. (2008) 'The endless tale of non-homologous end-joining.', *Cell research*, 18(1), pp. 114–124. doi: 10.1038/cr.2008.3.

Wilson, R. H. *et al.* (2013) 'Engineered DNA ligases with improved activities in vitro', *Protein Engineering, Design and Selection*. Protein Eng Des Sel, 26(7), pp. 471–478. doi: 10.1093/protein/gzt024.

Zheng, X. *et al.* (2017) 'An efficient system for deletion of large DNA fragments in *Escherichia coli* via introduction of both Cas9 and the non-homologous end joining system from *Mycobacterium smegmatis*', *Biochemical and Biophysical Research Communications*. Elsevier Ltd, 485(4), pp. 768–774. doi: 10.1016/j.bbrc.2017.02.129.

Supplementary Figures

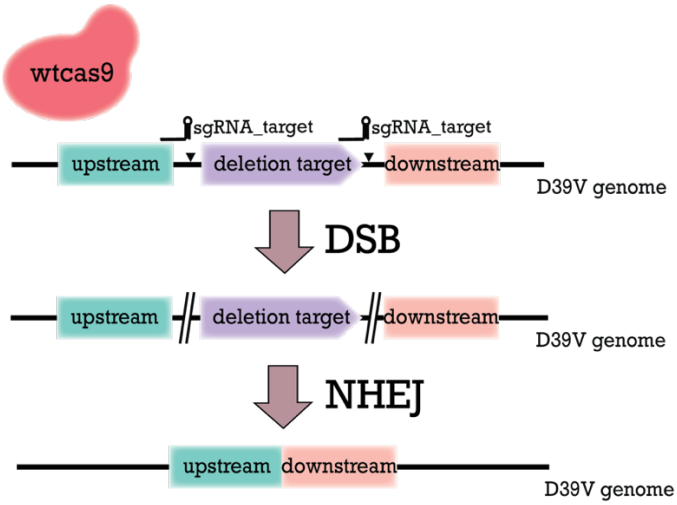


Figure S1: Suggested experimental approach for gene deletions using a combined CRISPR-Cas9 system and NHEJ system

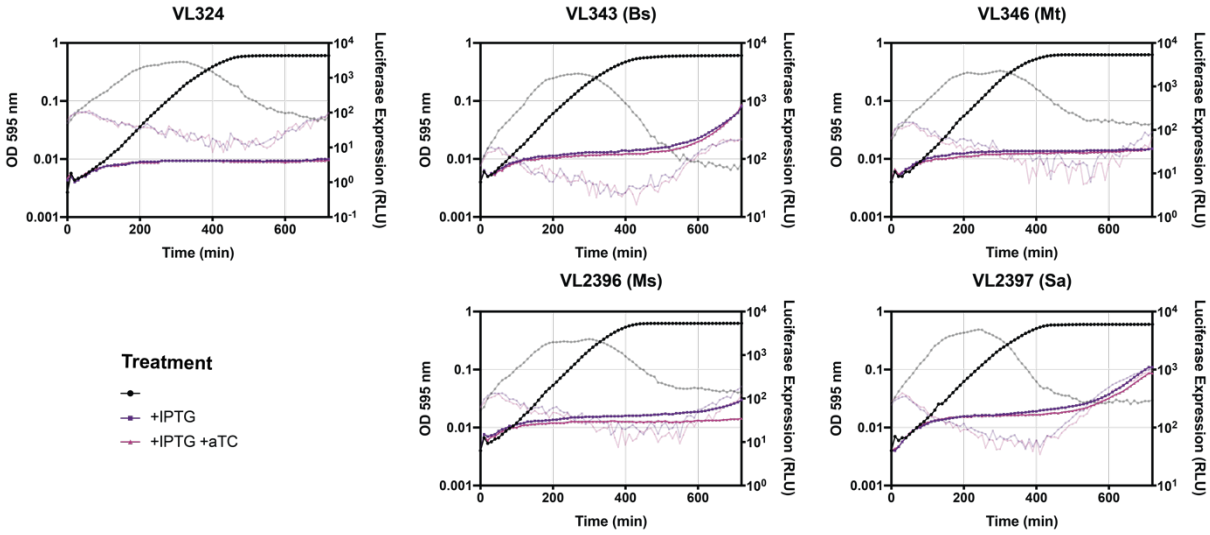


Figure S2*:* An IPTG-inducible CRISPR-cas9 system and an aTC inducible NHEJ system in S. pneumoniae. Cell density (OD595) (left y axis) and luciferase activity (shown as RLU) (right y axis) of the bacterial cultures were measured every 10 min. The values represent averages of three replicates. Experiments were repeated at least 3 times. A representative growth curve is shown. Control strain with no NHEJ system (strain VL324), NJEJ system of B. subtilis (strain VL343), NHEJ system of M. tuberculosis (strain VL344), NHEJ system of M. smegmatis (strain VL2396), and NHEJ system of S. ambofaciens (strain VL2397) were tested in three conditions. No induction, induction with 20 µM of IPTG and induction with 20 µM IPTG+500 ng/ml aTC. All strains have grown at 30°C.

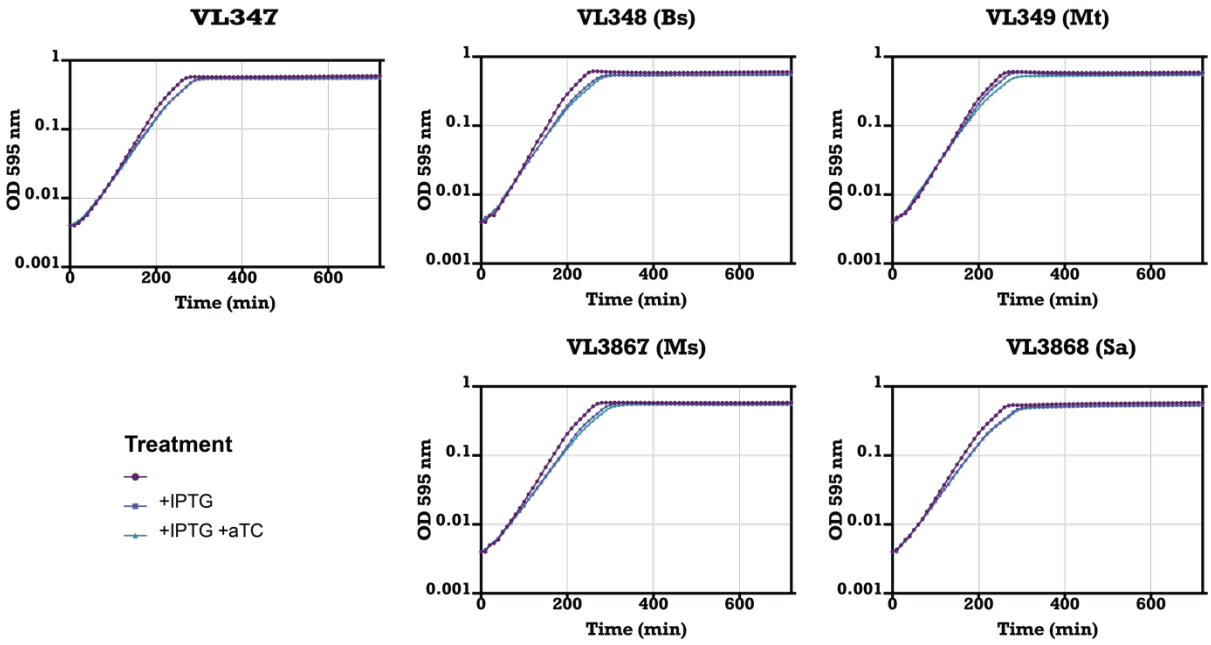


Figure S3: An IPTG-inducible I-SceI and an aTC inducible NHEJ system in *S. pneumoniae.* Cell density (OD595) (left y axis) and luciferase activity (shown as RLU) (right y axis) of the bacterial cultures were measured every 10 min. The values represent averages of three replicates. Experiments were repeated at least 3 times. A representative growth curve is shown. Control strain with no NHEJ system (strain VL347), NJEJ system of *B. subtilis* (strain VL348), NHEJ system of *M. tuberculosis* (strain VL349), NHEJ system of *M. smegmatis* (strain VL3867), and NHEJ system of *S. ambofaciens* (strain VL3868) were tested in three conditions. No induction, induction with 20 µM of IPTG and induction with 20 µM IPTG+500 ng/ml aTC. All strains have grown at 37°C.

Chapter 3: Harnessing CRISPR-Cas9 for genome editing in *Streptococcus pneumoniae* D39V

Authors: Dimitra Synefiaridou and Jan-Willem Veening

(D.S. designed research, performed experiments, analyzed data, and wrote this chapter;

published as Synefiaridou and Veening, Appl Environ Microbiol. 2021 Jan 4; AEM.02762-20. doi: 10.1128/AEM.02762-20)

Abstract

CRISPR-Cas systems provide bacteria and archaea with adaptive immunity against viruses and plasmids by detection and cleavage of invading foreign DNA. Modified versions of this system can be exploited as a biotechnological tool for precise genome editing at a targeted locus. Here, we developed a replicative plasmid that carries the CRISPR-Cas9 system for RNA-programmable, genome editing by counterselection in the opportunistic human pathogen *Streptococcus pneumoniae.* Specifically, we demonstrate an approach for making targeted, marker-less gene knockouts and large genome deletions. After a precise double-stranded break (DSB) is introduced, the cells' DNA repair mechanism of homology-directed repair (HDR) pathway is being exploited to select successful transformants. This is achieved through the transformation of a template DNA fragment that will recombine in the genome and eliminate recognition of the target of the Cas9 endonuclease. Next, the newly engineered strain can be easily cured from the plasmid that is temperature-sensitive for replication, by growing it at the non-permissive temperature. This allows for consecutive rounds of genome editing. Using this system, we engineered a strain with three major virulence factors deleted. The here developed approaches could be potentially transported to other Gram-positive bacteria.

Importance

Streptococcus pneumoniae (the pneumococcus) is an important opportunistic human pathogen killing over a million people each year. Having the availability of a system capable of easy genome editing would significantly facilitate drug discovery and efforts in identifying new vaccine candidates. Here, we introduced an easy-to-use system to perform multiple rounds of genome editing in the pneumococcus by putting the CRISPR-Cas9 system on a

temperature-sensitive replicative plasmid. The here used approaches will advance genome editing projects in this important human pathogen.

Introduction

Streptococcus pneumoniae (the pneumococcus) is a Gram-positive human commensal that colonizes asymptomatically the mucosal surfaces of the upper respiratory tract (UTR) (Kadioglu *et al.*, 2008). However, in susceptible groups like children, the elderly and the immunocompromised, it can occasionally cause disease, ranging from a mild upper respiratory tract infection, acute otitis media and sinusitis, to severe and potentially life-threatening conditions such as pneumonia, bacteremia and meningitis (Simell *et al.*, 2012). It is responsible for more than one million deaths annually (O'Brien *et al.*, 2009) and in 2017, the World Health Organization (WHO) classified *S. pneumoniae* as one of twelve priority pathogens for which new antibiotics are urgently needed.

Historically, *S. pneumoniae* research played a central role in advancing molecular biology. While trying to develop a vaccine against the pneumococcus, Griffith discovered natural transformation (Griffith, 1928). This was followed by research of Avery, MacLeod and McCarty to establish that DNA is the genetic material (Avery, Macleod and Mccarty, 1944). Over the last decade, the pneumococcus has become a valuable model to study the cell biology of ovoid-shaped bacteria and several cell biological tools such as integration vectors, fluorescent reporters, inducible promoters and CRISPR interference have been established for this organism (Massidda, Nováková and Vollmer, 2013; Liu *et al.*, 2017; Keller *et al.*, 2019). In addition, many selection and counterselection methods are available making it relatively easy to generate gene deletions, gene complementation mutants or point mutations in the pneumococcal genome (Sung *et al.*, 2001; Halfmann, Hakenbeck and Brückner, 2007; Li, Thompson and Lipsitch, 2014; Sorg *et al.*, 2020) . However, all current gene deletion methods established for *S. pneumoniae* are poorly scalable and often require a specific genetic background to function (e.g. the *rpsL*+ background in the janus system (Sung *et al.*, 2001).

In the case of gene replacement by selection markers, while powerful, this also has drawbacks, preventing further modifications of the genome when there are no further selectable markers available for additional strain development. Also, many important categories of gene mutation, such as missense substitutions and in-frame deletions, usually present no selectable phenotype (Sung *et al.*, 2001). To circumvent these issues, we here established CRISPR-Cas9 genome editing for use as counterselection in the pneumococcus.

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins are present in many bacteria and most archaea (Jansen *et al.*, 2002). Naturally, the system provides resistance against foreign genetic elements (e.g. phages or plasmids) via small noncoding RNAs that are derived from CRISPR loci (Barrangou

et al., 2007). In class 2 type II CRISPR systems, the mature crRNA that is base-paired to a trans-activating crRNA (tracrRNA) forms a two-RNA structure that directs the CRISPR-associated proteins (e.g. Cas9 from *Streptococcus pyogenes*) to introduce a double-stranded break (DSB) into the target DNA locus. Site-specific cleavage occurs at locations determined by both base-pairing complementarity between the crRNA and the target protospacer DNA and a short protospacer adjacent motif (PAM) (Jinek *et al.*, 2012). It has been demonstrated that the endonuclease can be programmed by engineering the mature dual-tracrRNA: crRNA as a single RNA chimera (sgRNA for single guide RNA), to cleave specific DNA sites. Thereby, modified versions of the system can be exploited as a biotechnological tool for precise, RNA-programmable genome targeting and editing (Jinek *et al.*, 2012).

After the DSB has been introduced by Cas9, the cell can utilize two major pathways in order to repair the break and survive: homologous recombination (HR) or non-homologous end-joining (NHEJ). In HR, a second intact copy of the broken chromosome segment, homologous to the DSB site, serves as a template for DNA synthesis across the break. In this mechanism, the crucial process of locating and recombining the homologous sequence is performed by RecA (Shuman and Glickman, 2007). NHEJ does not rely on a homologous DNA template, as the two DNA ends are rejoined directly together. Most bacteria such as *S. pneumoniae* cannot perform NHEJ, while it is capable to perform HR (Prudhomme, Libante and Claverys, 2002). DSB repair can be used as a way to generate mutants or desired changes to the genome by providing a HR template, and forms the basis of CRISPR-Cas engineering (Adli, 2018). Indeed, early work, using integrative vectors and tracrRNAs, showed that Cas9 can be used to make markerless gene deletions in *S. pneumoniae* (Jiang *et al.*, 2013).

In this study, we set out to establish a more concise CRISPR-Cas engineering framework for *S. pneumoniae* harnessing Cas9 from *S. pyogenes*. Specifically, we constructed a replicative plasmid containing a temperature-sensitive origin of replication (facilitating curing of the plasmid) carrying a genetic system for making targeted, marker-less gene knockouts and large genome deletions, which works with high efficiency in *S. pneumoniae*. The here developed plasmid system could potentially be transportable to other Gram-positive bacteria as the used origin of replication was shown to be functional in *L. lactis* and *B. subtilis* (Bijlsma *et al.*, 2007). While similar approaches have recently been undertaken to perform genome engineering in certain Gram-positive organisms such as *Enterococcus faecium* (de Maat *et al.*, 2019), Clostridium (Cañadas *et al.*, 2019), *Lactobacillus reuteri* (Oh and Van Pijkeren, 2014), *L. plantarum* (Zhou *et al.*, 2019) and *Lactococcus lactis* (Guo *et al.*, 2019), a concise CRISPR sgRNA:Cas9 gene editing system was not yet available for *S.*

pneumoniae and the here described vector has the advantage of being readily curable due to its temperature sensitive origin of replication.

Results

A *S. pneumoniae* replicative plasmid that carries the CRISPR-Cas9 system

First, a replicative plasmid was designed and constructed (Figure 1a), by combining PCR amplified genomic and plasmid parts. The main idea behind the choice for individual vector components relied in creating a platform with the CRISPR-Cas9 system in *S. pneumoniae* while at the same time allowing for plasmid propagation in both Gram-positive and Gram-negative hosts. The modular vector consists of six individual components. Two origins of replication; the high-copy pG$^+$host replicon, which is a replication thermosensitive derivative of pWV01 (Otto, de Vos and Gavrieli, 1982) that in *L. lactis*, (and other Gram-positive bacteria) replicates at 28°C but is lost above 37°C, and the low-copy *CloDF13* (CDF) replicon for propagation in *E. coli.* By combining these two origins of replication, it ensures low copy numbers at 37°C in *E. coli* thereby preventing toxicity of the CRISPR-Cas9 system while cloning. Additionally, it has the gene which encodes wtCas9 under the control of the Zinc-inducible promoter Pzn (Eberhardt *et al.*, 2009) (plasmid pDS05) and genes conferring spectinomycin (*E. coli*) and erythromycin (*S. pneumoniae*) resistance. Finally, it has the strong synthetic constitutive P3 promoter, which was optimized before specifically for *S. pneumoniae* (Sorg, Kuipers and Veening, 2015). Additionally, since analysis of the nucleotide distribution across all TSSs showed a strong preference for adenine as initiating nucleotide in *S. pneumoniae* (Slager, Aprianto and Veening, 2018), we have added an adenine as +1 for the sgRNA. The P3 promoter is driving the sgRNA sequence in which the 20 base-pairing region of the sgRNA is replaced by the *gfp* gene flanked by two *Bsa*I restriction sites. This allows for easy replacement of *gfp* by the spacer sequence of sgRNA with golden gate cloning. This way, successful cloning of the sgRNA allows for easy selection by absence of GFP fluorescence, giving us a versatile vector for different sgRNAs (see below).

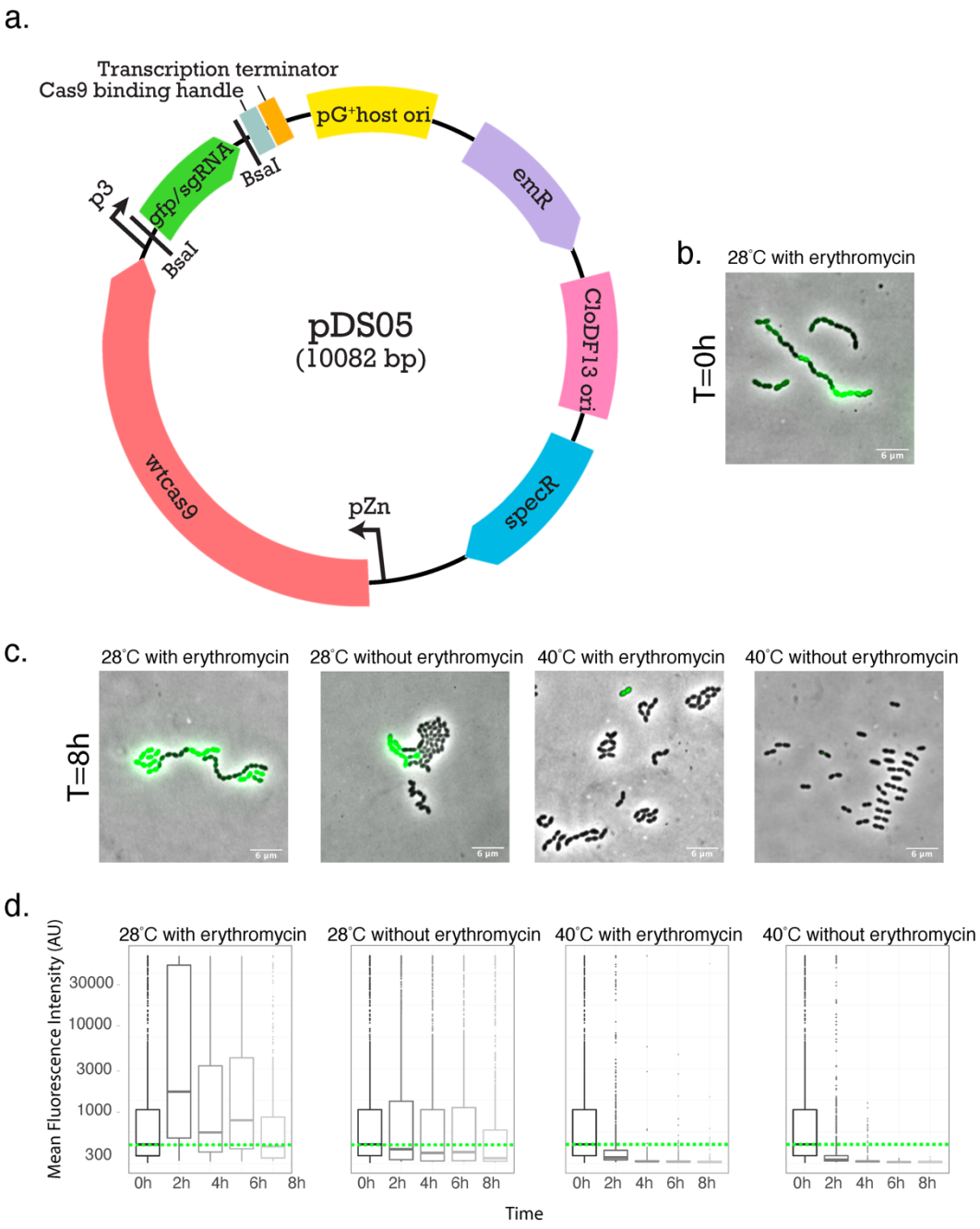


Figure 1: pDS05 is a temperature-sensitive plasmid that can be used for CRISPR-Cas9 genome editing in *S. pneumoniae*. (a) Schematic representation of plasmid pDS05. **(b)** Microscopy analysis of strain VL3655 (D39V, pDS05). Overlay of GFP signals with phase contrast image shows GFP expression. Note that the levels of fluorescence have been adjusted because of the large cell-to-cell variability in fluorescence (some cells appear dark but actually produce above background levels of fluorescence). Preculture grown at 28°C with erythromycin (T=0h), **(c)** Images are shown of cells grown for 8h as exponentially growing cells (balanced growth) in four different conditions: 28°C with erythromycin, 28°C without erythromycin, 40°C with erythromycin, and 40°C without erythromycin. Scale bar in all images = 6 µm. **(d)** Quantification of mean fluorescence intensity of GFP of cells grown under four different conditions: 28°C with erythromycin, 28°C without erythromycin, 40°C with

erythromycin, and 40°C without erythromycin in time points 0, 2, 4, 6 and 8 hours after dilution from the 28°C with erythromycin condition. Fluorescence microscopy of ±1000 cells per condition per time point were quantified and analyzed using MicrobeJ and BactMap and plotted as box plots, (box size and line represent the average intensity per cell) (see Materials and Methods). The green dotted horizontal line indicates the mean fluorescence of cells from the preculture harboring pDS05.

Successful plasmid propagation and plasmid curing in *S. pneumoniae*

To test that the newly constructed plasmid was being replicated and genes were expressed in *S. pneumoniae*, we grew strain VL3655 (carrying plasmid pDS05, see Figure 1a), which encodes GFP, in C+Y medium at 28°C in the presence of erythromycin. Fluorescence microscopy demonstrated that all cells produced GFP, although significant cell-to-cell variability was observed (Figure 1b). GFP intensity levels were determined in exponentially growing cells. Additionally, cells pre-grown at 28°C in presence of erythromycin (T=0) were split and grown under four different conditions. The permissive 28°C condition both with and without erythromycin in the growth medium and the non-permissive 40°C with and without erythromycin. Note that growth was balanced by re-diluting exponentially growing cells several times. Cells were being collected every two hours for 8 hours and GFP intensity levels were determined using fluorescence microscopy and images were analyzed using MicrobeJ and BactMAP (Ducret, Quardokus and Brun, 2016; Raaphorst, Kjos and Veening, 2020) (Figure 1c and d). The results show that GFP levels and, by extension plasmid copy number, stay stable at 28°C with erythromycin, and slowly decrease in the absence of antibiotic pressure. Furthermore, GFP levels decrease significantly in cells grown at 40°C, confirming that this is a non-permissive temperature for propagation of the plasmid. Absence of antibiotic pressure seems also to facilitate the decrease of the intensity levels of GFP, suggesting that the plasmid gets eliminated successfully under these conditions.

Since differences in plasmid copy number between cells can affect the levels of fluorescence, an additional time course experiment has been conducted, with cells growing in the same conditions as mentioned above, and we determined the cells that are resistant to erythromycin compared to the total viable cells by counting the colony forming units (CFUs) (Figure 1Sa). The results show that at the permissive temperature, most viable cells are erythromycin resistant, suggesting that they maintained the plasmid. Small discrepancies observed can be explained by the lack of separation of cells within the counted colony. At the non-permissive temperature, none of the cells were resistant, therefore they have lost the plasmid, supporting the previous microscopy findings. Finally, we also wanted to test whether Cas9 expression contributes to pDS05 plasmid instability. Therefore, we performed another fluorescence microscopy experiment (Figure 1Sb), at the permissive temperature, with and without antibiotic selection, and with and without Cas9 induction by Zn^{2+}. As shown in Figure

S1b, GFP levels and, by extension plasmid copy number, stay stable under all tested conditions suggesting that Cas9 induction in the absence of a targeting sgRNA does not cause plasmid loss.

CRISPR-Cas9-Mediated Counterselection

Once a deletion target has been selected, the plasmid with the specific sgRNA needs to be constructed. For the sgRNA design, several tools are available, but here the freely available online software Benchling was used [Biology Software] (2019). Benchling uses the latest algorithms to instantly assess off-target effects and on-target efficiency for guide RNAs. Via this platform we specify the guide parameters which are a single guide of 20bp length and an NGG PAM. Out of the list generated, sgRNAs without predicted mismatches and off targets were selected. As a second validation, the selected sgRNAs were evaluated using the R tool 'CRISPRi-seq target validation' (Liu *et al.*, 2021). The targeting of Cas9 to a locus of interest is achieved by cloning two annealed 24-bp DNA oligonucleotides (containing the 20 bp protospacer element) into the sgRNA backbone that matches the specified locus. First, *gfp* is removed from pDS05 by digesting with *Bsa*I. Complementary oligos that carry the spacer sequence are annealed together. They are designed in a way that after annealing, they have overhangs complementary with those left after digestion of the backbone, as previously described (Liu *et al.*, 2021) (Figure 2). The desired plasmid is obtained after ligation and transformed to *E. coli*. False positive transformants are easily identified, since they still carry the *gfp* and produce detectable fluorescent green colonies.

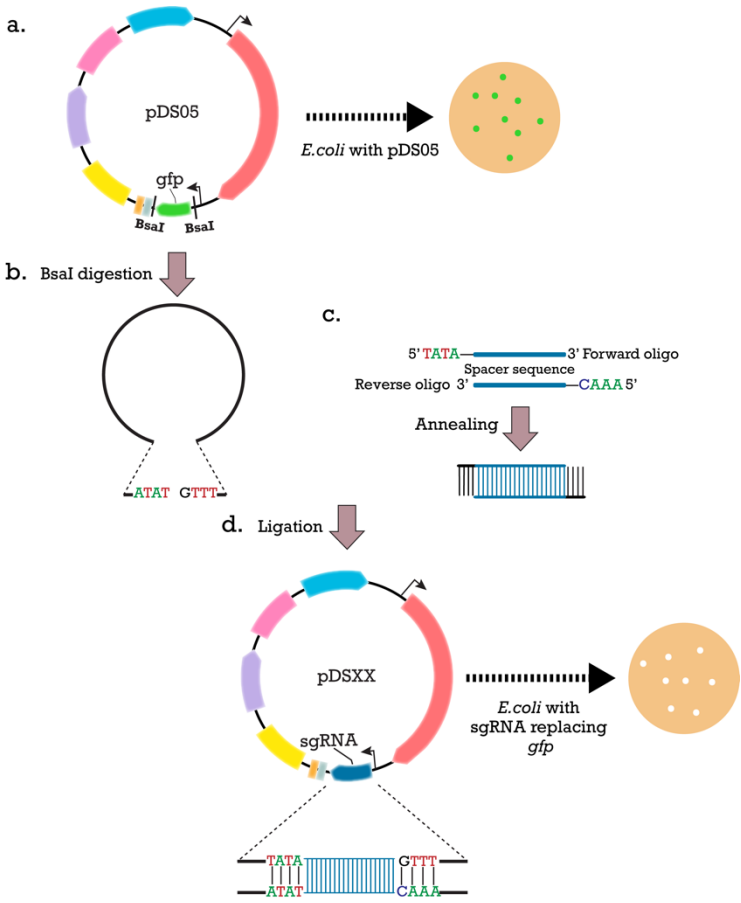


Figure 2: Workflow of sgRNA cloning. (a) pDS05 was designed to facilitate easy replacement of *gfp* by the spacer sequence of the desired sgRNA with golden gate cloning, allowing also for detection of false positive transformants. *gfp,* which encodes a green fluorescent protein, is in place of the spacer sequence of sgRNA and flanked by *Bsa*I sites. *E. coli* with pDS05 produces green fluorescent colonies **(b)** *Bsa*I digestion of the vector exposes 4 nt overhangs c. For each sgRNA, forward and reverse oligos were designed, as a reverse complement of each other, which after being annealed together, were containing the 20 bp spacer sequence and 4 nt overhangs, that can be specifically annealed with the digested vector. d. Ligation of the digested vector with the sgRNA annealed product was transformed into *E. coli*, producing white colonies.

After isolating the plasmid from *E. coli*, it needs to be used to transform *S. pneumoniae*. To achieve this, pneumococcal competence for transformation is induced by adding synthetic Competence Stimulating Peptide (CSP). Note that competence-dependent transformation with a replicative plasmid is less efficient than transforming with linear homologous DNA (Johnston *et al.*, 2014), so transformation efficiencies with pDS05 are typically low, with no evidence of differences between plasmids containing sham-sgRNA vs. sgRNA targeting the chromosome when Cas9 is not induced (Figure S2). Next, an HR template is constructed that consists of the upstream and downstream region of the deletion target. It has been shown that higher transformation efficiency is observed at higher donor DNA concentrations and with longer homology regions with a plateau of approximately 1000 bp of homology (Lee, Seok

and Morrison, 1998; Kurushima *et al.*, 2020). Therefore, each selected homology arm is designed to be at least 1000 bp, ensuring efficient homologous recombination. Following again induction of competence by CSP, the pneumococcal cells harboring the pDS05-derivative are transformed with this template. Transformants are selected by plating with $ZnCl_2$/$MnSO_4$, the inducer of Cas9, offering CRISPR-Cas9 -mediated counter-selection. Only cells that have taken up and integrated the HR template, thereby eliminated the recognition target of the sgRNA/Cas9 complex, would be able to survive, while untransformed cells will undergo DNA cleavage and die (Figure 3).

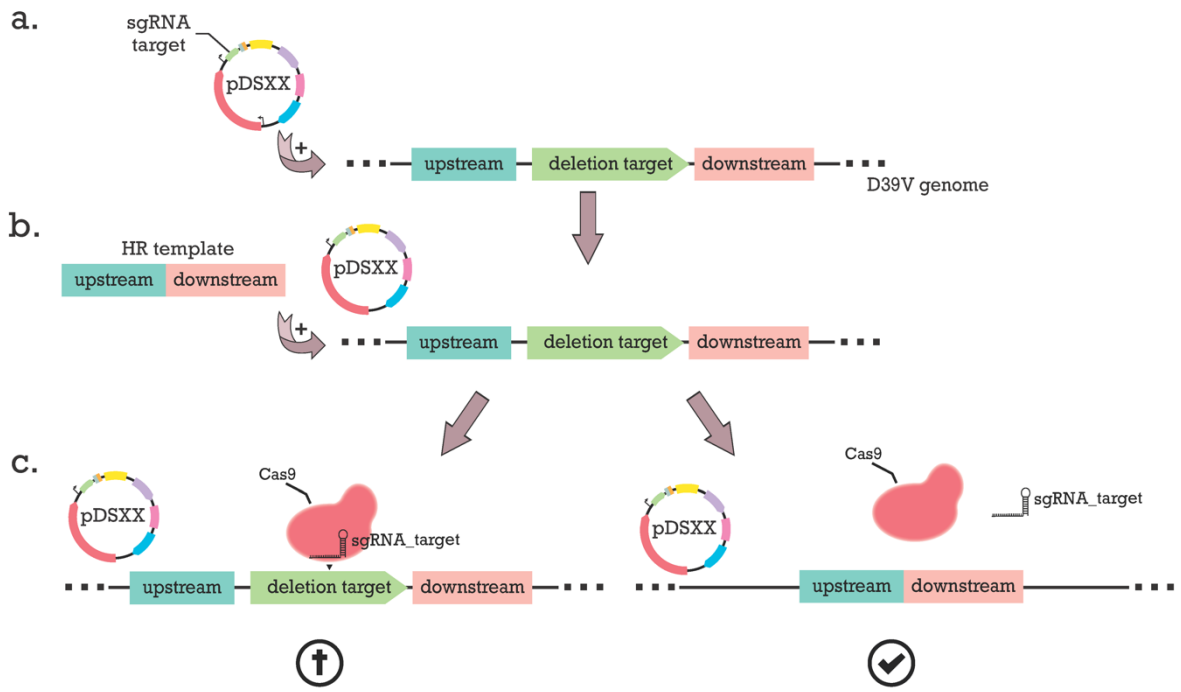


Figure 3: Workflow for markerless deletions. (a) Uptake of the plasmid by the strain with the sgRNA sequence for the desired deletion. **(b)** Transformation of a homology recombination (HR) template consisting of a fusion between the upstream and downstream region of the deletion target. **(c)** Plating transformants with Zn^{2+} to induce expression of Cas9. Only the cells that have taken up and integrated the HR template eliminating the recognition target are able to survive.

Deleting genes and large chromosomal regions from the *S. pneumoniae* genome

To assess the efficiency of the system, we first constructed a strain (strain VL3656; Figure 4a) in which we placed the *E. coli lacZ* gene under a constitutive promoter behind the *S. pneumoniae* D39V *cps* locus (encoding the capsule). *lacZ* encodes for a β-galactosidase that hydrolyzes X-gal to produce a blue product, allowing for 'blue/green' screening on plates. Colonies with blue color would still carry the *lacZ* gene, while colonies with the standard green (inside blood agar) color would indicate that the gene has been deleted from the chromosome.

Strain VL3656 also carries the pDS07 plasmid, which contains a sgRNA targeting *lacZ.* Next, we constructed an HR template that consisted of the 1000 bp upstream and 1000 bp downstream region of *lacZ* (excluding *lacZ*) and we transformed VL3656 with this template (Figure 4a). Transformants were selected by plating inside agar containing $ZnCl_2$/MnSO4 to induce expression of Cas9.

After transformation of strain VL3656 with the HR template, transformation efficiencies were calculated (Figure 4b). The CRISPR-Cas9 -mediated counter-selection, offered by the system, worked successfully. The selection efficiency was high, around 1.5×10^{-3}, and almost all colonies in the transformation where the HR template was given and Cas9 was induced had their original green color, indicating the *lacZ* gene has been successfully deleted.

Eight transformants were tested for correct deletion of *lacZ* by colony PCR. The primers used were binding 1000bp upstream and downstream of *lacZ*, setting the correct PCR product of the successful deletion at 2000bp (5000bp if *lacZ* was not deleted) (Figure S3). All tested colonies had the expected product of 2000bp, demonstrating successful deletion of *lacZ*, resulting in strain VL3657 (Figure S3). We note that typically selection of two or three transformants is sufficient since editing efficiency is over 99% (Figure 4).

Additionally, we also used the system to delete an even larger chromosomal fragment. For this, we targeted the operon that encodes the capsule and the *lacZ* gene that had been inserted downstream of it, which is around 24Kbp long, allowing for blue/green screening. Once again, selection efficiency was very high, approximately 4×10^{-3}, and almost all colonies had their original green color (Figure 4b). Colony PCR verified correct deletion of the *cps-lacZ* chromosomal region (see below).

Using the same HR template to delete *lacZ,* we also performed transformation assays without the counterselection offered, by inducing our CRISPR-Cas9 system, and we performed again a blue/green color colony screening (Figure 4c). More than 12.000 colonies needed to be screened to find three successful transformants with the original colony color, among the blue colonies. In contrast, by using the system, almost with absolute success rate, all the colonies on our plates were correct transformants, demonstrating how efficient our system is to easily select edited cells.

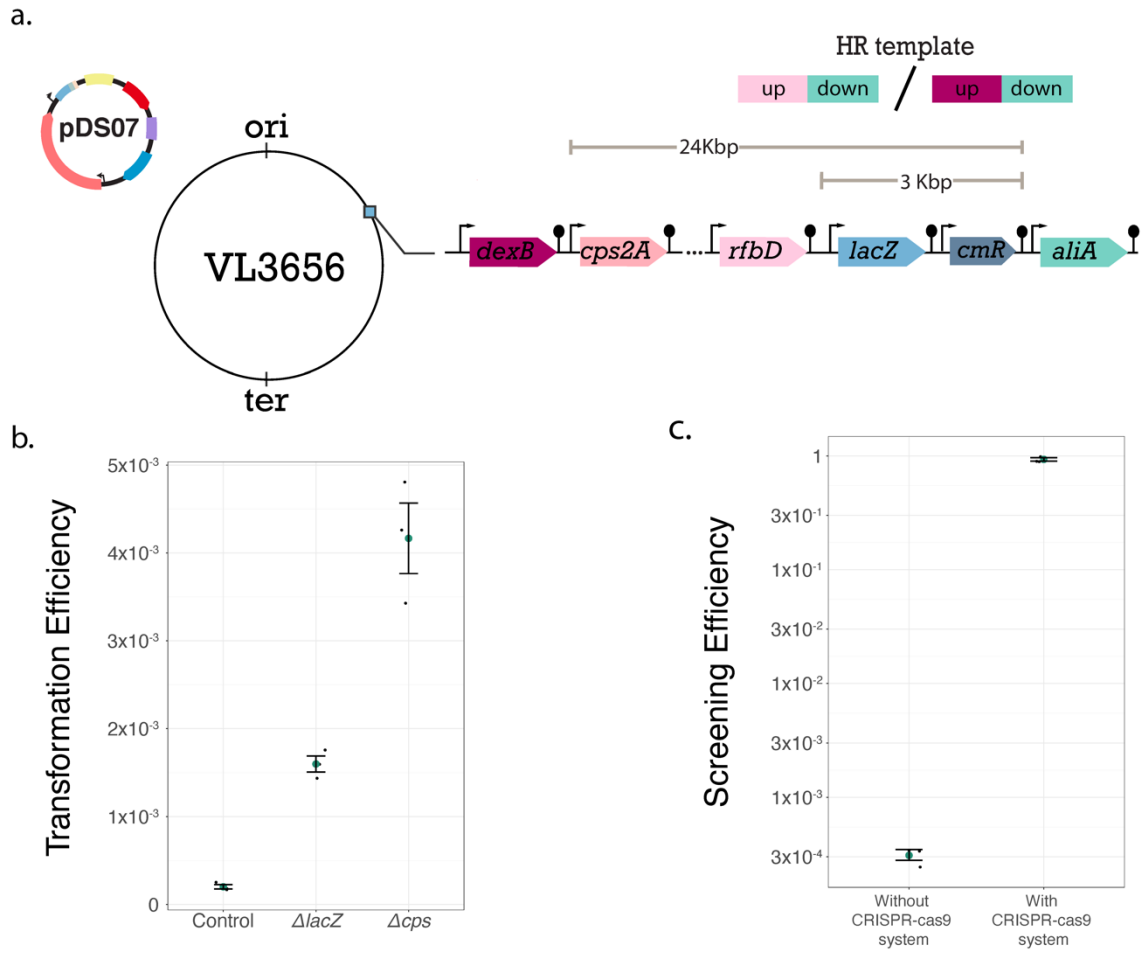


Figure 4: Genome editing in *S. pneumoniae* using CRISPR-Cas9. (a) Schematic representation of strain VL3656. The *lacZ* gene has been inserted downstream the capsule operon and a version of the plasmid with a sgRNA targeting *lacZ* has been transformed in the strain. **(b)** Transformation efficiency of *lacZ* and capsule operon deletion. The transformation efficiency was calculated by dividing the total number of cells as counted on plates without Cas9 inducer (1 mM $ZnCl_2$/$MnSO_4$) by the number of colonies, which were also green, in the presence of inducer. 'Control' is the transformation assay of strain VL3656 in the absence of HR template DNA. **(c)** Efficiency of successful transformants screened for integration of the *lacZ* deletion when using no selection and when using the CRISPR-Cas9 system. Data represent the average of three independent experiments (± SE).

Consecutive deletions of virulence factors of *S. pneumoniae*

Once the capsule operon and *lacZ* were removed from the chromosome, it was confirmed by colony PCR. All tested colonies demonstrated the expected PCR product. One such colony was picked resulting in strain VL3659. Next, we grew the new strain at the non-permissive temperature (40°C), eliminating the plasmid, resulting in strain VL3660 (*Δcps*).

To examine whether the system could be used in multiple rounds of genome editing, we attempted to delete the virulence factor pneumolysin. To delete the *ply* gene, we designed a sgRNA targeting pneumolysin and constructed plasmid pDS12, which we transformed into

VL3660. Following the same procedure as used to delete the *cps* operon, we deleted *ply*. Again, to confirm the successful deletion, the same principle for the primer set was used. All the colonies from the transformation plate had the expected PCR product demonstrating high selection efficiency using the CRISPR-Cas9 system. Finally, following the same strategy, we also deleted another important virulence factor, *lytA*, resulting in strain VL3665 (Figure 5; *Δcps, Δply, ΔlytA* + pDS13).

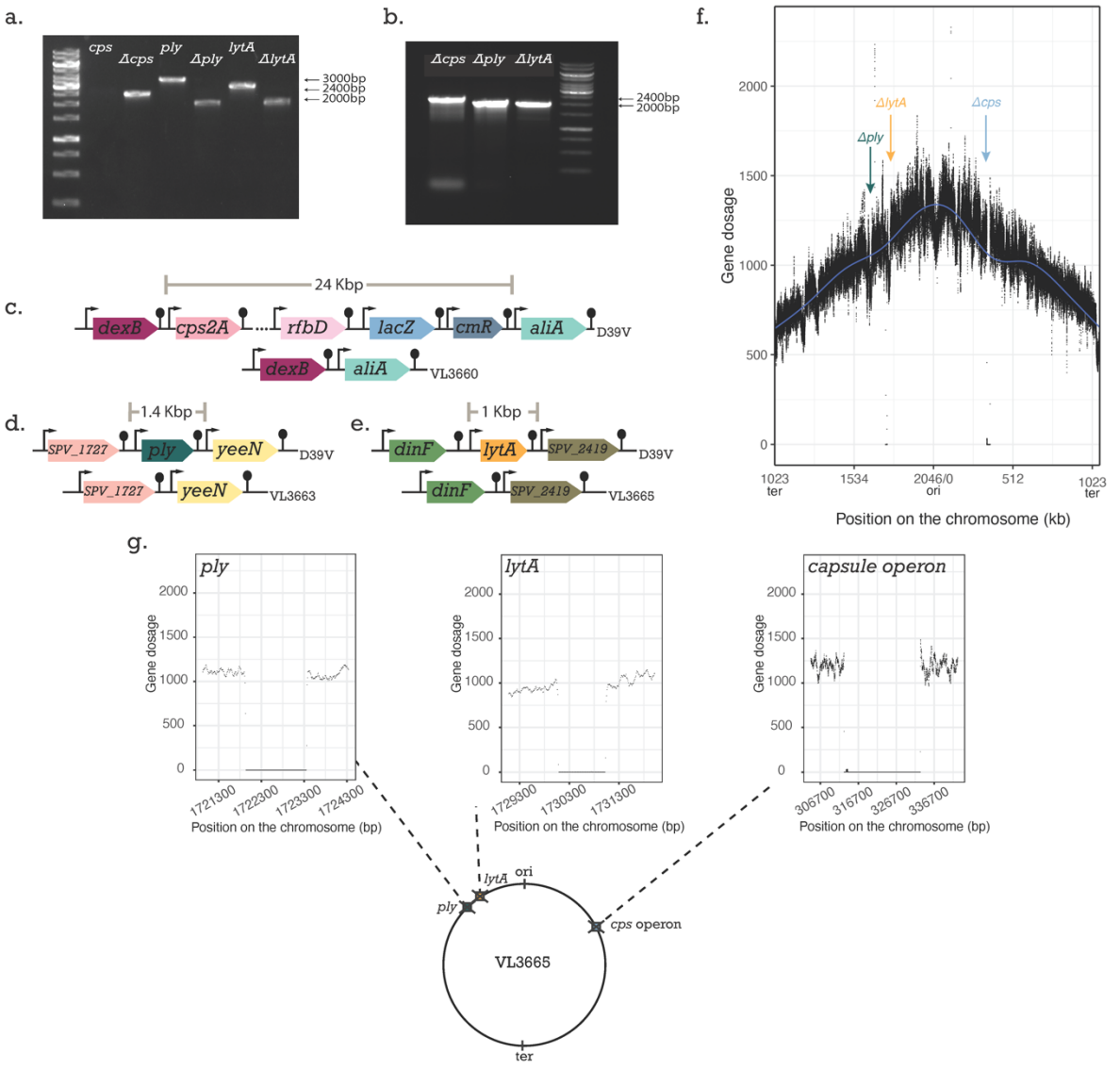


Figure 5: Genome analysis of the Δcps, Δply, ΔlytA triple mutant generated using CRISPR-Cas9 editing. (a) Colony PCR analysis of expected sizes for deletion of three virulence genes. WT vs VL3665. **(b)** Colony PCR analysis of three virulence gene deletions in the final strain VL3665. **(c)** Schematic representation of the D39V capsule operon region before deletion and after deletion, resulting in strain VL360. **(d)** Schematic representation of the ply region. **(e)** Schematic representation of the lytA region. **(f)** Whole genome marker frequency analysis of strain VL3665. The number of mapped reads (gene dosage) is plotted as a function of the position on the circular chromosome. **(g)**

Schematic representation of strain VL3665 with three virulence genes deleted. A zoom in of 10kb upstream and downstream of the deleted regions is shown.

Cas9-dependent genome editing was specific without evidence for off-target cutting in constructing a triple deletion *S. pneumoniae* mutant

After three consecutive deletions, using our plasmid with the CRISPR-Cas9 system, the final result was strain VL3665. It has been previously shown that Cas9 tolerates mismatches between guide RNA and target DNA at different positions in a sequence-dependent manner, resulting in off-target DSBs (Hsu *et al.*, 2013). To examine the fidelity of our CRISPR-Cas9 system and whether there were detectable genome-wide off-target effects, we performed whole-genome sequencing (WGS). The analysis detected one single SNP in the genome, in the gene *psaA* (*SPV_1463*), encoding a manganese ABC transporter. The mutation results in a D137E amino acid change. Using Sanger sequencing we confirmed that this SNP occurred only in the last strain of the consecutive deletions and it has not been present in the intermediate steps. This could have occurred as a result of needing to grow in the presence of $ZnSO_4$/$MnSO_4$ in the medium (Eijkelkamp *et al.*, 2014). However, comparing the growth of this strain with the ones that do not have the mutation, we do not observe any difference in fitness (Figure S3). Furthermore, there is no evidence to believe that this mutation is associated with an off-target effect as the sequence surrounding the SNP is completely different from the used sgRNA present in plasmid pDS13. Therefore, it is likely that this SNP occurred randomly during growth, without affecting fitness. We do note that the WGS approach only assesses off targeting for the carefully selected three sgRNAs and it remains to be seen how Cas9 specificity is with sub-optimal sgRNAs in *S. pneumoniae*.

Colony PCR on strain VL3665, confirmed that the three major virulence factors were deleted from the chromosome (Figure 5a-e). Additionally, reads from whole-genome sequencing were competitively mapped onto the reference genome, our wild type lab strain D39V (Figure 5f). Direct comparisons between the genomes revealed the three chromosomal positions at which the deletions have taken place, since at these positions, the read coverage drops. Therefore, we confirmed that we had successfully performed markerless deletions of these three genes (Figure 5g).

Discussion

Genetic manipulation of microorganisms has been pivotal for the development of biotechnological tools and the study of microorganisms themselves. CRISPR–Cas technology opens new avenues in genetic engineering applications by increasing expediency and efficiency in the generation of desired mutations, potentially without the

necessity of plasmid integration, extensive screening, or counter-selection (Selle and Barrangou, 2015). So far, CRISPR-Cas9 systems have been applied in a few bacterial species to edit their genomes, with more systems rapidly becoming available. Initially, the natural dual-RNA-guided CRISPR-Cas9 was used for genome editing (Jiang *et al.*, 2013; Oh and Van Pijkeren, 2014), but recently the engineered sgRNA- CRISPR-Cas9 has been employed. In different approaches, the two components, Cas9 and sgRNA can be either in two different plasmids (Jiang *et al.*, 2015), or in a single plasmid (Altenbuchner, 2016) and efficient genome editing has been achieved, like deletions, insertions and introduction of point mutations in *Escherichia coli* (Jiang *et al.*, 2015) and in *Bacillus subtilis* (Altenbuchner, 2016). In Streptomyces, a single plasmid system, with a temperature-sensitive pSG5 replicon was used for curing the plasmid after a successful series of deletions (Cobb, Wang and Zhao, 2015). Similarly, in *Staphylococcus aureus* a temperature-sensitive, two-vector system enables conditional recombineering and CRISPR-Cas9-mediated counterselection (Penewit *et al.*, 2018). To date, a number of studies have established CRISPR-Cas9 tools for genetic engineering in different bacteria, and this catalogue likely will continue to expand, enabling an increasing number of applications, with advantages and limitations. In *S. pneumoniae*, a dual-RNA:Cas9 system, integrated in the chromosome, has been shown to perform insertions, deletions or scarless single-nucleotide substitutions(Jiang *et al.*, 2013).

In this study, we have developed a replicative plasmid with a broad host-range temperature-sensitive origin of replication carrying a concise CRISPR-Cas9-based system for advanced and markerless genome engineering in the bacterium *S. pneumoniae.* In particular, we demonstrate that we have successfully deleted genes and large chromosomal regions in a precise and sequential way.

The here designed plasmid has the temperature sensitive origin of replication *pG$^+$host*, which is a derivative of pWV01 of *L. lactis* and can be successfully propagated in pneumococcus at 30°C, while it is not stable at 37°C or 40°C. Indeed, we show that our *pG$^+$host* derivative, pDS05, is rapidly lost at 37°C and at 40°C (Figure 1, Figure S1). We used this feature to eliminate the plasmid from the strains, upon the desired deletion. The fact that the copies of the plasmid vary per cell does not affect our system, since even one copy of *cas9* seems to be sufficient to perform the DSB (van Raaphorst, Kjos and Veening, 2017).

Specifically, our approach is to harness this CRISPR-Cas9 system and the homologous recombination system, to perform CRISPR-Cas9-Mediated Counterselection. Successful transformants survive the CRISPR-Cas9 induced DSB only if they uptake the rescue HR template. At this point we do not know whether a CRISPR-Cas9-induced DSB also promotes RecA-dependent HR with the user defined template. Regardless, our CRISPR-Cas9 system

manages to select for transformants in which single genes or even large chromosomal regions were deleted with very high efficiency. Comparing this to just performing natural transformation without counterselection, which would be an alternative for clean deletions, we show the advantages of our system (Figure 4c). Without counterselection, we would need to screen many colonies to find correct transformants, depending on the target. This will have to be done by colony PCR, since in most cases, the desired deletion will not give any phenotypic difference in the colonies of the successful transformant, which is a costly and time demanding process. On the other hand, with CRISPR-Cas9-mediated counterselection, nearly all the colonies that we obtained were the desired transformant, since very few false positives have been observed.

Since we are ultimately interested to remove multiple genes and chromosomal regions from the genome, we also needed to demonstrate that our system is capable of consecutive deletions. The key for this was to easily eliminate the plasmid from the newly constructed strain. By growing the strain still carrying the plasmid at the high, non-permissive temperature, we manage to easily cure it. Next, we can transform a new plasmid and proceed further with our deletions. Specifically, after we deleted the capsule, we next deleted virulence factors *ply* and *lytA,* proving that our CRISPR-Cas9 system has flexibility in genetic manipulation of the bacterial genome. Apart from the above demonstrated consecutive deletions, the plasmid can be used in future studies for other genome editing options, like SNPs, insertions or replacements. Additionally, by including more than one sgRNA in the backbone, higher level of multiplexing, like concurrent or sequential deletions can be performed.

Together, the here described plasmid and approach will be useful for the pneumococcal research community and may be applicable to other Gram-positive bacteria as well. It is likely that the zinc-inducible promoter used to drive Cas9 expression here is specific to pneumococci and would need to be exchanged by an alternative inducible promoter for the species of interest. Plasmid pDS05 is available from Addgene (157913).

Materials and Methods

Bacterial strains, transformations and growth conditions

All pneumococcal strains used in this study are derivatives of the serotype 2 *S. pneumoniae* strain D39V (Avery, Macleod and Mccarty, 1944; Slager, Aprianto and Veening, 2018). All plasmids were cloned in *NEB® Turbo Competent E. coli* (catalog number C2984; New England BioLabs). All the strains are shown in Table 1.

Table 3: Strain and plasmid list

S. pneumoniae Strains	Relevant Genotype	Reference
D39V	Serotype 2 strain, wild type	Avery et al 1944, Slager et al 2018
VL321	*SPV_2146*-P_{32}-*lacZ-chl-aliA*	This study
VL588	*ssbB-luc-kan, cps::chl*	Lab collection
VL2172	pPEPZ-read1-P_3-*Bsa*I-*gfp-Bsa*I-read2-N701-p7	Lab collection
VL3655	D39V +pDS05 [*pG⁺host ori*(Ts)-*ermR-cloDF13ori-specR-*P_{Zn}*_wtcas9-*P_3*_gfp-sgRNA*]	This study
VL3656	*SPV_2146-lacZ-chl-aliA* +pDS07 [*pG⁺host ori*(Ts)-*ermR-cloDF13ori-specR-pZn_wtcas9-*P_3*_sgRNA lacZ*]	This study
VL3657	*ΔlacZ* +pDS07 [*pG⁺host ori*(Ts)-*ermR-cloDF13ori-specR-pZn_wtcas9-*P_3*_sgRNA lacZ*]	This study
VL3658	*ΔlacZ*	This study
VL3659	*Δcps* + pDS07 [*pG⁺host ori*(Ts)-*ermR-cloDF13ori-specR-*P_{Zn}*_wtcas9-*P_3*_sgRNA lacZ*]	This study
VL3660	*Δcps*	This study
VL3661	*Δcps* + pDS12 [*pG⁺host ori*(Ts)-*ermR-cloDF13ori-specR-*P_{Zn}*_wtcas9-* P_3*-sgRNA ply*]	This study
VL3662	Δcps, Δply + pDS12 [*pG⁺host ori*(Ts)-*ermR-cloDF13ori-specR-*P_{Zn}*_wtcas9-* P_3*-sgRNA ply*]	This study
VL3663	*Δcps, Δply*	This study

VL3664	*Δcps, Δply* + pDS13 [*pG⁺host ori*(Ts)-*ermR-cloDF13ori-specR-* *P_{Zn}_wtcas9-P_3_sgRNA lytA*]	This study
VL3665	*Δcps, Δply, ΔlytA* + pDS13 [*pG⁺host ori*(Ts)-*ermR-cloDF13ori-specR-* *P_{Zn}_wtcas9-P_3_sgRNA lytA*]	This study
Plasmids		
pDS05	*pG⁺host ori*(Ts)-*ermR-cloDF13ori-specR-* *P_{Zn}_wtcas9-P_3_gfp-sgRNA*	This study
pDS06	*pG⁺host ori*(Ts)-*ermR-cloDF13ori-specR-* *P_{Zn}_wtcas9-P_3_sgRNA luc*	
pDS07	*pG⁺host ori*(Ts)-*ermR-cloDF13ori-specR-* *P_{Zn}_wtcas9-P_3_sgRNA lacZ*	This study
pDS12	*pG⁺host ori*(Ts)-*ermR-cloDF13ori-specR- P_{Zn}* *_wtcas9-P_3_sgRNA ply*	This study
pDS13	*pG⁺host ori*(Ts)-*ermR-cloDF13ori-specR-* *P_{Zn}_wtcas9-P_3_sgRNA lytA*	This study
pRAS2	pJWV01-*P_{32}-lacZ*	Lab collection

S. pneumoniae was grown either at 30°C, 37°C or 40°C (indicated) without shaking in liquid C+Y medium adapted from Adams and Roe (Martin *et al.*, 1995) and contained the following compounds: adenosine (68.2 µM), uridine (74.6 µM), L-asparagine (302 µM), L-cysteine (84.6 µM), L-glutamine (137 µM), L-tryptophan (26.8 µM), casein hydrolysate (4.56 g L-1), BSA (729 mg L-1), biotin (2.24 µM), nicotinic acid (4.44 µM), pyridoxine (3.10 µM), calcium pantothenate (4.59 µM), thiamin (1.73 µM), riboflavin (0.678 µM), choline (43.7 µM), CaCl2 (103 µM), K2HPO4 (44.5 mM), MgCl2 (2.24 mM), FeSO4 (1.64 µM), CuSO4 (1.82 µM), ZnSO4 (1.58 µM), MnCl2 (1.29 µM), glucose (10.1 mM), sodium pyruvate (2.48 mM), saccharose (861 µM), sodium acetate (22.2 mM) and yeast extract (2.28 g L-1).

E. coli strains were cultivated in LB at 37°C with shaking. When appropriate, 100 µg/ml spectinomycin (spec) was added.

Transformation

To transform the different plasmid variants into *S. pneumoniae*, cells were grown in C+Y medium (pH 6.8) at 37 °C to an OD_{595} of 0.1. Then, cells were treated for 12 min at 37°C with synthetic CSP-1 (100 ng mL^{-1}) and incubated for 20 min at 30°C with the plasmid. After incubation, cells were grown in C+Y medium at the permissive temperature of 30°C for 120

min. *S. pneumoniae* transformants were selected by plating inside Columbia agar supplemented with 3% of defibrinated sheep blood (Thermo Scientific) with antibiotic selection, using erythromycin at concentration 0.25 μg/ml. Plates were incubated at 30°C.

To transform the HR template, cells were grown in C+Y medium (pH 6.8) at 30°C with 0.1 μg/ml erythromycin to an OD_{595} of 0.1. Then, cells were treated for 12 min at 37°C with synthetic CSP^{-1} (100 ng mL^{-1}) and incubated for 20 min at 30°C with the HR template. After incubation, cells were grown in C+Y medium at 30°C for 20 min. Transformants were selected by plating inside molten Columbia agar supplemented with 3% of defibrinated sheep blood (Thermo Scientific), with CRISPR-Cas9 -mediated counter selection, using a mixture of 1 mM $ZnCl_2$/$MnSO_4$. $MnSO_4$ was added to counteract potential toxicity from $ZnCl_2$.

Plates were incubated at 30°C. Correct transformation was verified by PCR and sequencing. Working stocks of cells were prepared by growing cells in C+Y (pH 6.8), until an OD_{595} of 0.4. Cells were collected by centrifugation (1595 × *g* for 10 min) and resuspended in fresh C+Y medium with 15% glycerol and stored at −80°C.

Plasmid curing

After the plasmid has been transformed into pneumococcus and successful deletion has been performed with the HR template and CRISPR-Cas9 -mediated counter selection, the plasmid can be eliminated from the pneumococcal cells. To achieve that, we first grow the strain with the plasmid at the non-permissive temperature, 40°C in C+Y, with a starting inoculum 1/10.000. Next, we plate the liquid culture in Columbia blood agar in several dilutions, to obtain single colonies after overnight incubation at 40°C. Single colonies were screened and 99% of them had successfully cured the plasmid from the strain.

Recombinant DNA techniques

Oligonucleotides were ordered from Sigma and are listed in Table 2. Phanta Max Super-Fidelity DNA Polymerase (Vazyme) was used in PCR amplifications, restriction enzymes (ThermoFisher Scientific) were used for digestions and T4 DNA Ligase (Vazyme) was used for ligations.

Table 4: Oligonucleotides used in this study

Name	**Sequence (5′–3′); restriction site (underlined)**
OVL4739_pGh F	CTCTCACACCTGCCTGTCAATCGCAACATCAAACCAAAATAAAAAC
OVL4740_pGh R	CTCTCACACCTGCCTGTTTCAAAAGCGACTCATAGAATTATTTC

OVL4741_pCDF-1b F	CTCTCACACCTGCCGTATGAATCTAGAGCGGTTCAGTAGAAAAG
OVL4742_pCDF-1b R	CTCTCACACCTGCCGTATACTTGAACGAATTGTTAGACATTATTTG
OVL4743_wtcas9 F	AGATGGCACCTGCCAGAAGTACAAGCACTTTGGGACGTTCTCCCTTAG
OVL4744_wtcas9 R	AGATGGCACCTGCCAGAACGCTAAATACGCTTCACAGTTTCTTCTTC
OVL4745_gRNA F	CTCTCACACCTGCTCACGCGTATAAGAGACAGCCATTCTACAG
OVL4746_gRNA R	CTCTCACACCTGCTCACATTGAGACAGAAAAAAAGCACCGACTC
OVL2130_GG-luc-R	AAACGGCGCCATTCTATCCTCTA
OVL2131_GG-luc-F	TATATAGAGGATAGAATGGCGCC
OVL2132_GG-lacZ-F	TATAGGATGAAGACCAGCCCTTCC
OVL2133_GG-lacZ-R	AAACGGAAGGGCTGGTCTTCATCC
OVL2142_lin pGh R	CCTAGGTCTCATATAGTTATTATACCAGGG
OVL2143_lin pGh F	GTAAGGTCTCGGTTTAAGAGCTATG
OVL2250_GG-ply-F	TATACCGAGTTGTAACAGGCAAGG
OVL2251_GG-ply-R	AAACCCTTGCCTGTTACAACTCGG
OVL2813_GG-sgRNAlytA-F	TATAAACCAAAGAAGAGTTCATGA
OVL2814_GG-sgRNAlytA-R	AAACTCATGAACTCTTCTTTGGTT
rfbD-F	TCATGACCTACCTAGCTGAAAATCG
rfbD-R+BglII	GGCCAGATCTAAGCGCCCAATAACGAAGTATATTG
P32-lacZ-BglII	ATGCAGATCTAGGCCGGCCGATATGATAAG
lacZ_R-AscI	ATCACGGGCGCGCCTTATTTTTGACACCAGACCAACTG

Plasmid and Strain construction

Plasmid pDS05 (*pG⁺host ori(Ts)-ermR-cloDF13ori-specR-P_{Zn}_wtcas9-P_3_gfp-sgRNA*). Gram-positive, temperature sensitive origin of replication *pG⁺host* (Maguin *et al.*, 1992) and gene *ermR*, which confers resistance to erythromycin, were amplified from plasmid

pGh9::ISS1 (Maguin *et al.*, 1996) using the primers OVL4739_pGh F and OVL4740_pGh R (fragment 1). Gram-negative origin of replication *CloDF13* (CDF) and gene *specR*, which confers resistance to spectinomycin, were amplified from plasmid pCDF-1b (Nijkamp *et al.*, 1986) with primers OVL4741_pCDF-1b F and OVL4742_pCDF-1b R (fragment 2). The gene which encodes wtCas9 under the control of the Zinc-inducible promoter was amplified from plasmid pJWV102-spCas9wt (van Raaphorst, Kjos and Veening, 2017), using the primers OVL4743_wtcas9 F and OVL4744_wtcas9 R (fragment 3). The sgRNA sequence in which the 20 base-pairing region of the sgRNA is replaced by the *gfp* gene was amplified from strain VL2172 with primers OVL4745_gRNA F and OVL4746_gRNA R (fragment 4). The four fragments were digested all together with restriction enzyme *AarI* and ligated. The ligation product was transformed into *E. coli* NEB Turbo and transformants were selected on LB agar with spectinomycin. Correct assembly was confirmed by PCR and sequencing.

Plasmid pDS06 (*pG$^+$host ori(Ts)-ermR-cloDF13ori-specR-P_{Zn}_wtcas9-P_3_sgRNA luc*). pDS05 was amplified with primers OVL2143_lin pGh F and OVL2142_lin pGh R (fragment 1). Spacer sequence of *sgRNA lacZ* was constructed by annealing primers OVL2131_GG-luc-F and OVL2130_GG-luc-R. Amplified pDS05 was digested with restriction enzyme *BsaI* and ligated with the annealed oligos. The ligation product was transformed into *E. coli* NEB Turbo and transformants were selected on LB agar with spectinomycin. Correct assembly was confirmed by PCR and sequencing. Construction of plasmids pDS07, pDS12 and pDS13 are described in the Supplementary Information.

S. pneumoniae strain VL321 (*SPV_2146*-P_{32}-*lacZ-chl-aliA*). *rfbD* and *SPV_2146* were amplified from chromosomal DNA with primers rfbD-F and rfbD-R+BglII (fragment 1). P_{32}-*lacZ* was amplified from pRAS2 (lab collection) with primers P32-lacZ-BglII and lacZ_R-AscI (fragment 2). Chloramphenicol resistance marker and *aliA* were amplified from strain VL588 with primers cam-F+AscI and del_CSP.dn-R (fragment 3). Fragment 1 was digested with restriction enzyme *Bgl*II, fragment 2 was digested with restriction enzymes *Bgl*II and *Asc*I and fragment 3 was digested with restriction enzyme *Asc*I. All three fragments were ligated together, and the ligation product was transformed into D39V. Correct assembly was confirmed by PCR and sequencing. Construction of plasmid-containing strains VL3655, VL3656, VL3657, VL3659, VL3660, VL3661, VL3662, VL3664 and VL3665 and plasmid-cured strains VL3658, VL3660 and VL3663, are described in the Supplementary Information.

Microscopy

S. pneumoniae cells were stored as exponential phase frozen cultures. Frozen stock was inoculated 1:100 in C+Y medium and pre-grown to OD600 ~ 0.1. Cells were diluted once

again 1:100 in fresh C+Y (with antibiotic, if applicable) and grown to exponential phase to achieve balanced growth.

Cells were grown as described above to achieve balanced growth and subsequently concentrated and brought onto a multi test slide carrying a thin layer of 1.2% agarose in C+Y. Imaging was performed on Fluorescence microscopy was performed on a Leica DMi8 through a 100x phase contrast objective (NA 1.40) with a SOLA Light Engine (Lumencor) light source. Light was filtered through external excitation filters 470/40 nm (Chroma ET470/40x) for visualization of GFP. Light passed through a cube (Leica 11536022) with a GFP/RFP polychroic mirror (498/564 nm). External emission filters used were from Chroma ET520/40m. Images were captured using LasX software (Leica) and exported to ImageJ (Schneider, Rasband and Eliceiri, 2012) for final preparation.

Cell outlines were detected using MicrobeJ (Ducret, Quardokus and Brun, 2016). For all microscopy experiments, random image frames were used for analysis. The cell outline, object detection, and fluorescent intensity data were further processed using the R-package BactMAP (Raaphorst, Kjos and Veening, 2020).

Transformation efficiency assays

To calculate the transformation efficiency, 1 µg/ml of PCR product of DNA fragment containing the upstream and the downstream region of the deletion target was added. Serial dilutions were plated either with or without erythromycin and with or without 1 mM $ZnCl_2$/$MnSO_4$, depending on the counterselection method, and the transformation efficiency was calculated by dividing the number of transformants by the total viable count. All transformation efficiency values are averages of three biologically independent experiments.

Plating assays

To determine the number of colonies that are resistant to erythromycin and, consequently, harbor the plasmid, cells were grown in C+Y culture, as described above. At each time point, the same amount of a dilution was plated with or without erythromycin inside Columbia blood agar and the plates were incubated O/N at the same temperature as the liquid culture they were growing in.

Whole genome sequencing and variant analysis

Genomic DNA was isolated using the FastPure Bacteria DNA Isolation Mini Kit (Vazyme) according to the manufacturers' protocol and sent for whole genome sequencing. Illumina library prep and sequencing were carried out by Novogene (sequencing in PE150 mode). Reads were trimmed using Trimmomatic (Bolger, Lohse and Usadel, 2014), then assembled

using SPAdes (Nurk *et al.*, 2013) and contigs were reordered in Mauve (Darling *et al.*, 2004) using D39V as a reference (Slager, Aprianto and Veening, 2018). Reads were mapped onto the scaffold using bwa (Li and Durbin, 2009), read depth was determined in samtools (Li *et al.*, 2009), and plotted in R (Team, 2014). In order to detect small variants, raw reads were mapped onto the reference using bwa, and variants were detected using Freebayes (Garrison and Marth, 2012). Potential variants with PHRED scores greater than 30 were filtered out on DP >5 using vcflib (Garrison, no date), then intersected with the D39V annotation using Bedtools (Quinlan and Hall, 2010)

References

Adli, M. (2018) 'The CRISPR tool kit for genome editing and beyond', *Nature Communications*. Nature Publishing Group, pp. 1–13. doi: 10.1038/s41467-018-04252-2.

Altenbuchner, J. (2016) 'Editing of the *Bacillus subtilis* genome by the CRISPR-Cas9 system', *Applied and Environmental Microbiology*. American Society for Microbiology, 82(17), pp. 5421–5427. doi: 10.1128/AEM.01453-16.

Avery, O. T., Macleod, C. M. and Mccarty, M. (1944) 'Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus Type III', *The Journal of experimental medicine*, 79(2), pp. 137–58. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2135445&tool=pmcentrez&rendertype=abstract (Accessed: 16 October 2014).

Barrangou, R. *et al.* (2007) 'CRISPR provides acquired resistance against viruses in prokaryotes.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 315(5819), pp. 1709–12. doi: 10.1126/science.1138140.

Bijlsma, J. J. E. *et al.* (2007) 'Development of genomic array footprinting for identification of conditionally essential genes in *Streptococcus pneumoniae*.', *Applied and environmental microbiology*. American Society for Microbiology (ASM), 73(5), pp. 1514–24. doi: 10.1128/AEM.01900-06.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: a flexible trimmer for Illumina sequence data', *Bioinformatics (Oxford, England)*. 2014/04/01. Oxford University Press, 30(15), pp. 2114–2120. doi: 10.1093/bioinformatics/btu170.

Cañadas, I. C. *et al.* (2019) 'RiboCas: A Universal CRISPR-Based Editing Tool for *Clostridium*', *ACS Synthetic Biology*. ACS Synth Biol, 8(6), pp. 1379–1390. doi: 10.1021/acssynbio.9b00075.

Cobb, R. E., Wang, Y. and Zhao, H. (2015) 'High-Efficiency Multiplex Genome Editing of *Streptomyces* Species Using an Engineered CRISPR/Cas System', *ACS Synthetic Biology*. American Chemical Society, 4(6), pp. 723–728. doi: 10.1021/sb500351f.

Darling, A. C. E. *et al.* (2004) 'Mauve: multiple alignment of conserved genomic sequence with rearrangements.', *Genome research*. Cold Spring Harbor Laboratory Press, 14(7), pp. 1394–403. doi: 10.1101/gr.2289704.

de Maat, V. *et al.* (2019) 'CRISPR-Cas9-mediated genome editing in vancomycin-resistant *Enterococcus faecium'*, *FEMS Microbiology Letters*. Oxford Academic, 366(22). doi: 10.1093/femsle/fnz256.

Ducret, A., Quardokus, E. M. and Brun, Y. V. (2016) 'MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis', *Nature microbiology*. NIH Public Access, 1(7), p. 16077. doi: 10.1038/NMICROBIOL.2016.77.

Eberhardt, A. *et al.* (2009) 'Cellular localization of choline-utilization proteins in *Streptococcus pneumoniae* using novel fluorescent reporter systems', *Molecular Microbiology*. John Wiley & Sons, Ltd, 74(2), pp. 395–408. doi: 10.1111/j.1365-2958.2009.06872.x.

Eijkelkamp, B. A. *et al.* (2014) 'Extracellular zinc competitively inhibits manganese uptake and compromises oxidative stress management in *Streptococcus pneumoniae*', *PLoS ONE*. PLoS One, 9(2). doi: 10.1371/journal.pone.0089427.

Garrison, E. (no date) *GitHub - vcflib/vcflib: C++ library and cmdline tools for parsing and manipulating VCF files*.

Garrison, E. and Marth, G. (2012) 'Haplotype-based variant detection from short-read sequencing'.

Griffith, F. (1928) 'The Significance of Pneumococcal Types.', *The Journal of hygiene*. Cambridge University Press, 27(2), pp. 333–339. doi: 10.1055/s-0028-1106478.

Guo, T. *et al.* (2019) 'A rapid and versatile tool for genomic engineering in *Lactococcus lactis'*, *Microbial Cell Factories*. BioMed Central, 18(1), p. 22. doi: 10.1186/s12934-019-1075-3.

Halfmann, A., Hakenbeck, R. and Brückner, R. (2007) 'A new integrative reporter plasmid for *Streptococcus pneumoniae*', *FEMS Microbiology Letters*. Oxford Academic, 268(2), pp. 217–224. doi: 10.1111/j.1574-6968.2006.00584.x.

Hsu, P. D. *et al.* (2013) 'DNA targeting specificity of RNA-guided Cas9 nucleases', *Nature Biotechnology*. Nature Publishing Group, 31(9), pp. 827–832. doi: 10.1038/nbt.2647.

Jansen, R. *et al.* (2002) 'Identification of genes that are associated with DNA repeats in prokaryotes.', *Molecular microbiology*, 43(6), pp. 1565–75. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11952905 (Accessed: 13 May 2018).

Jiang, W. *et al.* (2013) 'RNA-guided editing of bacterial genomes using CRISPR-Cas systems.', *Nature biotechnology*. Nature Publishing Group, 31(3), pp. 233–9. doi: 10.1038/nbt.2508.

Jiang, Y. *et al.* (2015) 'Multigene editing in the Escherichia coli genome via the CRISPR-Cas9 system', *Applied and Environmental Microbiology*. American Society for Microbiology, 81(7), pp. 2506–2514. doi: 10.1128/AEM.04023-14.

Jinek, M. *et al.* (2012) 'A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.', *Science (New York, N.Y.)*, 337(6096), pp. 816–21. doi: 10.1126/science.1225829.

Johnston, C. *et al.* (2014) 'Bacterial transformation: distribution, shared mechanisms and divergent control', *Nature Reviews Microbiology*. Nature Publishing Group, 12(3), pp. 181–196. doi: 10.1038/nrmicro3199.

Kadioglu, A. *et al.* (2008) 'The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease.', *Nature reviews. Microbiology*, 6(4), pp. 288–301. doi: 10.1038/nrmicro1871.

Keller, L. E. *et al.* (2019) 'Three new integration vectors and fluorescent proteins for use in the opportunistic human pathogen *Streptococcus pneumoniae*', *Genes*. MDPI AG, 10(5). doi: 10.3390/genes10050394.

Kurushima, J. *et al.* (2020) 'Unbiased homeologous recombination during pneumococcal transformation allows for multiple chromosomal integration events', *eLife*. eLife Sciences Publications, Ltd, 9. doi: 10.7554/elife.58771.

Lee, M. S., Seok, C. and Morrison, D. A. (1998) 'Insertion-duplication mutagenesis in *Streptococcus pneumoniae*: Targeting fragment length is a critical parameter in use as a random insertion tool', *Applied and Environmental Microbiology*. American Society for Microbiology, 64(12), pp. 4796–4802. doi: 10.1128/aem.64.12.4796-4802.1998.

Li, H. *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*, 25(16), pp. 2078–2079. doi: 10.1093/bioinformatics/btp352.

Li, H. and Durbin, R. (2009) 'Fast and accurate short read alignment with Burrows-Wheeler transform', *Bioinformatics*, 25(14), pp. 1754–1760. doi: 10.1093/bioinformatics/btp324.

Li, Y., Thompson, C. M. and Lipsitch, M. (2014) 'A modified Janus cassette (sweet Janus) to improve allelic replacement efficiency by high-stringency negative selection in *Streptococcus pneumoniae'*, *PLoS ONE*, 9(6), pp. 2–7. doi: 10.1371/journal.pone.0100510.

Liu, X. *et al.* (2017) ' High-throughput CRISPRi phenotyping identifies new essential genes in *Streptococcus pneumoniae* ', *Molecular Systems Biology*. EMBO, 13(5), p. 931. doi: 10.15252/msb.20167449.

Liu, X. *et al.* (2021) 'Exploration of Bacterial Bottlenecks and *Streptococcus pneumoniae* Pathogenesis by CRISPRi-Seq', *Cell Host and Microbe*. Elsevier Inc., 29(1), pp. 107-120.e6. doi: 10.1016/j.chom.2020.10.001.

Maguin, E. *et al.* (1992) 'New thermosensitive plasmid for gram-positive bacteria.', *Journal*

of bacteriology. American Society for Microbiology (ASM), 174(17), pp. 5633–8. doi: 10.1128/jb.174.17.5633-5638.1992.

Maguin, E. *et al.* (1996) 'Efficient insertional mutagenesis in lactococci and other gram-positive bacteria.', *Journal of Bacteriology*, 178(3), pp. 931 LP – 935. doi: 10.1128/jb.178.3.931-935.1996.

Martin, B. *et al.* (1995) 'The recA gene of *Streptococcus pneumoniae* is part of a competence-induced operon and controls lysogenic induction.', *Molecular microbiology*, 15(2), pp. 367–79. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7538190 (Accessed: 27 October 2014).

Massidda, O., Nováková, L. and Vollmer, W. (2013) 'From models to pathogens: how much have we learned about *Streptococcus pneumoniae* cell division?', *Environmental Microbiology*. John Wiley & Sons, Ltd, 15(12), pp. 3133–3157. doi: 10.1111/1462-2920.12189.

Nijkamp, H. J. J. *et al.* (1986) 'The complete nucleotide sequence of the bacteriocinogenic plasmid CloDF13', *Plasmid*, 16(2), pp. 135–160. doi: 10.1016/0147-619X(86)90072-7.

Nurk, S. *et al.* (2013) 'Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads', in. Springer, Berlin, Heidelberg, pp. 158–170. doi: 10.1007/978-3-642-37195-0_13.

O'Brien, K. L. *et al.* (2009) 'Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates.', *Lancet*. Elsevier Ltd, 374(9693), pp. 893–902. doi: 10.1016/S0140-6736(09)61204-6.

Oh, J. H. and Van Pijkeren, J. P. (2014) 'CRISPR-Cas9-assisted recombineering in *Lactobacillus reuteri*', *Nucleic Acids Research*. Oxford University Press, 42(17). doi: 10.1093/nar/gku623.

Otto, R., de Vos, W. M. and Gavrieli, J. (1982) 'Plasmid DNA in *Streptococcus cremoris* Wg2: Influence of pH on Selection in Chemostats of a Variant Lacking a Protease Plasmid.', *Applied and environmental microbiology*. American Society for Microbiology (ASM), 43(6), pp. 1272–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16346027 (Accessed: 21 October 2019).

Penewit, K. *et al.* (2018) 'Efficient and scalable precision genome editing in *Staphylococcus aureus* through conditional recombineering and CRISPR/Cas9-mediated counterselection', *mBio*. American Society for Microbiology, 9(1). doi: 10.1128/mBio.00067-18.

Prudhomme, M., Libante, V. and Claverys, J. P. (2002) 'Homologous recombination at the border: Insertion-deletions and the trapping of foreign DNA in *Streptococcus pneumoniae'*,

Proceedings of the National Academy of Sciences of the United States of America. National Academy of Sciences, 99(4), pp. 2100–2105. doi: 10.1073/pnas.032262999.

Quinlan, A. R. and Hall, I. M. (2010) 'BEDTools: a flexible suite of utilities for comparing genomic features', *Bioinformatics*. Narnia, 26(6), pp. 841–842. doi: 10.1093/bioinformatics/btq033.

Raaphorst, R., Kjos, M. and Veening, J. (2020) 'BactMAP: An R package for integrating, analyzing and visualizing bacterial microscopy data', *Molecular Microbiology*, 113(1), pp. 297–308. doi: 10.1111/mmi.14417.

van Raaphorst, R., Kjos, M. and Veening, J.-W. (2017) 'Chromosome segregation drives division site selection in *Streptococcus pneumoniae*', *Proceedings of the National Academy of Sciences*, 114(29), pp. E5959–E5968. doi: 10.1073/pnas.1620608114.

Schneider, C. A., Rasband, W. S. and Eliceiri, K. W. (2012) 'NIH Image to ImageJ: 25 years of image analysis.', *Nature methods*, 9(7), pp. 671–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22930834 (Accessed: 17 July 2014).

Selle, K. and Barrangou, R. (2015) 'Harnessing CRISPR-Cas systems for bacterial genome editing.', *Trends in microbiology*. Elsevier, 23(4), pp. 225–32. doi: 10.1016/j.tim.2015.01.008.

Shuman, S. and Glickman, M. S. (2007) 'Bacterial DNA repair by non-homologous end joining', *Nature Reviews Microbiology*, 5(11), pp. 852–861. doi: 10.1038/nrmicro1768.

Simell, B. *et al.* (2012) 'The fundamental link between pneumococcal carriage and disease.', *Expert review of vaccines*, 11(7), pp. 841–55. doi: 10.1586/erv.12.53.

Slager, J., Aprianto, R. and Veening, J. W. (2018) 'Deep genome annotation of the opportunistic human pathogen *Streptococcus pneumoniae* D39', *Nucleic Acids Research*. Oxford University Press, 46(19), pp. 9971–9989. doi: 10.1093/nar/gky725.

Sorg, R. A. *et al.* (2020) 'Synthetic gene-regulatory networks in the opportunistic human pathogen *Streptococcus pneumoniae*', *Proceedings of the National Academy of Sciences*. National Academy of Sciences, p. 201920015. doi: 10.1073/pnas.1920015117.

Sorg, R. A., Kuipers, O. P. and Veening, J.-W. (2015) 'Gene Expression Platform for Synthetic Biology in the Human Pathogen *Streptococcus pneumoniae*', *ACS Synthetic Biology*, 4(3), pp. 228–239. doi: 10.1021/sb500229s.

Sung, C. K. *et al.* (2001) 'An rpsL Cassette , Janus , for Gene Replacement through Negative Selection in Streptococcus pneumoniae An rpsL Cassette , Janus , for Gene Replacement through Negative Selection in *Streptococcus pneumoniae*', *Applied and*

environmental microbiology, 67(11), pp. 5190–5196. doi: 10.1128/AEM.67.11.5190.

Team, R. core (2014) 'R: A language and environment for statistical computing', *R Foundation for Statistical Computing, Vienna, Austria*.

Zhou, D. *et al.* (2019) 'CRISPR/Cas9-assisted seamless genome editing in *Lactobacillus plantarum* and its application in N-acetylglucosamine production', *Applied and Environmental Microbiology*. American Society for Microbiology, 85(21). doi: 10.1128/AEM.01367-19.

Supplementary Figures

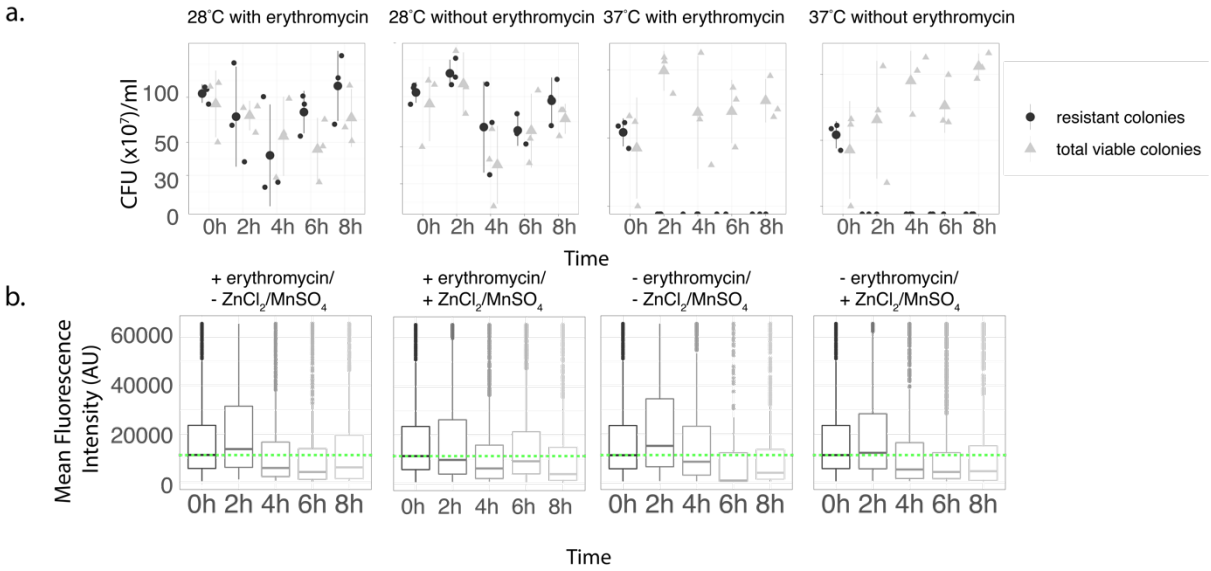


Figure S1: (a) Colony forming units (CFU) count of VL3655 cells harboring plasmid pDS05 grown under four different conditions; 28°C with erythromycin, 28°C without erythromycin, 37°C with erythromycin, and 37°C without erythromycin in time points 0, 2, 4, 6 and 8 hours after dilution from the 28°C with erythromycin condition. Each time point, for each condition, a dilution of the liquid culture was plated with or without erythromycin and incubated O/N at 37°C to determine the number of resistant colonies compared to the total viable colonies. **(b)** Quantification of mean fluorescence intensity of GFP of VL3655 cells grown under four different conditions; at 28°C: with erythromycin and without Zn^{2+}, with erythromycin and with Zn^{2+}, without erythromycin and with Zn^{2+}, and without erythromycin and with Zn^{2} in time points 0, 2, 4, 6 and 8 hours after dilution from the 28°C with erythromycin without Zn^{2+}condition. Fluorescence microscopy of ±1000 cells per condition per time point were quantified and analyzed using MicrobeJ and BactMap and plotted as box plots (box size and line represent the average intensity per cell) (see Materials and Methods). The green dotted horizontal line indicates the mean fluorescence of cells from the preculture harboring pDS05

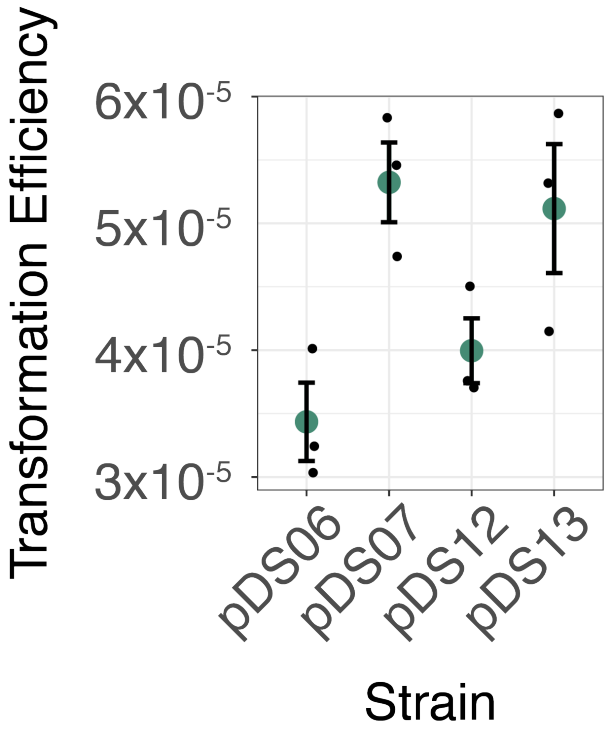


Figure S2: Transformation efficiency of four plasmids with a different sgRNA, transformed in D39V selected on erythromycin without Zn^{2+} so Cas9 is not induced. Plasmids pDS06 and pDS07 contain sgRNAs that do not target the chromosome, while pDS12 and pDS13 contain sgRNA that target the chromosome. The transformation efficiency was calculated by dividing the number of transformants by the total viable count. Black dots represent the mean of each biological replicate and the big green dot is representing the mean of the three replicates. Error bars represent the standard error of the mean

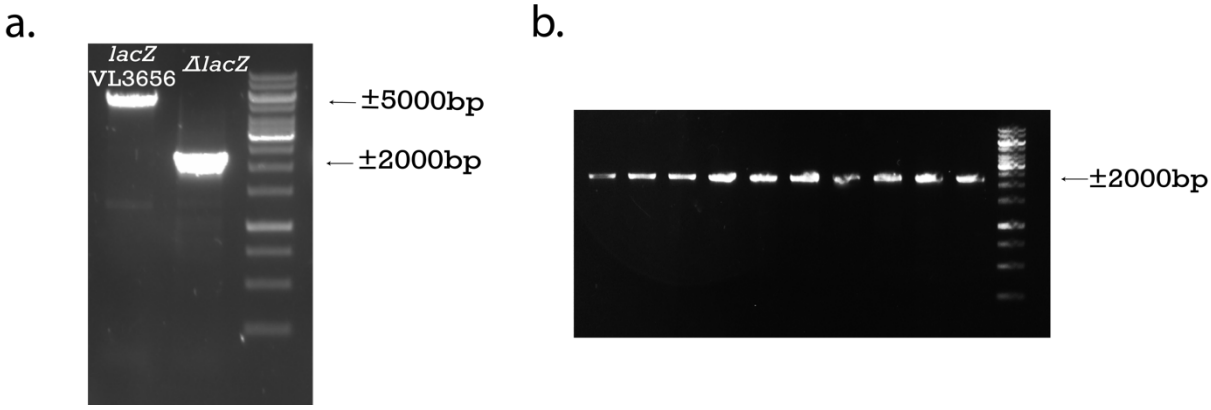


Figure S3: (a) Colony PCR analysis of expected sizes. **(b)** Eight randomly selected transformants of *lacZ* deletion.

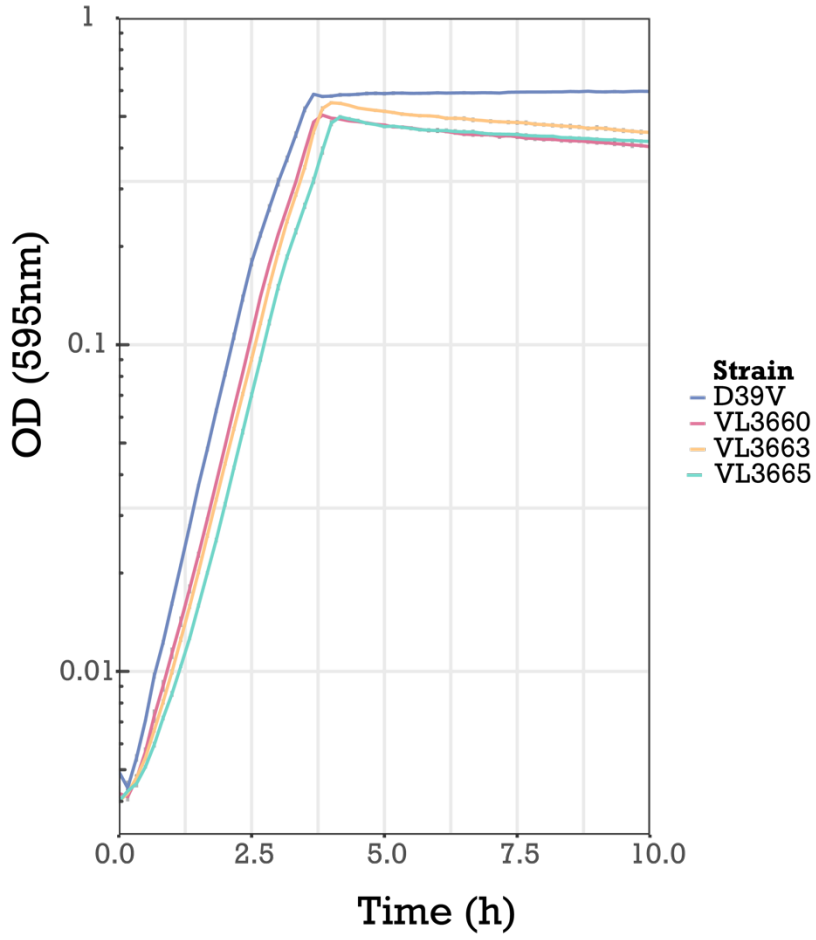


Figure S4: Cell density (OD595) of the bacterial cultures were measured every 10 min. A typical experimental outcome is shown (repeated at least 3 times). The values represent averages of three replicates. D39V = wild type, VL3660 = Δcps, VL3663 = Δcps, Δply, VL3665 = Δcps, Δply, ΔlytA

Supplementary methods

Plasmid and Strain construction

pDS07 (*pG⁺host ori(Ts)-ermR-cloDF13ori-specR-P_{Zn}_wtcas9-P_3_sgRNA lacZ*). pDS05 was amplified with primers OVL2143_lin pGh F and OVL2142_lin pGh R (fragment 1). Spacer sequence of *sgRNA lacZ* was constructed by annealing primers OVL2132_GG-lacZ-F and OVL2133_GG-lacZ-R. Amplified pDS05 was digested with restriction enzyme *BsaI* and ligated with the annealed oligos. The ligation product was transformed into *E. coli* NEB Turbo and transformants were selected on LB agar with spectinomycin. Correct assembly was confirmed by PCR and sequencing.

pDS12 (*pG⁺host ori(Ts)-ermR-cloDF13ori-specR- P_{Zn} _wtcas9-P_3_sgRNA ply*). pDS05 was amplified with primers OVL2143_lin pGh F and OVL2142_lin pGh R. Spacer sequence of *sgRNA ply* was constructed by annealing primers OVL2250_GG-ply-F and OVL2251_GG-

ply-R. Amplified pDS05 was digested with restriction enzyme *BsaI* and ligated with the annealed oligos. The ligation product was transformed into *E. coli* NEB Turbo and transformants were selected on LB agar with spectinomycin. Correct assembly was confirmed by PCR and sequencing.

pDS13 (*pG$^+$host ori(Ts)-ermR-cloDF13ori-specR- P_{Zn} _wtcas9-P_3_sgRNA lytA*). pDS05 was amplified with primers OVL2143_lin pGh F and OVL2142_lin pGh R. Spacer sequence of *sgRNA lytA* was constructed by annealing primers OVL2813_GG-sgRNAlytA-F and OVL2814_GG-sgRNAlytA-R. Amplified pDS05 was digested with restriction enzyme *BsaI* and ligated with the annealed oligos. The ligation product was transformed into *E. coli* NEB Turbo and transformants were selected on LB agar with spectinomycin. Correct assembly was confirmed by PCR and sequencing.

VL3655 (D39V + pDS05 [*pG$^+$host ori(Ts)-ermR-cloDF13ori-specR-P_{Zn}_wtcas9-P_3_gfp-sgRNA*]). Plasmid pDS05 was transformed into D39V and transformants were selected on Columbia blood agar with erythromycin to produce the strain VL3655. The presence of the plasmid was confirmed by PCR and plasmid extraction.

VL3656 (*SPV_2146-lacZ-chl-aliA* + pDS07 [*pG$^+$host ori(Ts)-ermR-cloDF13ori-specR-p_{Zn}_wtcas9-P_3_sgRNA lacZ*]). Plasmid pDS07, was transformed into VL321(*SPV_2146-lacZ-chl-aliA*) and transformants were selected on Columbia blood agar with erythromycin to produce the strain VL3656. The presence of the plasmid was confirmed by PCR and plasmid extraction.

VL3657 (Δ*lacZ* + pDS07 [*pG$^+$host ori(Ts)-ermR-cloDF13ori-specR-P_{Zn}_wtcas9-P_3_sgRNA lacZ*]). HR template Δ*lacZ* was transformed into VL3656 and transformants were selected on Columbia blood agar with $ZnCl_2$/$MnSO_4$ to produce the strain VL3657. Correct integration was confirmed by PCR.

VL3658 (*ΔlacZ*). Strain VL3657 was cured from the plasmid, as described, resulting in strain VL3658.

VL3659 (Δ*cps* + pDS07 [*pG$^+$host ori(Ts)-ermR-cloDF13ori-specR-P_{Zn}_wtcas9-P_3_sgRNA lacZ*]). HR template Δ*cps* was transformed into VL3656 and transformants were selected on Columbia blood agar with $ZnCl_2$/$MnSO_4$ to produce the strain VL3659. Correct integration was confirmed by PCR.

VL3660 (*Δcps*). Strain VL3659 was cured from the plasmid, as described, resulting in strain VL3660.

VL3661 (*Δcps* + pDS12 [*pG$^+$host ori(Ts)-ermR-cloDF13ori-specR-P_{Zn}_wtcas9- P_3-sgRNA ply*]). Plasmid pDS12 was transformed into VL3660 and transformants were selected on

Columbia blood agar with erythromycin to produce the strain VL3661. The presence of the plasmid was confirmed by PCR and plasmid extraction.

VL3662 (*Δcps, Δply* + pDS12 [*pG*$^{+}$*host ori(Ts)-ermR-cloDF13ori-specR-*P_{Zn}*_wtcas9-*P_3*_sgRNA ply*]). HR template Δ*ply* was transformed into VL3661 and transformants were selected on Columbia blood agar with $ZnCl_2$/$MnSO_4$ to produce the strain VL3662. Correct integration was confirmed by PCR.

VL3663 (*Δcps, Δply*). Strain VL3662 was cured from the plasmid, as described, resulting in strain VL3663.

VL3664 (*Δcps, Δply* + pDS13 [*pG*$^{+}$*host ori(Ts)-ermR-cloDF13ori-specR-*P_{Zn}*_wtcas9-*P_3*_sgRNA lytA*]). Plasmid pDS13 was transformed into VL3663 and transformants were selected on Columbia blood agar with erythromycin to produce the strain VL3664. The presence of the plasmid was confirmed by PCR and plasmid extraction.

VL3665 (*Δcps, Δply, ΔlytA* + pDS13 [*pG*$^{+}$*host ori(Ts)-ermR-cloDF13ori-specR-*P_{Zn}*_wtcas9-*P_3*_sgRNA lytA*]). HR template Δ*lytA* was transformed into VL3664 and transformants were selected on Columbia blood agar with $ZnCl_2$/$MnSO_4$ to produce the strain VL3665. Correct integration was confirmed by PCR.

Chapter 4: ComM promotes wall teichoic acid production in *Streptococcus pneumoniae* resulting in reduced susceptibility to autolysins

(D.S. aided in writing, research design and in performing experiments; unpublished)

Authors: Arnau Domenech[1,#], Dimitra Synefiaridou[1,#], Daniel Straume[2,#], Gonzalo Felipe Cebrero[3], Xue Liu[1], Camilo Perez[3], Nicolas Gisch[4], Leiv Sigve Håvarstein[2*] and Jan-Willem Veening[1*]

#contributed equally

Abstract

(DS wrote this)

Streptococcus pneumoniae is a bacterium naturally competent for genetic transformation. Competent bacteria produce so called 'fratricins' that kill non-competent pneumococci sharing the same ecological niche. Competent cells are protected against fratricide by producing the immunity protein ComM. The molecular mechanisms underlying ComM-dependent immunity, however, remains unknown. To unravel this, we performed a genome-wide CRISPRi-seq screen on a pooled sgRNA library growing under competence-permissive and non-permissive conditions. By evaluating our results, we found that teichoic acids proteins are essential during competence. We demonstrate that ComM, acting jointly with LytR, provides immunity against fratricins by switching the flux of teichoic acids from the membrane to the cell wall. Finally, we suggest a model in which activation of competence during infection is crucial to transfer important virulence factors out of the pneumococcal cell, ensuring better adherence or immune evasion.

Introduction

Streptococcus pneumoniae (the pneumococcus) is a member of the commensal microbiota of the human nasopharynx. However, it is a major public health problem because it can cause severe life-threatening infections such as sepsis, pneumonia and meningitis (Wahl et al. 2018).

The pneumococcus is a naturally transformable bacterium, which is able to take up or assimilate exogenous DNA (Chewapreecha et al. 2014; Johnston et al. 2014; Sia and Dawson 1931). This phenomenon is an important mechanism of genome plasticity and is largely responsible for the acquisition and spread of antibiotic resistance and virulence factors such as the capsule (Croucher et al. 2011; Wyres et al. 2013). The transformation process in the pneumococcus requires the induction of a physiological state named competence, which involves about 10% of the pneumococcal genome (Claverys, Martin, and Polard 2009; Slager, Aprianto, and Veening 2019).

Competence is induced by a classical two-component quorum sensing system in which the *comC*-encoded competence-stimulating peptide (CSP), is cleaved and exported by the membrane transporter ComAB to the extracellular space (Hui and Morrison 1991; Håvarstein, Coomaraswamy, and Morrison 1995; Havarstein, Diep, and Nes 1995). Upon a certain threshold of CSP accumulation, it stimulates the autophosphorylation of the membrane-bound histidine-kinase ComD, which subsequently activates its cognate response regulator ComE (Pestova, Håvarstein, and Morrison 1996; Martin et al. 2013). Phosphorylated ComE induces a positive feedback to fully activate competence. One of the genes regulated by ComE, *comX*, encodes a sigma factor, which activates the genes required for DNA repair, DNA uptake, and transformation (Martin et al. 2013). Although competence is mostly associated with DNA transformation, only 22 of the genes induced during competence are related to transformation (Peterson et al. 2004; Guiral et al. 2006). Other expressed genes are involved in biofilm formation (Aggarwal et al. 2018), bacteriocin production (Kjos et al. 2016) or siblings' fratricide (Steinmoen et al. 2002).

As pneumococci do not discriminate between homologous and foreign DNA, it is believed that fratricide serves to increase and facilitate the exchange of DNA between pneumococci in nature. Thus, competent pneumococci lyse and subsequently release nutrients and DNA from a subfraction of the population that does not become competent (Steinmoen et al. 2002). Three choline-binding proteins (CBPs) constitute the lysis mechanism: CbpD, LytA, and LytC. CbpD is the main driver of fratricide as the process cannot commence in its absence (Kausmally et al. 2005; Eldholm et al. 2009a). CbpD, LytA and LytC contain choline binding domains, which can bind to the choline residues in the cell wall teichoic acids (WTA) or the

lipid-anchored lipoteichoic acids (LTA) in the membrane (Kausmally et al. 2005). As bacteria without choline in their cell wall will not undergo lysis by these fratricins, this process enhances the probability for accessing homologous DNA (Johnsborg et al. 2008). The pneumococcus is special in its teichoic acid biosynthesis pathway, as the same substrate is used for LTA and WTA (Figure 1, Denapaite et al. 2012). It was recently shown that TacL is responsible for producing LTA from the precursor teichoic acid (Heß et al. 2017; Flores-Kim et al. 2019). The three putative enzymes of the LCP-protein family Cps2A, Psr, and LytR are predicted to anchor the precursor to the peptidoglycan (PGN) to create WTA, but it remains unclear which is the dominant route and how the levels of WTA and LTA are controlled (Figure 1) (Kawai et al. 2011; Eberhardt et al. 2012; Denapaite et al. 2012). In addition, most pneumococcal strains anchor their exopolysaccharide capsule on the same residue within the PGN as WTA, thus there must also be cross-talk between these two processes (Figure 1;Larson and Yother 2017; Fischer et al. 1993).

CbpD, which is only expressed during competence, consists of four domains: the N-terminal cysteine/histidine-dependent amidohydrolase-peptidase (CHAP) domain, which most likely cleaves peptide bonds of the PGN chain (Eldholm et al. 2009b)). Then two src-homology 3b (SH3b) domains, which recognize and bind to the PGN, and the C-terminal end consisting of four repeating choline binding sequences, which direct CbpD to the septal region of the pneumococcal cell (Eldholm et al. 2010). The initial cell wall damage by CbpD most likely facilitates the accessibility of LytA and LytC, enhancing the fratricide (or allolysis) process. Interestingly, several key virulence factors such as CbpA (involved in adherence) and PspA (inhibits opsonisation) are also CBPs (Kadioglu et al. 2008).

To avoid committing suicide, competent cells produce ComM, an immunity protein (Håvarstein et al. 2006). During competence, ComM (ComE-dependent) is produced (~ 5 min) earlier than the CbpD autolysin (ComX-dependent). Although ComM is an integral transmembrane protein, its structure is unknown, and how it provides immunity remains to be elucidated.

In addition, competence is also activated when cells are adhering to human epithelial cells (Aprianto et al. 2016) and during invasive disease in a mouse model (Lin et al. 2020). It has recently been suggested that competence is a key pathogenic factor in a murine pneumococcal meningitis model (Schmidt et al. 2019). However, it is unclear how competence and fratricide play a role in pneumococcal virulence.

Here, we found that teichoic acids are essential during competence. We demonstrate that ComM provides immunity against fratricins by switching the flux of teichoic acids from the membrane to the cell wall. We show that this is achieved in concert with LytR and that it is

LytR that is the main enzyme responsible for WTA production in the pneumococcus. Depletion of any of the steps involved in WTA will lead to increased susceptibility to CbpD and fratricide. Activation of competence leads to an increase of WTA, reducing the efficiency of CbpD. In addition, this newly formed WTA can now act as a landing platform for the CBPs, including several virulence factors, which are otherwise hidden from the host as they are initially bound to the LTA. The data presented here suggest a model in which competence activation during infection is crucial to shuttle important virulence factors to the outside of the pneumococcal cell to ensure better adherence or immune evasion. The potential ramifications of these findings are discussed.

Results

Teichoic acid biosynthesis is essential during pneumococcal competence as identified by CRISPRi-seq

While it has been shown that the competence-induced ComM protein is required for immunity against fratricins (Håvarstein et al. 2006), the molecular mechanism underlying immunity remains unclear. To investigate this, we performed a genome-wide CRISPRi-seq screen (Figure 1C-D). We screened a pooled library of pneumococcal strains harboring an inducible dCas9 and a constitutively expressed sgRNA, targeting in total 1499 operons in the genome of strain D39V (Liu et al. 2021). The rationale behind this screen is that bacteria undergoing transcriptional downregulation of genes important for immunity against fratricins will be outcompeted or lysed by fratricins produced by strains in which the sgRNA targets a neutral gene (Figure 1B). The pooled library was grown under competence-permissive and non-permissive conditions by the addition of synthetic CSP_1. To confirm that observed fitness costs were due to competence induction and not because of the essentiality of the targeted operon, four different conditions were tested (Figure S1A): (I) control (C+Y, pH 6.8; non-permissive for natural competence development), (II) library induction in absence of competence (+ IPTG), (III) competence induction (+ CSP_1) and (IV) both library and competence induction (+ IPTG and CSP_1). After Illumina sequencing (Liu et al. 2021), the fitness of targets and quality of the replicates were then evaluated by DEseq2 analysis (Love, Huber, and Anders 2014), and fold change of the abundance of sgRNAs between the four groups was analyzed (Figure S1B-C).

Fourteen sgRNAs targeting ten operons were significantly less abundant in condition IV, suggesting that they are essential or become more essential during competence (Figures 1D and S2A, and Table S1). As expected, *comCDE* and *comM* were among the top hits, demonstrating that when cells are unable to activate competence or ComM, they are rapidly outcompeted or lysed by competent (and CbpD) producing competitors. Strikingly, most of the other hits were operons related to teichoic acid (TA) synthesis (Figure 1D).

Next, we examined in more detail all the sgRNAs targeting genes involved in both competence and TA synthesis pathways (Figure S2). All but two operons related to TA synthesis (*SPV_1620-aatB*, and *tacL*) were underrepresented (Figure S2A). The first operon was less present when competence and the library were induced, although the differences were not significant; however, the sgRNA targeting *tacL*, which incorporates TAs chains into the membrane (LTA), did not show any fitness cost during competence (Figure S2B). The three remaining sgRNAs targeted *secG* and two hypothetical proteins: *SPV_0713* and *SPV_0564*.

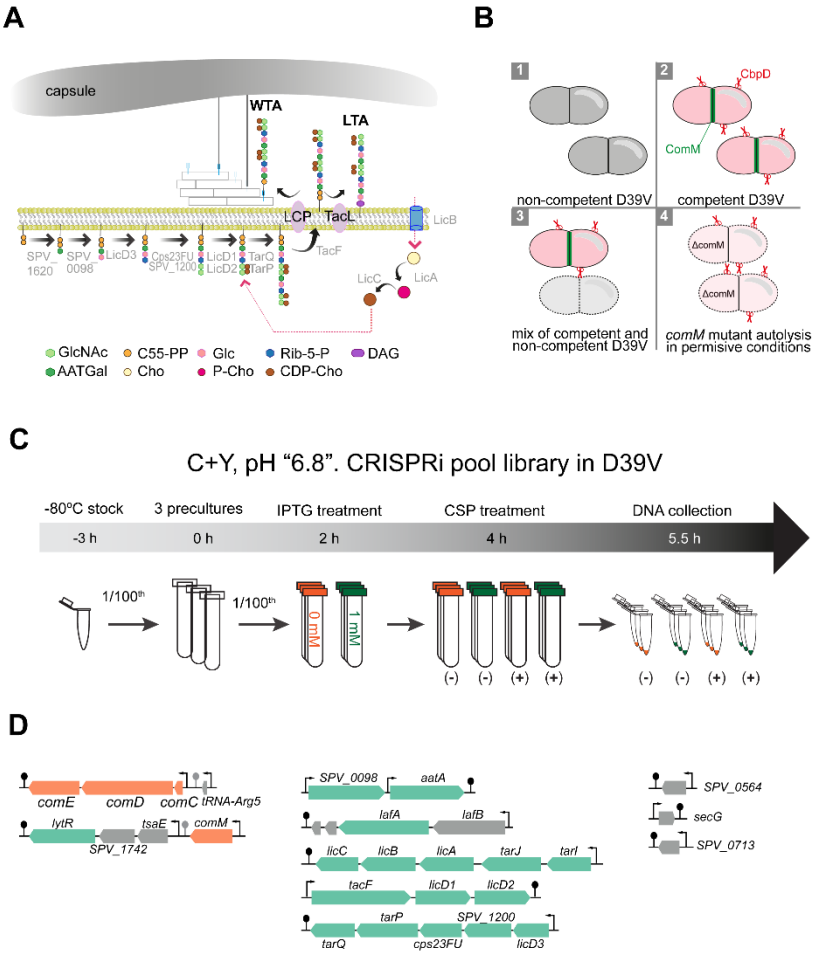


Figure 1: (a) Proposed TA synthesis pathway in Streptococcus pneumoniae. Abbreviations: LCP (LytR-Cps2A-Psr family), C55-PP (Undecaprenyl diphosphate), Glc (Glucose), Rib-5-P (ribitol 5-phosphate), DAG (Diacylglycerol), GlcNAc (N-acetylgalactosamine), UDP-KDG (UDP-4-keto-6-deoxy-D-glucose), AATGal (2-acetamido-4-amino-2,4,6-trideoxygalactose), Cho (Choline), P-Cho (phosphorylcholine). Note that the capsule and the WTA compete for the same anchoring position on the peptidoglycan (O-6 of MurNAc).**(b)** Pneumococcal fratricide. 1) in non-permissive conditions for competence, D39V cells do not become competent, thus, the fratricin CbpD and the immunity protein ComM are not expressed. 2) In high permissive conditions for natural competence, all cells become competent, producing CbpD but also ComM, thus all cells are protected and there is no fratricide. 3) In conditions where only a subpopulation of cells becomes competent, non-competent pneumococci lyse and subsequently release nutrients and DNA. 4) Deletion of comM results in autolysis when competence is activated via CbpD production. **(c)** Protocol used to detect essential genes during competence. Three independent precultures of the CRISPRi pooled library (Liu et al. 2021) were grown until OD 0.1 in C+Y pH 6.8 to avoid spontaneous competence. Cells were then diluted 100X in C+Y supplemented with 0 or 1 mM of IPTG. When cells reached OD595 nm 0.1 again, 100 ng/ml CSP1 was added to the indicated conditions and cells were grown for two more generations (OD595 nm ~ 0.4). Cells were collected, the DNA was isolated and the library was prepared for Ilumina sequencing (Liu et al. 2021). **(d)** Operons with fitness

cost during competence. In orange, competence-related genes; in green, genes involved in the teichoic acid synthesis pathway.

Teichoic acids are not required for competence development

(DS aided in these experiments)

One hypothesis that could explain why strains carrying an sgRNA targeting a TA gene are underrepresented in the CRISPRi-seq screen is that competence is not activated, and hence ComM is not expressed, and cells are susceptible to fratricide. To test whether TAs are required for competence activation, we generated non-polar depletion strains for each individual gene belonging to the 10 operons (Table S2). The deletions containing an erythromycin resistance cassette were transformed in the corresponding strain harboring an IPTG-inducible ectopic copy of the gene (P_{lac} promoter), in order to control the expression of the genes (e.g. *comM* $^{-/Plac-}$) (Figure S3). High-content microscopy of the depletion strains showed different phenotypes depending on the targeted gene, as shown before (Figure S4)(Liu et al. 2017). We used the competence-specific induced *ssbB* promoter fused to firefly luciferase (P_{ssbB}-*luc,* Slager et al. 2014) to quantify competence activation. Cells were grown in presence of different concentrations of IPTG to deplete or overproduce the TA gene products (0, 0.005 mM, 0.01 mM, 0.1 mM and 1 mM), and luciferase activity as well as optical density were regularly measured along the growth (Figure S5). As expected, in absence of *comCDE*, competence was not triggered. However, depletion of all TA-related genes except *tacL,* did not affect natural competence development under conditions where cell growth was not compromised (Figure S4). Interestingly, although the absence of TacL inhibited natural competence, somehow cells remained protected from fratricide (Figure S6), explaining why *tacL* depletion did not show any fitness cost in the CRISPRi experiment (Figure S2).

Teichoic acids protect from cell lysis during competence

(DS aided in these experiments)

As TAs are not required for competence induction (Figure S6), another hypothesis is that TAs provide resistance to fratricins during competence. In this case, depletion of TA-related genes should result in autolysis when competence is triggered. To test this, we used the SYTOX™ Green Dead Cell Stain (ThermoFisher scientific) to evaluate cell lysis in presence and absence of CSP_1-induced competence. A *comM* mutant was used as a control strain that should show lysis upon competence induction. As shown in Figure S6, a rapid lysis after CSP_1 addition was observed for the *comM* mutant and nearly all the TA-related genes. This data suggests that, besides ComM, TAs are essential to maintain the integrity of the cells during 'attack' of competence-induced fratricins (i.e., CbpD, LytA). Indeed, it was previously shown

that cells depleted for TAs become more susceptible towards autolysis by LytA (Bonnet et al. 2018).

lytR is in the same operon as *comM* and is upregulated during competence by transcriptional read-through

(DS aided in these experiments)

As both CbpD and LytA bind to the choline units on the TAs, we focused on the external part of the pathway: the anchoring of TAs into cell wall (WTA) and membrane (LTA), respectively (Figure 2A). TacL is the only known protein responsible for anchoring TAs to lipids, while LCP family proteins (i.e., LytR, Cps2A and Psr) are the putative proteins anchoring TA chains to the cell wall thereby producing WTA. The fact that *cps2A* and *psr* were not found in the CRISPRi-seq screen (Figure S2), and that LytR forms part of the same operon as ComM (Figure 2B), suggests a main role for LytR for attaching TAs into the cell wall and subsequent protection from fratricins.

While TacL is located in a single operon, the *comM-tsaE-spv_1742-lytR* operon shows an interesting complexity (Slager, Aprianto, and Veening 2018). Transcription of *comM* is under competence control by ComE, and its imperfect terminator is only about 61% efficient (Figure 2B). The three downstream genes are regulated by an internal transcription start site (TSS) with basal expression. However, when competence is activated, the expression of all three genes is highly increased due to the c*omM* terminator inefficiency (Figure 2C; Aprianto et al. 2018). Little is known about the role of TsaE and SPV_1742, however their individual deletion did not affect cell integrity during competence (Figures 2D and S5). To further exclude their role in fratricide, we constructed a double *ΔtsaE ΔSPV_1742* deletion mutant. As shown in Figure 2D, the *ΔtsaE ΔSPV_1742* double mutant was not susceptible to lysis during competence, suggesting they do not play a role in TA synthesis.

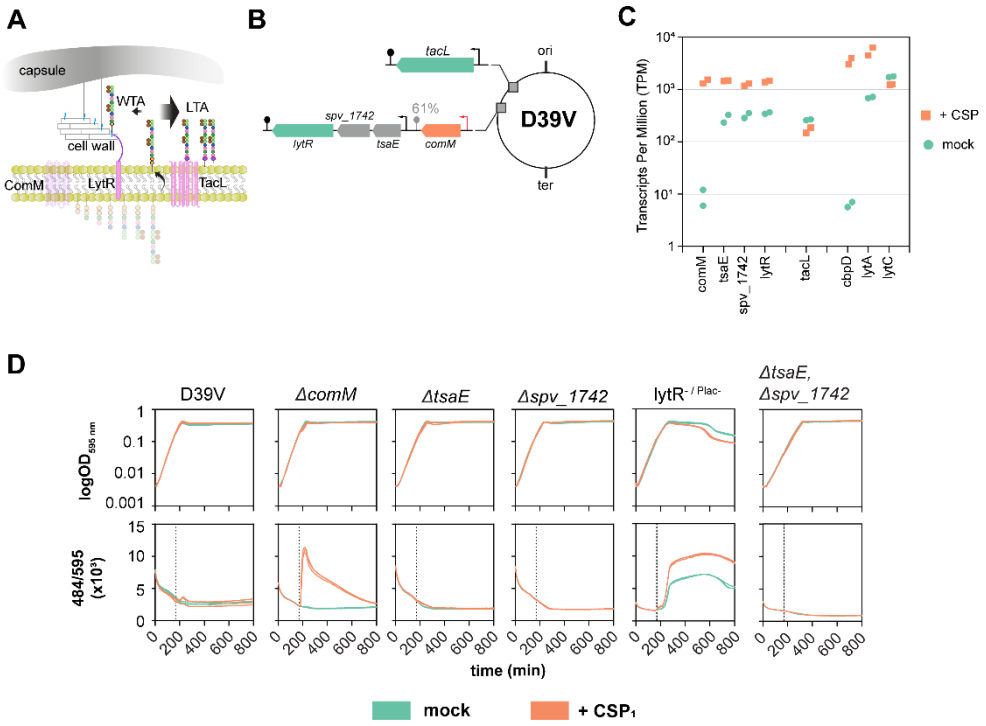


Figure 2. Role of the *comM-tsaE-SPV_1742-lytR* and *tacL* operons. **(a)** In *S. pneumoniae* the same TA chains can be incorporated either to the membrane (by TacL) or to the cell wall (mainly by LytR). During exponential phase TacL ensures a steady flux of LTAs (Flores-Kim et al. 2019) as depicted by the large rightwards pointing arrow. **(b)** Schematic representation of the *tacL*, *comM* and *lytR* operons. Note the ComE-activated promoter (orange arrow) upstream of *comM* and the imperfect transcriptional terminator downstream of *comM* (Slager, Aprianto, and Veening 2018). **(c)** Expression levels in competence induced and non-induced cells. RNA-seq data from Aprianto et al. 2018; Slager, Aprianto, and Veening 2019. Duplicates for each condition are shown **(d)** Cell lysis evaluation. Individual strains were grown in C+Y pH 6.8 to avoid natural competence development in presence of SYTOX™ Green Dead Cell Stain dye. When cell cultures reached $OD_{595\,nm}$ 0.1 (~ after 170 min), 100 ng/ml of CSP_1 was added to induce competence. Three biological replicates per condition are shown. As LytR is essential, the depletion strain (lytR $^{-\,/\,Plac-}$) was used in absence of IPTG, to deplete for LytR.

LytR is required for ComM-mediated immunity to CbpD

(DS aided in these experiments, except for the one of Fig 4C)

While *tsaE* and *spv_1742* did not play a role in immunity to fratricins, we were unable to construct a *lytR* deletion strain in the D39V genetic background, suggesting it is an essential gene under our experimental conditions. Nevertheless, we were able to construct a complementation strain in which *lytR* is expressed from an IPTG-inducible promoter at the ZIP locus (Keller, et al, 2019) while deleted from its native locus (Figure S3). To evaluate how LytR protects cells from competence-related lysis, we performed three complementary approaches: evaluation of cell lysis in the deletion of the fratricins (Figure 3A), overexpression of *cbpD*

(Figure 3B) and addition of exogenous CbpD (Figure 3C; pure CbpD was obtained as described by Straume et al. 2020). In all three approaches, absence of competence or *comM* depletion resulted in a rapid cell lysis when CbpD was present (either induced or added exogenously to the medium). Importantly, depletion of *lytR* strongly increased susceptibility to CbpD (Figure 3C). Deletion of both *cbpD* and *lytA* in lytR$^{-/Plac-}$ cells does not protect them from lysis, confirming the essential role of LytR, not only during competence (Figure 3A). Interestingly, LytR overexpression in the presence of basal levels of ComM showed some level of protection against CbpD-mediated lysis, which was more evident when both ComM and LytR were overexpressed (Figure 3C). Indeed, overproduction of both ComM and LytR provided better protection from CbpD-induced lysis than overproduction of ComM alone (Figure 3C). In contrast, native expression and overexpression of LytR in the absence of ComM (*comM* depletion), showed the same levels of lysis as a *comM* mutant, suggesting that both proteins act together to protect from fratricide.

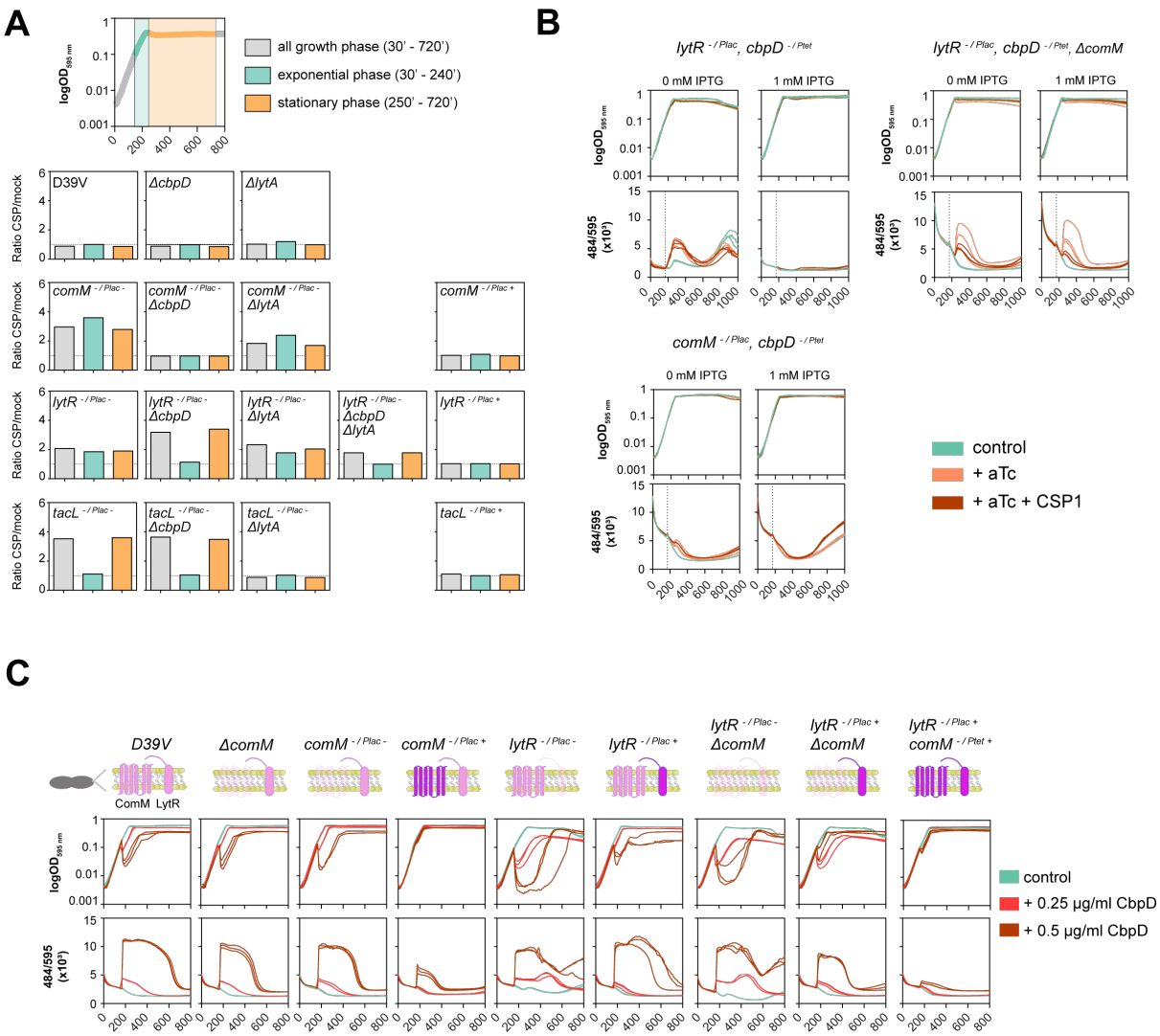


Figure 3. Both LytR and ComM are required for immunity to CbpD. (a) Cell lysis evaluation in presence and absence of the fratricins. Top: growth phase was split in three periods: exponential phase (green; from CSP1 addition to stationary phase), early stationary phase (orange) and late stationary phase (blue). Individual strains were grown in C+Y pH 6.8 to avoid natural competence development in presence of SYTOX™ Green Dead Cell Stain dye. When cell cultures reached OD595 nm 0.1 (~ after 170 min), 0 and 100 ng/ml CSP1 was added to the medium, and the fluorescence ratio (+CSP1 / mock) was calculated. Three biological replicates per condition are shown. As fratricide occurs shortly after competence induction, it should be detected in the first period, while autolysis via LytA is detected during early and/or late stationary phase. **(b)** Induced cell lysis by overexpression of cbpD (cbpD-/Ptet+). Individual strains were grown in C+Y pH 6.8 to avoid natural competence development in the presence of SYTOX™ Green Dead Cell Stain dye. When cell cultures reached OD595 nm 0.1 (~ after 170 min), 0.5 µg/ml of aTc (orange) or 0.5 µg/ml of aTc plus 100 ng/ml of CSP1 (red) was added to induce cbpD expression and/or competence. Three biological replicates per condition are shown. **(c)** Induced cell lysis by addition of purified CbpD. Individual strains were grown in C+Y pH 6.8 to avoid natural competence development in presence of SYTOX™ Green Dead Cell Stain dye. When cell cultures reached OD595 nm 0.1 (~ after 170 min), 0.25 µg/ml (red) or 0.5 µg/ml of purified CbpD (garnet) was added to the medium. Three biological replicates per condition are shown.

ComM is required for LytR activity

(DS aided in these experiments)

Our results show that ComM and LytR play a major role protecting cells during competence. During exponential growth, the action of TacL ensures plentiful production of LTAs (Flores-Kim et al. 2019) (Figure 2A). Competence activation, and subsequent upregulation of ComM and LytR, might change the ratio between LTA and WTA (Figure 4 - Model, see comments in Figure 2) and this increased WTA might be the mechanism underlying the protection from the PGN hydrolytic activity of CbpD. To test this hypothesis, we measured the incorporation of TAs in the membrane and cell wall using radioactive Methyl-^{3}H-Choline, as it is only incorporated in TAs (Tomasz 1967) (Figure 5A).

Twenty minutes after competence induction in D39V, a significant increase in the WTA/LTA ratio was observed (Figure 5B, Table S3). This shift was abolished when competence was induced in absence of ComM (we used the *ΔcomM ΔcbpD* double mutant to avoid fratricide, Figure S7A). In this strain, *comM* was replaced by a promoterless kanamycin-resistance cassette in frame, thus, levels of LytR are still highly increased after CPS_1 addition. Indeed, an increase in WTA levels was observed in the *ΔcomM ΔcbpD* mutant strain when competence was triggered; however, those levels are significantly lower compared to D39V (Figure 5B), confirming that ComM is important for LytR activity.

To analyse the role of both ComM and LytR in more detail, we tested ^{3}H-Cho incorporation in strains with different expression levels of both genes (Figure 5C). Interestingly, *comM* overexpression resulted in a reduction of LTA amounts relative to WTA, independent of the *lytR* expression levels (Figure 5C). Contrary, in absence of *comM* or in its native expression, the amounts of LTA remained similar than in D39V. When cells were slightly depleted for LytR in native (low) *comM* levels, no significant differences were observed (Figure 5C). Interestingly, overexpression of ComM promotes a significant increase of WTAs in both native and overexpression levels of *lytR*. In the same line, overexpression of LytR also resulted in increased amounts of WTA in basal levels of ComM (Figure 5C, Table S3).

To complement these results, we also tested the depletion and overexpression of *tacL* (Figure 5D). In absence of TacL, TAs cannot be anchored in the membrane, thereby providing more substrate for LytR to produce WTA (Figure 4 - model). In support of this model, we find a shifted ratio of WTA/LTA in favour of WTA when TacL was depleted from the cells (Figure 5D). However, in the membrane fraction we still detected the TAs that are not yet anchored in the cell wall, in addition to the precursors anchored in the inner part of the membrane (Figure 1C). As expected, overexpression of TacL does not have any impact on LTA production during exponential growth, since most of the TAs are already anchored to the (Flores-Kim et al. 2019) (Figure 5D, Table S3).

Together, this data shows that ComM and LytR are essential for WTA production. The exact mechanism by which ComM works to aid LytR is unknown. One possibility is that ComM shuttles TA precursors from the TA flippase (TacF) to LytR, ComM stimulates the activity of LytR, or that ComM inhibits the activity of TacL thereby increasing the flux of TA precursor available for LytR (Figure 4 - model).

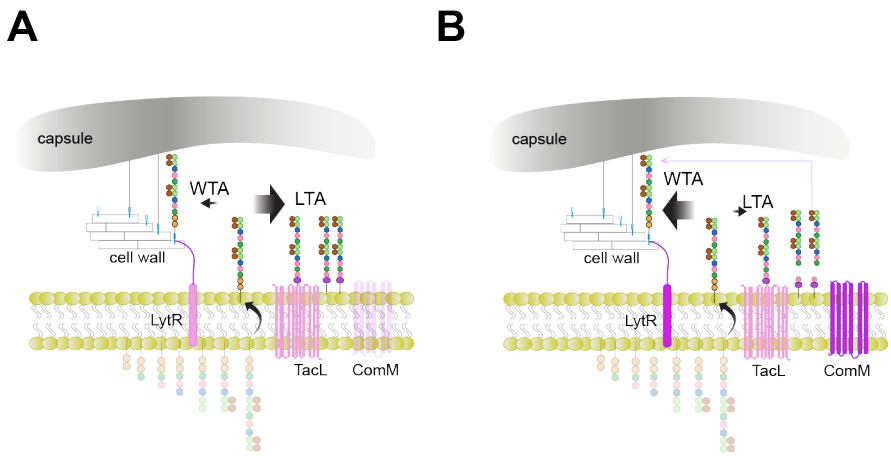


Figure 4. Proposed model for TA biosynthesis. (a) During exponential growth, TacL anchors TA to the membrane glycolipids and produces LTAs **(b)** Upon competence activation,

ComM expression and ComM-stimulated activation of LytR shift the ratio between LTA and WTA, producing more WTA.

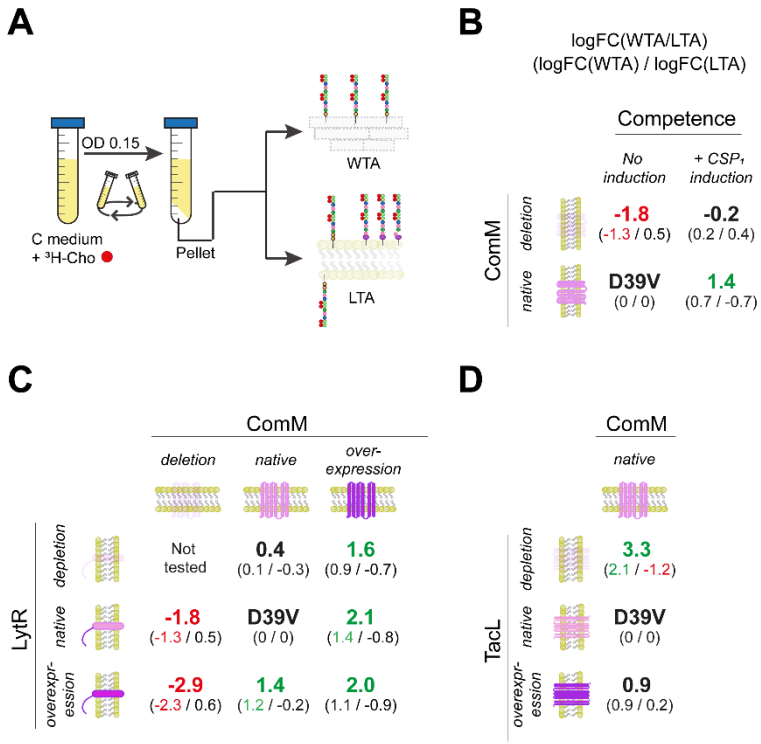


Figure 5. Incorporation of 3H-Cho into teichoic acids. (a) Protocol to detect 3H-Cho. Cells were grown in C medium supplemented with 3H-Cho until OD595 nm 0.15. The pellet was resuspended in 2 ml 50 mM MES buffer and divided in two equally sized aliquots, one for membrane (including LTAs and TA precursors) and one for cell wall isolation (see methods section for more details). **(b)** 3H-Cho detection after competence induction. Membranes and cell walls were isolated in order to quantify the amount of 3H-Cho incorporated. Average of three independent biological replicates and two technical replicates is shown (raw data in table S3). In large is shown the ratio in log fold change between WTA and LTA normalized to wt cells D39V (logFC(WTA0/logFC(LTA) = 0/0). In small is shown the relative abundancy of WTA or LTA compared to D39V. Red indicates a significantly smaller ratio between WTA/LTA in that conditions compared to wt, whereas green shows a significantly greater ratio between WTA/LTA. **(c)** 3H-Cho detection in different LytR and ComM conditions. Double lytR depletion and comM deletion condition was not tested due to the growth defect and increased lysis of the strain (data not shown). LytR was depleted for ~ 2h to not cause too much cell lysis due to its essentiality. Average of three independent biological replicates and two technical replicates is shown (raw data in table S2). **(d)** 3H-Cho detection in different TacL conditions. Average of three independent biological replicates and two technical replicates is shown (raw data in table S2).

Discussion

Here we identified genes that become essential when competence is triggered in *S. pneumoniae*. Using CRISPRi-seq, targeting 1499 TSS in *S. pneumoniae* D39V, we found 14 sgRNAs (targeting 10 operons) underrepresented when both the library and competence were induced. Two competence-related operons, *comCDE* and *comM*, showed an expected fitness cost as the so-called immunity protein ComM cannot be produced by downregulation of both operons, and thereby fratricines rapidly lyse them. Contrary, the *comAB* operon did not show any fitness cost as cells can still sense exogenous CSP produced by competing cells and activate competence (Figure S2B). The fact that the sgRNA targeting *comX* was present in all conditions, confirms that fratricide immunity is conferred due to ComM (which is ComE-P dependent and not ComX dependent), rather than other competence-related proteins (Figure S2B).

Interestingly, most of the targeted operons were related to the TA synthesis pathways, suggesting their important role protecting cells from competence-related autolysis (Figure 1B). Other genes known to play a role in cell wall synthesis were also underrepresented during competence development but did not make the statistical cut-off (Table S1). This suggests that perturbations in general cell wall homeostasis also sensitizes pneumococci for fratricide but to a lesser extent than downregulation of TA genes. Contrary to other Gram-positive bacteria, TAs biosynthesis pathway in *S. pneumoniae* is identical for LTAs and WTAs (Denapaite, MDR 2012). As all but two TA-related operons (*SPV_1620-aatB*, and *tacL*) become more essential during competence (Figure 2B, Table S1), we demonstrate that the levels, or relative amounts, of LTA and WTA are altered when competence is triggered. This evidence is supported by the fact that TA synthesis is not affected during competence (the expression levels of the genes involved in this pathway are constant; Figure S6), and that TacL (the enzyme responsible to anchoring TAs to the membrane glycolipids) is not essential for normal growth (Heß et al. 2017).

Our results suggest a model by which competence is triggered, e.g. by host adherence or environmental stress (Aprianto et al 2016; Slager et al 2014). Upon competence activation, ComM and CbpD are expressed. The release of CbpD in the environment might aid in the release of DNA from non-competent pneumococci thereby strengthening the extracellular matrix within the microcolony/biofilm and the possibility to acquire beneficial traits or template for DNA repair. CbpD-producing cells are immune to the murimase activity by ComM-stimulated activation of LytR, which we here show cross-links TA to the cell wall thereby forming relatively more WTA. We note that WTA and the capsule compete for the same substrate on the PGN (Denapaite et al. 2012), thus activation of competence essentially will also reduce capsule linkage to the cell wall. In addition, by having more WTA relative to LTA,

the CBPs such as key adherence factor CbpA, is now better positioned on the outside of the cell to fulfill its function. Future research is needed to test this hypothesis to see whether CBPs are indeed more surface displayed during competence in a ComM-dependent manner. In addition, while it has been shown that comCDE mutants are attenuated (Hammerschmid mouse model, Jim et al. unpublished), it will be interesting to see if this effect can be entirely accounted by ComM. In any case, this work highlights the power of genome-wide CRISPRi-seq screens to unravel new functions of essential genes, such as the genes involved in TA synthesis and this would have not been possible with traditional Tn-seq-based approaches.

Materials & Methods

Bacterial strains and growth conditions

All pneumococcal strains used in this study are derivatives of the clinical isolate *S. pneumoniae* D39V (Avery, Macleod, and Mccarty 1944; Slager, Aprianto, and Veening 2018, Public Health England NCTC14078). Bacterial strains are listed in Table S2. Growth conditions of bacterial cells were described previously d. Briefly, *S. pneumoniae* was grown in C+Y medium (pH 6.8, non-permissive conditions for natural competence induction), at 37 °C and stored at –80 °C in C+Y with 14.5% glycerol at $OD_{595\ nm}$ of 0.4.

Competence assays

(DS participated in conducting this experiments)

Competence development was monitored in strains containing a transcriptional fusion of the firefly luciferase gene (*luc*) with the late competence gene *ssbB*. The *S. pneumoniae* strains were cultured in a Tecan Infinite F200 PRO allowing for real-time monitoring of competence induction *in vitro*. A pre-culture was diluted to initial OD_{595nm} 0.004 in C+Y pH 7.8 +/- 0.05 (permissive conditions for natural competence induction) containing 0.45 mg/ml of luciferine and then incubated in 96-wells microtiter plates with no shaking. Growth ($OD_{595\ nm}$) and luciferase activity (RLU) were measured every 10 min during 14 h. Expression of the *luc* gene (only if competence is activated) results in the production of luciferase and thereby in the emission of light (Domenech et al. 2020). An average of three replicates and the standard error of the mean (SEM) are shown unless indicated.

Pooled CRISPRi library

The pooled CRISPRi library containing 1499 sgRNAs targeting all the transcription start sites in *S. pneumoniae* D39V strain (Liu et al., 2021) was used to identify essential genes during competence. A pre-culture of the pooled library was grown at in C+Y medium pH 6.8 to avoid natural competence induction, at 37 °C until $OD_{595\ nm}$ of 0.1 (Figure 1A). Then, the pre-culture was diluted at $OD_{595\ nm}$ 0.005 in presence or absence of 1 mM IPTG to induce the expression of *dCas9*. After 2 h ($OD_{595\ nm}$ ~ 0.1) of library induction, 100 ng/ml CSP_1 was added in half of the samples to induce competence. Cells were incubated at 37 ºC until $OD_{595\ nm}$ 0.4, and DNA was isolated using a commercial kit (Promega®). Preparation of the Ilumina library by one-step PCR with commercial oligos and library sequencing were performed following the manufacturer instructions and explained before (Liu et al., 2021).

The fastq files generated from sequencing are available on NCBI (accession number pending). The 20 bp base-pairing sequences were trimmed out from read 1 according to their

position with Trimmomatic Version 0.36 (Bolger, Lohse, and Usadel 2014). Then the trimmed reads were mapped to a pseudogenome containing all sgRNAs sequence (Liu et al., 2021) using Bowtie 2 (Langmead and Salzberg 2012). The sgRNAs were counted with featureCounts (Liao, Smyth, and Shi 2014). The count data of sgRNAs were then analyzed with the DESeq2 package in R (https://github.com/veeninglab/CRISPRi-seq) for evaluation of fitness cost of each sgRNA. We tested against a log2FC of 1, with an alpha of 0.05.

Construction of inducible strains

(DS wrote this and constructed these strains)

For every gene related to teichoic acid synthesis, an ectopic IPTG-inducible strain was created using the pASR130 (pPEPZ::P_{lac}-TmpR-blaR) plasmid (Keller et al. 2019). Primers used to amplify the genes are listed in table S4. PCR products were digested with BsmBI, BsaI or SapI (depending on the presence of restriction sites incompatible with the cloning) and were ligated with similarly digested pPEPZ plasmid containing the P_{lac} promoter. The ligation was transformed into strain ADP95 (D39V, *prs1*::P_{F6}-*lacI-tetR, bgaA*::P_{ssbB}-*luc*; Domenech et al., 2020). All transformants were selected on Columbia blood agar containing 10 µg/ml of trimethoprim and correct colonies were verified by PCR and sequencing.

Construction of the native deletions

(DS wrote this and constructed these strains)

Every gene related to teichoic acid synthesis was replaced by an erythromycin-resistant marker, in frame with the rest of the genes in the same operon (without promoter and terminator). Primers used to amplify the genes are listed in table S4. Upstream region (~ 1Kb), downstream region (~ 1Kb) and the promoterless erythromycin marker were digested with the indicated restriction enzyme (BsmBI or BsaI, depending on the presence of restriction sites incompatible with the cloning) and were ligated. The ligation was transformed into the corresponding strain carrying the ectopic inducible construct. All transformants were selected on Columbia blood agar containing 0.5 µg/ml of erythromycin and 1 mM IPTG, and correct colonies were verified by PCR and sequencing.

In vivo radiolabelling of TA

(DS participated in conducting this experiments)

S. pneumoniae strains (D39V derivatives) were grown in C medium at pH 6.8 (to avoid natural competence development) until $OD_{595\ nm}$ of 0.1 in absence of the inducers. The pre-cultures were diluted to initial $OD_{595\ nm}$ of 0.001 in 10 ml of C medium at pH 6.8 and

supplemented with 43.7 mM of ^{3}H-choline. Final concentrations of 1 mM of IPTG and/or 500 ng/ml of aTc were added when indicated. Cells were grown at 37 °C without shaking until $OD_{595\ nm}$ of 0.15. For the conditions that required competence induction, 100 ng/ml of CSP_1 was added at $OD_{595\ nm}$ 0.10 and incubated for 20 minutes at 37 ºC. Cells were centrifuged for 5 min at 7000 x *g* and 2 ml of supernatant was collected and stored at –80 ºC. The pellet was resuspended in 2 ml of 50 mM MES and distributed in two fractions. Membrane and cell wall isolates were obtained as described before (Flores-Kim et al. 2019). Each growth condition was performed with experimental triplicates.

Radioactivity quantification

Radioactivity of cell wall and membrane isolates was measured by scintillation counting using a Liquid Scintillation Analyzer Tri-Carb 4910 TR (PerkinElmer). To do this, 2.5 ml of Ultima Gold™ XR LSC Cocktail (PerkinElmer, Waltham, MA) were mixed either with 100 ml of isolated membrane or 200 ml of cell wall isolate. Each experimental triplicate was measured twice. For background correction, 500 ml of cell culture supernatant was treated and measured in the same way.

Quantification and statistical analysis

Data analysis was performed using GraphPad Prism, Microsoft Excel, R version 2.15.1 and RStudio Version 1.0.136.

Data shown in plots are represented as mean of at least three replicates ± SEM, as stated in the figure legends. Exact number of replicates for each experiment are enclosed in their respective figure legends.

Supplementary tables

Supplementary File named S_Tables.xlsx

References

Aggarwal, Surya D., Rory Eutsey, Jacob West-Roberts, Arnau Domenech, Wenjie Xu, Iman Tajer Abdullah, Aaron P. Mitchell, Jan-Willem Veening, Hasan Yesilkaya, and N. Luisa Hiller. 2018. “Function of BriC Peptide in the Pneumococcal Competence and Virulence Portfolio.” Edited by Timothy J. Mitchell. *PLOS Pathogens* 14 (10): e1007328. https://doi.org/10.1371/journal.ppat.1007328.

Aprianto, Rieza, Jelle Slager, Siger Holsappel, and Jan Willem Veening. 2016. “Time-Resolved Dual RNA-Seq Reveals Extensive Rewiring of Lung Epithelial and Pneumococcal Transcriptomes during Early Infection.” *Genome Biology* 17 (1): 198. https://doi.org/10.1186/s13059-016-1054-5.

———. 2018. “High-Resolution Analysis of the Pneumococcal Transcriptome under a Wide Range of Infection-Relevant Conditions.” *Nucleic Acids Research* 46 (19): 9990–10006. https://doi.org/10.1093/nar/gky750.

Avery, O T, C M Macleod, and Maclyn Mccarty. 1944. “Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III.” *The Journal of Experimental Medicine* 79 (2): 137–58. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2135445&tool=pmcentrez&rendertype=abstract.

Bolger, Anthony M, Marc Lohse, and Bjoern Usadel. 2014. “Trimmomatic: A Flexible Trimmer for Illumina Sequence Data.” *Bioinformatics (Oxford, England)* 30 (15): 2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Bonnet, J., C. Durmort, I. Mortier-Barrière, N. Campo, M. Jacq, C. Moriscot, D. Straume, et al. 2018. “Nascent Teichoic Acids Insertion into the Cell Wall Directs the Localization and Activity of the Major Pneumococcal Autolysin LytA.” *Cell Surface* 2 (June): 24–37. https://doi.org/10.1016/j.tcsw.2018.05.001.

Chewapreecha, Claire, Simon R. Harris, Nicholas J. Croucher, Claudia Turner, Pekka Marttinen, Lu Cheng, Alberto Pessia, et al. 2014. “Dense Genomic Sampling Identifies Highways of Pneumococcal Recombination.” *Nature Genetics* 46 (3): 305–9. https://doi.org/10.1038/ng.2895.

Claverys, Jean-Pierre, Bernard Martin, and Patrice Polard. 2009. “The Genetic Transformation Machinery: Composition, Localization, and Mechanism.” *FEMS Microbiology Reviews* 33 (3): 643–56. https://doi.org/10.1111/j.1574-6976.2009.00164.x.

Croucher, Nicholas J, Simon R Harris, Christophe Fraser, Michael a Quail, John Burton, Mark van der Linden, Lesley McGee, et al. 2011. “Rapid Pneumococcal Evolution in Response to Clinical Interventions.” *Science (New York, N.Y.)* 331 (6016): 430–34. https://doi.org/10.1126/science.1198545.

Denapaite, Dalia, Reinhold Brückner, Regine Hakenbeck, and Waldemar Vollmer. 2012. “Biosynthesis of Teichoic Acids in *Streptococcus pneumoniae* and Closely Related Species: Lessons from Genomes.” *Microbial Drug Resistance*. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA . https://doi.org/10.1089/mdr.2012.0026.

Domenech, Arnau, Ana Rita Brochado, Vicky Sender, Karina Hentrich, Birgitta Henriques-Normark, Athanasios Typas, and Jan Willem Veening. 2020. “Proton Motive Force Disruptors Block Bacterial Competence and Horizontal Gene Transfer.” *Cell Host and Microbe* 27 (4): 544-555.e3. https://doi.org/10.1016/j.chom.2020.02.002.

Eberhardt, Alice, Christopher N. Hoyland, Daniela Vollmer, Stephanie Bisle, Robert M. Cleverley, Ola Johnsborg, Leiv S. Håvarstein, Richard J. Lewis, and Waldemar Vollmer.

2012. “Attachment of Capsular Polysaccharide to the Cell Wall in *Streptococcus pneumoniae*.” *Microbial Drug Resistance* 18 (3): 240–55. https://doi.org/10.1089/mdr.2011.0232.

Eldholm, Vegard, Ola Johnsborg, Kristine Haugen, Hilde Solheim Ohnstad, and Leiv Sigve Havastein. 2009a. “Fratricide in *Streptococcus pneumoniae:* Contributions and Role of the Cell Wall Hydrolases CbpD, LytA and LytC.” *Microbiology* 155 (7): 2223–34. https://doi.org/10.1099/mic.0.026328-0.

———. 2009b. “Fratricide in Streptococcus Pneumoniae: Contributions and Role of the Cell Wall Hydrolases CbpD, LytA and LytC.” *Microbiology* 155 (7): 2223–34. https://doi.org/10.1099/mic.0.026328-0.

Eldholm, Vegard, Ola Johnsborg, Daniel Straume, Hilde Solheim Ohnstad, Kari Helene Berg, Juan A. Hermoso, and Leiv Sigve Håvarstein. 2010. “Pneumococcal CbpD Is a Murein Hydrolase That Requires a Dual Cell Envelope Binding Specificity to Kill Target Cells during Fratricide.” *Molecular Microbiology* 76 (4): 905–17. https://doi.org/10.1111/j.1365-2958.2010.07143.x.

Fischer, Werner, Thomas Behr, Rudolf Hartmann, Jasna Peter-Katalinić, and Heinz Egge. 1993. “Teichoic Acid and Lipoteichoic Acid of S*treptococcus pneumoniae* Possess Identical Chain Structures: A Reinvestigation of Teichoid Acid (C Polysaccharide).” *European Journal of Biochemistry* 215 (3): 851–57. https://doi.org/10.1111/j.1432-1033.1993.tb18102.x.

Flores-Kim, Josué, Genevieve S. Dobihal, Andrew Fenton, David Z. Rudner, and Thomas G. Bernhardt. 2019. “A Switch in Surface Polymer Biogenesis Triggers Growth-Phase-Dependent and Antibiotic-Induced Bacteriolysis.” *ELife* 8. https://doi.org/10.7554/eLife.44912.

Guiral, Sébastien, Vincent Hénard, Chantal Granadel, Bernard Martin, and Jean Pierre Claverys. 2006. “Inhibition of Competence Development in *Streptococcus pneumoniae* by Increased Basal-Level Expression of the ComDE Two-Component Regulatory System.” *Microbiology* 152 (2): 323–31. https://doi.org/10.1099/mic.0.28425-0.

Håvarstein, Leiv Sigve, Gowri Coomaraswamy, and Donald A. Morrison. 1995. “An Unmodified Heptadecapeptide Pheromone Induces Competence for Genetic Transformation in *Streptococcus pneumoniae*.” *Proceedings of the National Academy of Sciences of the United States of America* 92 (24): 11140–44. https://doi.org/10.1073/pnas.92.24.11140.

Havarstein, Leiv Sigve, Dzung Bao Diep, and Ingolf F. Nes. 1995. “A Family of Bacteriocin ABC Transporters Carry out Proteolytic Processing of Their Substrates Concomitant with Export.” *Molecular Microbiology* 16 (2): 229–40. https://doi.org/10.1111/j.1365-2958.1995.tb02295.x.

Håvarstein, Leiv Sigve, Bernard Martin, Ola Johnsborg, Chantal Granadel, and Jean Pierre Claverys. 2006. “New Insights into the Pneumococcal Fratricide: Relationship to Clumping and Identification of a Novel Immunity Factor.” *Molecular Microbiology* 59 (4): 1297–1037. https://doi.org/10.1111/j.1365-2958.2005.05021.x.

Heß, Nathalie, Franziska Waldow, Thomas P. Kohler, Manfred Rohde, Bernd Kreikemeyer, Alejandro Gómez-Mejia, Torsten Hain, et al. 2017. “Lipoteichoic Acid Deficiency Permits Normal Growth but Impairs Virulence of *Streptococcus pneumoniae.*” *Nature Communications* 8 (1). https://doi.org/10.1038/s41467-017-01720-z.

Hui, F. M., and D. A. Morrison. 1991. “Genetic Transformation in Streptococcus Pneumoniae: Nucleotide Sequence Analysis Shows ComA, a Gene Required for Competence Induction, to Be a Member of the Bacterial ATP-Dependent Transport Protein Family.” *Journal of Bacteriology* 173 (1): 372–81. https://doi.org/10.1128/jb.173.1.372-381.1991.

Johnsborg, Ola, Vegard Eldholm, Martha Langedok Bjørnstad, and Leiv Sigve Håvarstein. 2008. “A Predatory Mechanism Dramatically Increases the Efficiency of Lateral Gene Transfer in *Streptococcus pneumoniae* and Related Commensal Species.” *Molecular Microbiology* 69 (1): 245–53. https://doi.org/10.1111/j.1365-2958.2008.06288.x.

Johnston, Calum, Bernard Martin, Gwennaele Fichant, Patrice Polard, and Jean Pierre Claverys. 2014. “Bacterial Transformation: Distribution, Shared Mechanisms and Divergent Control.” *Nature Reviews Microbiology*. Nature Publishing Group. https://doi.org/10.1038/nrmicro3199.

Kadioglu, Aras, Jeffrey N Weiser, James C Paton, and Peter W Andrew. 2008. “The Role of *Streptococcus pneumoniae* Virulence Factors in Host Respiratory Colonization and Disease.” *Nature Reviews. Microbiology* 6 (4): 288–301. https://doi.org/10.1038/nrmicro1871.

Kausmally, Louise, Ola Johnsborg, Merete Lunde, Eivind Knutsen, and Leiv Sigve Håvarstein. 2005. “Choline-Binding Protein D (CbpD) in *Streptococcus pneumoniae* Is Essential for Competence-Induced Cell Lysis.” *Journal of Bacteriology* 187 (13): 4338–45. https://doi.org/10.1128/JB.187.13.4338-4345.2005.

Kawai, Yoshikazu, Jon Marles-Wright, Robert M Cleverley, Robyn Emmins, Shu Ishikawa, Masayoshi Kuwano, Nadja Heinz, et al. 2011. “A Widespread Family of Bacterial Cell Wall Assembly Proteins.” *The EMBO Journal* 30 (24): 4931–41. https://doi.org/10.1038/emboj.2011.358.

Keller, Lance E., Anne Stéphanie Rueff, Jun Kurushima, and Jan Willem Veening. 2019. “Three New Integration Vectors and Fluorescent Proteins for Use in the Opportunistic Human Pathogen *Streptococcus pneumoniae*.” *Genes* 10 (5). https://doi.org/10.3390/genes10050394.

Kjos, Morten, Eric Miller, Jelle Slager, Frank B. Lake, Oliver Gericke, Ian S. Roberts, Daniel E. Rozen, and Jan-Willem Veening. 2016. “Expression of *Streptococcus pneumoniae* Bacteriocins Is Induced by Antibiotics via Regulatory Interplay with the Competence System.” Edited by Andreas Peschel. *PLOS Pathogens* 12 (2): e1005422. https://doi.org/10.1371/journal.ppat.1005422.

Langmead, Ben, and Steven L. Salzberg. 2012. “Fast Gapped-Read Alignment with Bowtie 2.” *Nature Methods* 9 (4): 357–59. https://doi.org/10.1038/nmeth.1923.

Larson, Thomas R., and Janet Yother. 2017. “*Streptococcus pneumoniae* Capsular Polysaccharide Is Linked to Peptidoglycan via a Direct Glycosidic Bond to β-D- *N* -Acetylglucosamine.” *Proceedings of the National Academy of Sciences* 114 (22): 5695–5700. https://doi.org/10.1073/pnas.1620431114.

Liao, Y., G. K. Smyth, and W. Shi. 2014. “FeatureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features.” *Bioinformatics* 30 (7): 923–30. https://doi.org/10.1093/bioinformatics/btt656.

Lin, Jingjun, Pyunghun Park, Hua Li, Myung Whan Oh, Iwona T. Dobrucki, Wawrzyniec Dobrucki, and Gee W. Lau. 2020. “*Streptococcus pneumoniae* Elaborates Persistent and Prolonged Competent State during Pneumonia-Derived Sepsis.” *Infection and Immunity* 88 (4). https://doi.org/10.1128/IAI.00919-19.

Liu, Xue, Clement Gallay, Morten Kjos, Arnau Domenech, Jelle Slager, Sebastiaan P Kessel, Kèvin Knoops, Robin A Sorg, Jing-Ren Zhang, and Jan-Willem Veening. 2017. “ High-throughput CRISPRi Phenotyping Identifies New Essential Genes in *Streptococcus pneumoniae* .” *Molecular Systems Biology* 13 (5): 931. https://doi.org/10.15252/msb.20167449.

Liu, Xue, Jacqueline M. Kimmey, Laura Matarazzo, Vincent de Bakker, Laurye Van Maele, Jean Claude Sirard, Victor Nizet, and Jan Willem Veening. 2021. “Exploration of Bacterial

Bottlenecks and *Streptococcus pneumoniae* Pathogenesis by CRISPRi-Seq." *Cell Host and Microbe* 29 (1): 107-120.e6. https://doi.org/10.1016/j.chom.2020.10.001.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15 (12): 550. https://doi.org/10.1186/s13059-014-0550-8.

Martin, Bernard, Anne-Lise Soulet, Nicolas Mirouze, Marc Prudhomme, Isabelle Mortier-Barrière, Chantal Granadel, Marie-Françoise Noirot-Gros, Philippe Noirot, Patrice Polard, and Jean-Pierre Claverys. 2013. "ComE/ComE~P Interplay Dictates Activation or Extinction Status of Pneumococcal X-State (Competence)." *Molecular Microbiology* 87 (2): 394–411. https://doi.org/10.1111/mmi.12104.

Pestova, E. V., L. S. Håvarstein, and D. A. Morrison. 1996. "Regulation of Competence for Genetic Transformation in S*treptococcus pneumoniae* by an Auto-Induced Peptide Pheromone and a Two-Component Regulatory System." *Molecular Microbiology* 21 (4): 853–62. https://doi.org/10.1046/j.1365-2958.1996.501417.x.

Peterson, Scott N., Chang Kyoo Sung, Robin Cline, Bhushan V. Desai, Erik C. Snesrud, Ping Luo, Jennifer Walling, et al. 2004. "Identification of Competence Pheromone Responsive Genes in S*treptococcus pneumoniae* by Use of DNA Microarrays." *Molecular Microbiology* 51 (4): 1051–70. https://doi.org/10.1046/j.1365-2958.2003.03907.x.

Schmidt, Frank, Niamatullah Kakar, Tanja C. Meyer, Maren Depke, Ilias Masouris, Gerhard Burchhardt, Alejandro Gómez-Mejia, et al. 2019. "In Vivo Proteomics Identifies the Competence Regulon and AliB Oligopeptide Transporter as Pathogenic Factors in Pneumococcal Meningitis." Edited by Paul M. Sullam. *PLOS Pathogens* 15 (7): e1007987. https://doi.org/10.1371/journal.ppat.1007987.

Sia, Richard H.P., and Martin H. Dawson. 1931. "In Vitro Transformation of Pneumococcal Types: II. the Nature of the Factor Responsible for the Transformation of Pneumococcal Types." *Journal of Experimental Medicine* 54 (5): 701–10. https://doi.org/10.1084/jem.54.5.701.

Slager, Jelle, Rieza Aprianto, and Jan Willem Veening. 2018. "Deep Genome Annotation of the Opportunistic Human Pathogen *Streptococcus pneumoniae* D39." *Nucleic Acids Research* 46 (19): 9971–89. https://doi.org/10.1093/nar/gky725.

———. 2019. "Refining the Pneumococcal Competence Regulon by RNA Sequencing." *Journal of Bacteriology* 201 (13). https://doi.org/10.1128/JB.00780-18.

Slager, Jelle, Morten Kjos, Laetitia Attaiech, and Jan-Willem Veening. 2014. "Antibiotic-Induced Replication Stress Triggers Bacterial Competence by Increasing Gene Dosage near the Origin." *Cell* 157 (2): 395–406. https://doi.org/10.1016/j.cell.2014.01.068.

Steinmoen, Hilde, Eivind Knutsen, and Leiv Sigve Håvarstein. 2002. "Induction of Natural Competence in *Streptococcus pneumoniae* Triggers Lysis and DNA Release from a Subfraction of the Cell Population." *Proceedings of the National Academy of Sciences of the United States of America* 99 (11): 7681–86. https://doi.org/10.1073/pnas.112464599.

Straume, Daniel, Katarzyna Wiaroslawa Piechowiak, Silje Olsen, Gro Anita Stamsås, Kari Helene Berg, Morten Kjos, Maria Victoria Heggenhougen, Martín Alcorlo, Juan A. Hermoso, and Leiv Sigve Håvarstein. 2020. "Class A PBPs Have a Distinct and Unique Role in the Construction of the Pneumococcal Cell Wall." *Proceedings of the National Academy of Sciences of the United States of America* 117 (11): 6129–38. https://doi.org/10.1073/pnas.1917820117.

Tomasz, Alexander. 1967. "Choline in the Cell Wall of a Bacterium: Novel Type of Polymer-Linked Choline in Pneumococcus." *Science* 157 (3789): 694–97. https://doi.org/10.1126/science.157.3789.694.

Wahl, Brian, Katherine L. O'Brien, Adena Greenbaum, Anwesha Majumder, Li Liu, Yue Chu, Ivana Lukšić, et al. 2018. "Burden of S*treptococcus pneumoniae* and *Haemophilus influenzae* Type b Disease in Children in the Era of Conjugate Vaccines: Global, Regional, and National Estimates for 2000–15." *The Lancet Global Health* 6 (7): e744–57. https://doi.org/10.1016/S2214-109X(18)30247-X.

Wyres, K. L., L. M. Lambertsen, N. J. Croucher, L. McGee, A. von Gottberg, J. Linares, M. R. Jacobs, et al. 2013. "Pneumococcal Capsular Switching: A Historical Perspective." *Journal of Infectious Diseases* 207 (3): 439–49. https://doi.org/10.1093/infdis/jis703.

Supplementary Figures

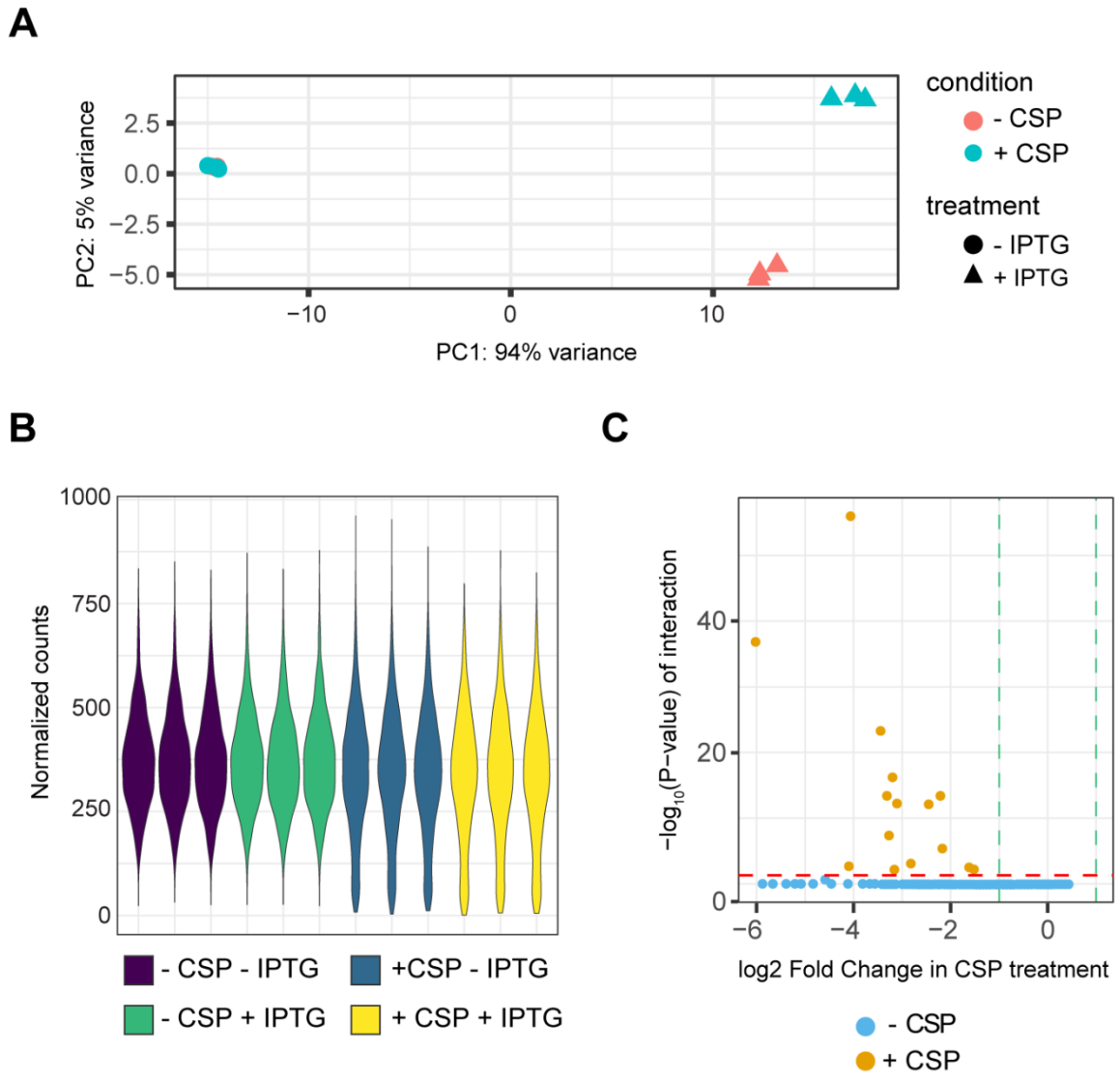


Figure S1: Evaluation of fitness cost during competence using CRISPRi pool screen. **(a-b**) Variance of the replicates. **(c)** IPTG main effect in competence by interaction of p−values. Tested against fold change of 2 (green lines) with alpha of 0.05 (red line).

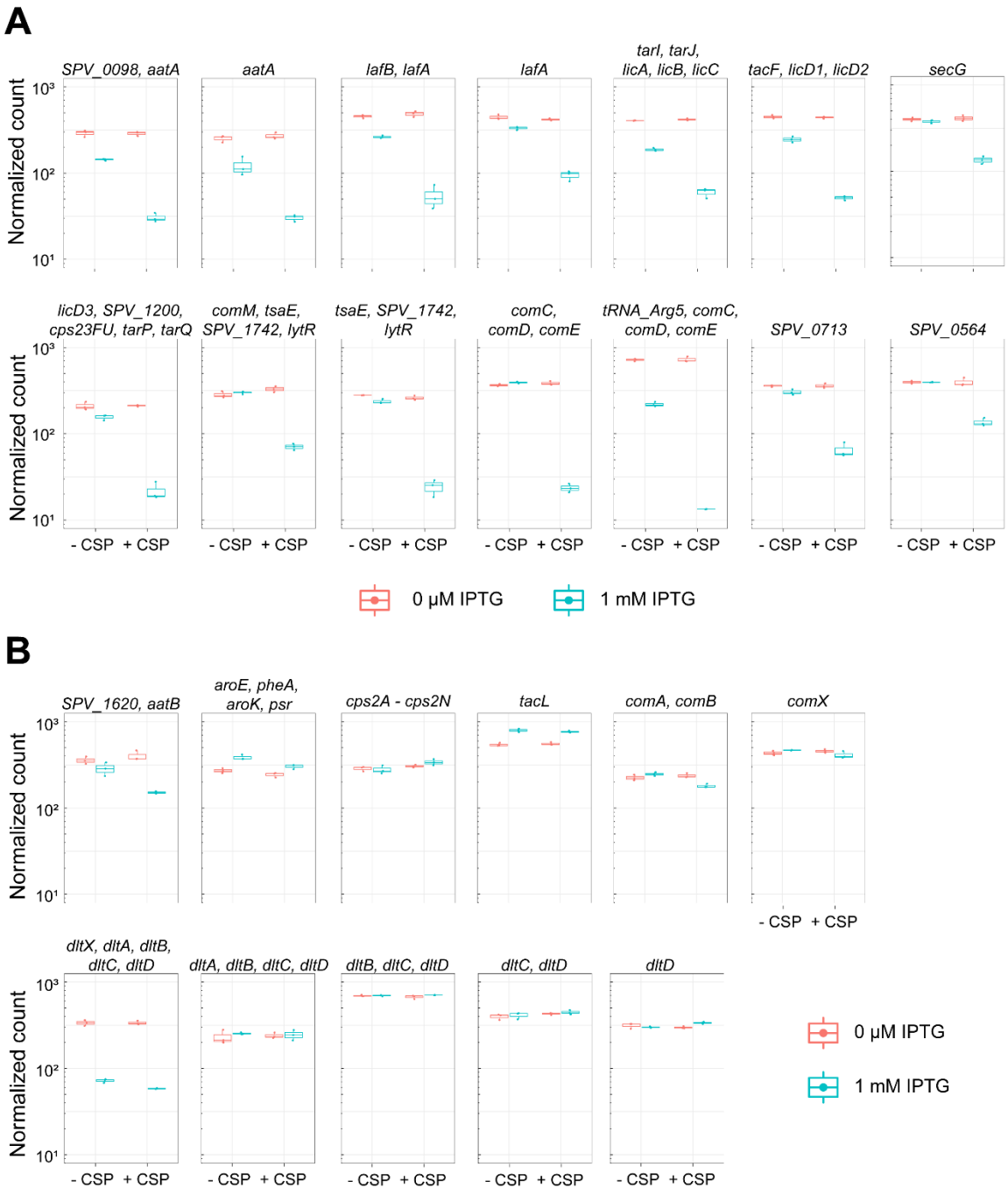
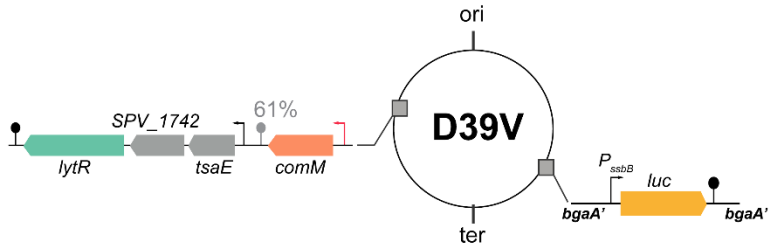
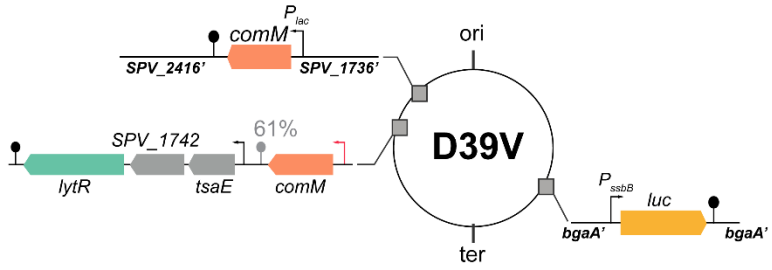


Figure S2: Normalized counts of sgRNAs related to competence and teichoic acid (TA) synthesis. (a) sgRNAs with a significant fitness cost during competence**. (b)** other sgRNAs related to competence or TA synthesis with no fitness cost. Fitness cost was evaluated (see methods for more details).

1. Introduction of the competence-specific induced *ssbB* promoter fused to firefly luciferase (P_{ssbB}-*luc*) in *bgaA* locus



2. Introduction of an IPTG-inducible ectopic copy of the gene of election (P_{lac} promoter), in ZIP region (e.g. *comM*)



3. Introduction of an erythromycin resistance cassette (*ermB*) to delete the native gene (e.g. *comM::ery*)

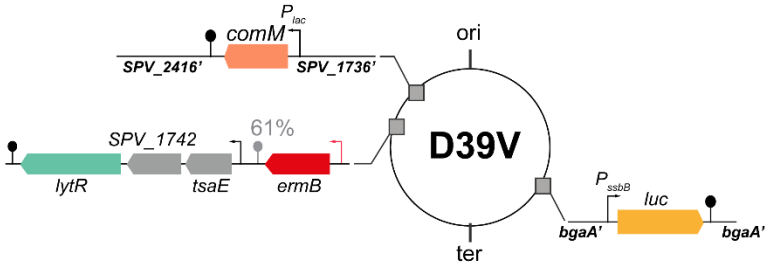
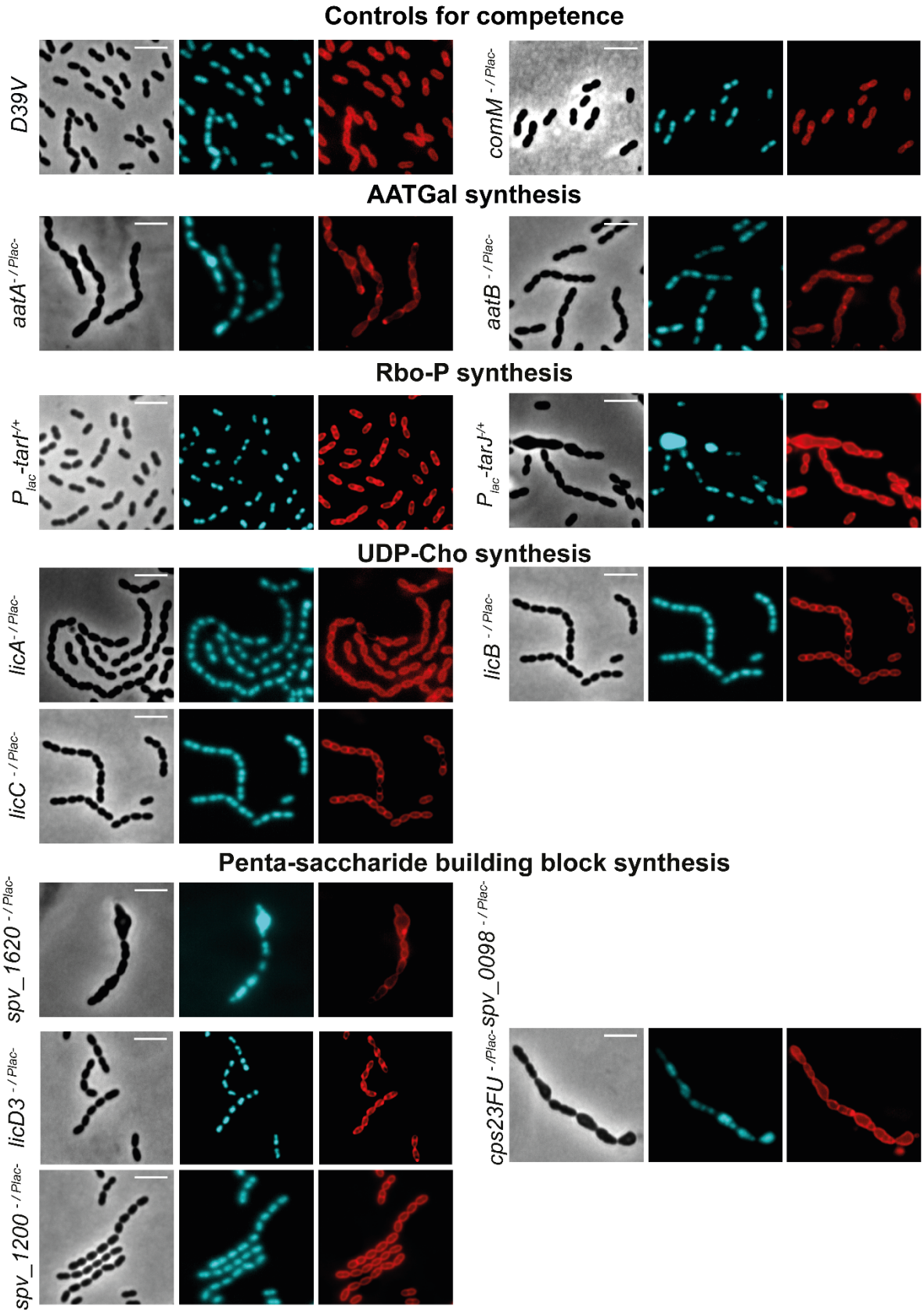


Figure S3: Design of inducible systems.


Controls for competence
D39V
comM -/Plac-
AATGal synthesis
aatA -/Plac-
aatB -/Plac-
Rbo-P synthesis
Plac-tarI+/+
Plac-tarJ+/+
UDP-Cho synthesis
licA -/Plac-
licB -/Plac-
licC -/Plac-
Penta-saccharide building block synthesis
spv_1620 -/Plac-
licD3 -/Plac-
spv_1200 -/Plac-
cps23FU -/Plac-
spv_0098 -/Plac-

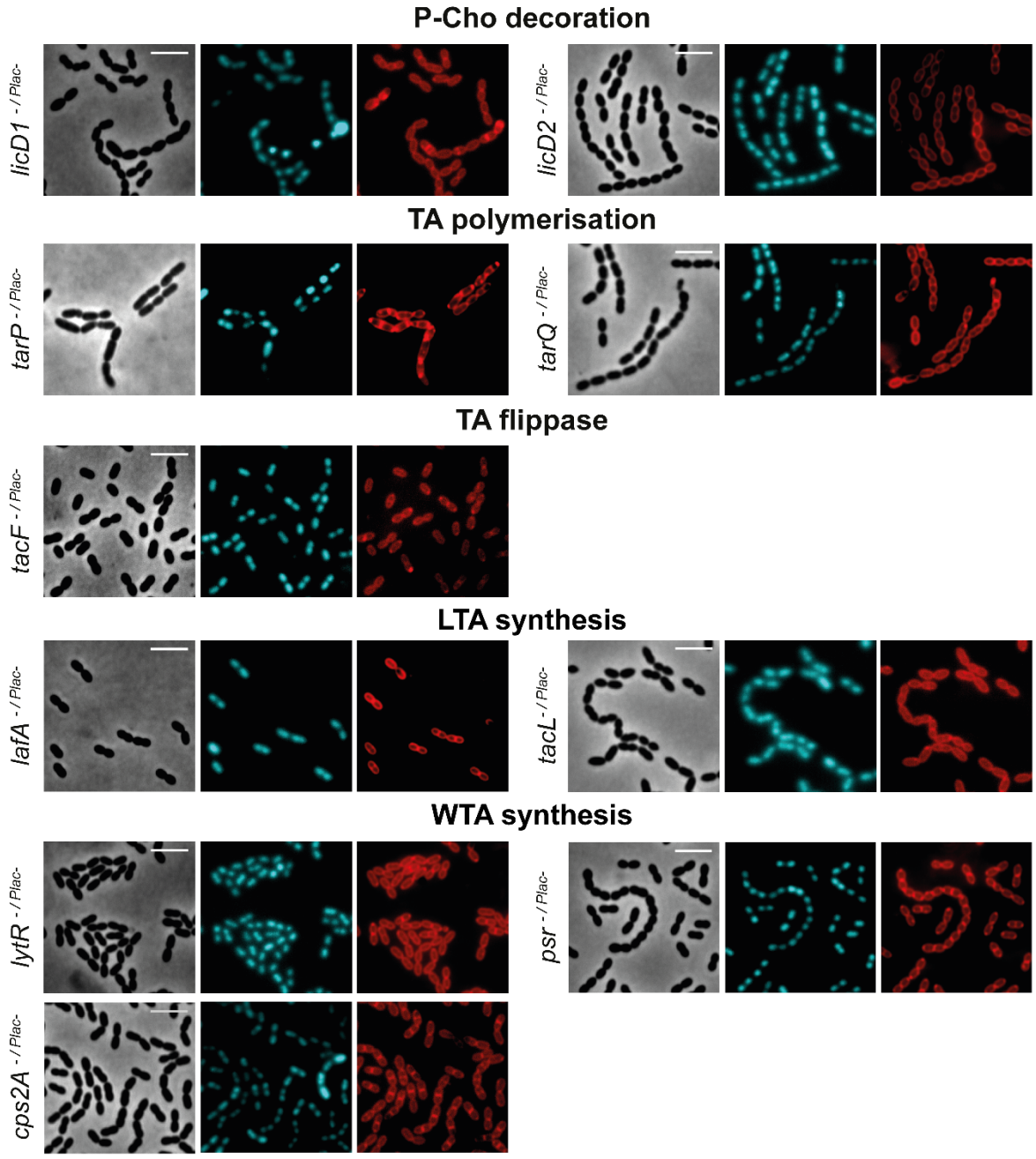


Figure S4: Morphological changes were examined with fluorescence microscopy, and representative micrographs are shown. Phase contrast, DAPI staining, and Nile red staining are displayed.

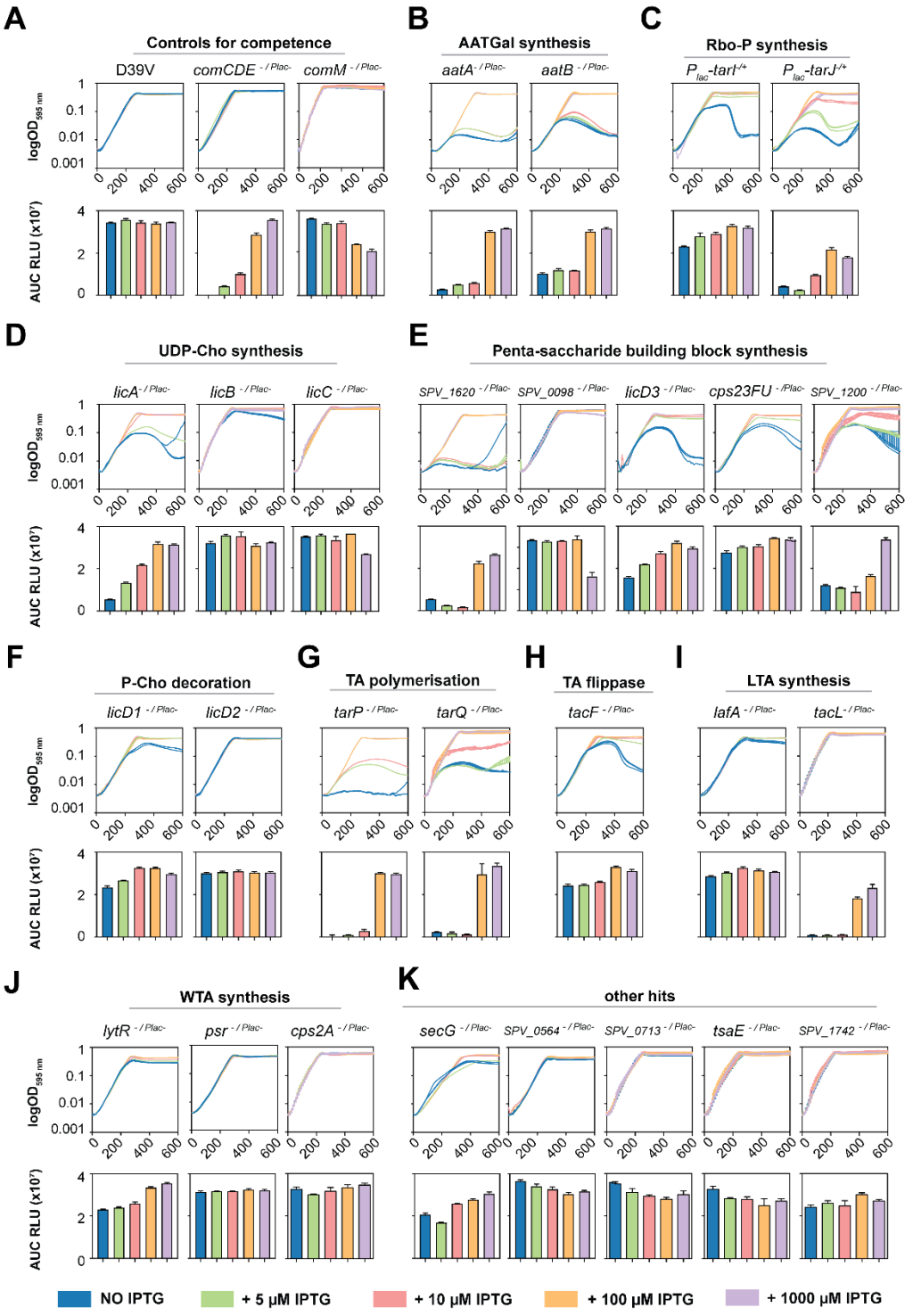


Figure S5: Natural competence development in competence and teichoic acid related genes. Every strain contains a depletion system by ectopically expressing the indicated gene under control of the Plac IPTG-inducible promoter, and the deletion of the gene from its native location. Top, growth curves in absence of the protein (blue = No IPTG) or with different IPTG concentrations. Bottom, area under the curve (AUC) of relative luminescence values (RLU).

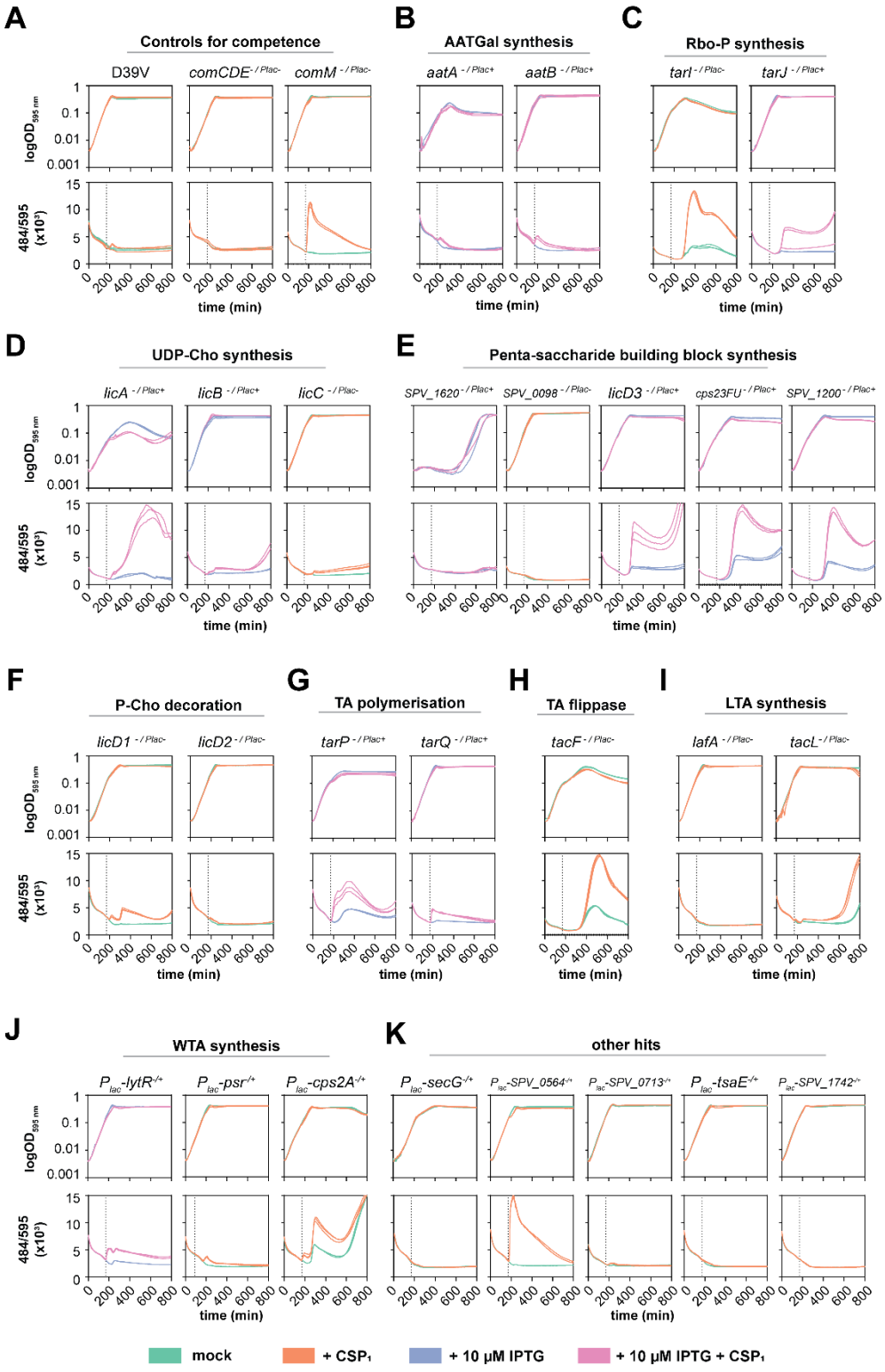


Figure S6: Detection of cell lysis. Individual strains were grown in C+Y pH 6.8 to avoid natural competence development in presence of SYTOX™ Green Dead Cell Stain dye. When $OD_{595\,nm}$ reached ~ 0.1, 100 ng/ml of CSP_1 was added to induce competence. For those genes showing less cell lysis in our setup, experiment was performed in absence of IPTG (green and orange). For those essential with severe growth defect, 10 µM IPTG was added to maintain a mild expression of the protein and avoid cell lysis in absence of CSP due to the essentiality of the gene (blue and pink).

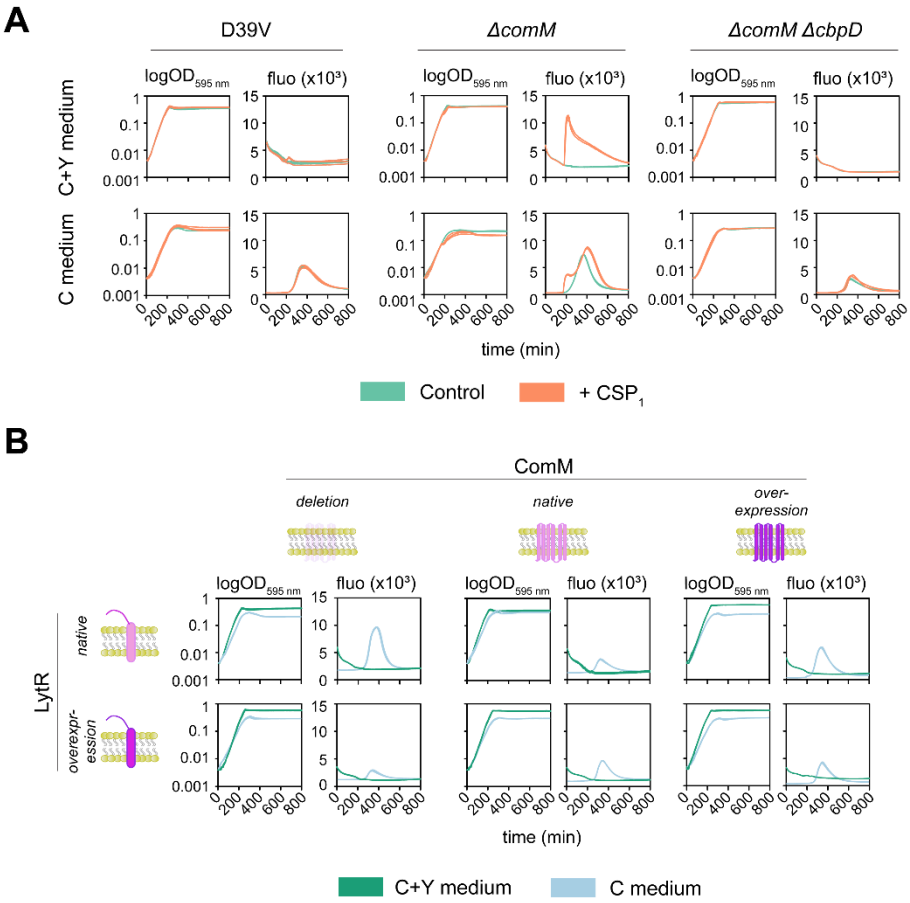


Figure S7: Cell lysis detection in C+Y and C media. A) Cell lysis detection in D39V, *ΔcomM* and double *ΔcomM ΔcbpD* strains. Individual strains were grown in C+Y medium (top) or C medium (bottom) at pH 6.8 to avoid natural competence development in presence of SYTOX™ Green Dead Cell Stain dye. When cell cultures reached $OD_{595\,nm}$ 0.1 (~ after 170 min), 100 ng/ml of CSP_1 was added to induce competence (orange lines). Three biological replicates per condition are shown. **B)** Evaluation of cell lysis in the strains used for the radioactive assay (Figure 4). Cells were grown in the indicated medium in presence of SYTOX™ Green Dead Cell Stain dye. Three biological replicates per condition are shown.

Chapter 5: Summary and Thesis Discussion

(D.S. wrote this chapter)

From science fiction to reality

The ability to make virtually any targeted change in the genome of any living cell or organism is a longstanding aspiration of the life sciences. Precise nucleic acid editing technologies are valuable for studying cell biology and potential novel therapeutics. Until recently, zinc-finger nucleases (ZFNs) were the only practical option available to researchers who are interested in targeted genome engineering in higher eukaryotes (Bibikova et al. 2003). In 2012, the proposition of programmable gene editing using CRISPR-Cas9 (Jinek et al. 2012), made this genome engineering ambition a reality. This not only caused enabled scientists to engineer higher eukaryotes, CRISPR-Cas systems also enabled genetic research in non-model microorganisms such as the here studied bacterium *Streptococcus pneumoniae* (this thesis).

CRISPR-Cas9 in pneumococcal studies

CRISPR–Cas systems are a diverse group of RNA-guided nucleases (Ishino et al. 1987) that naturally defend prokaryotes against viral invaders (Garneau et al. 2010; Barrangou et al. 2007). Currently, editing tools based on programmable CRISPR-associated nucleases, have been widely adapted for an increasing number of organisms. *Streptococcus pneumoniae,* an important human pathogen (Kadioglu et al. 2008) and a valuable model to study the cell biology of ovoid-shaped bacteria (Massidda, Nováková, and Vollmer 2013; Keller et al. 2019; Liu et al. 2017), is one of the first organisms in which RNA-guided editing by CRISPR-Cas9 system was used to introduce precise mutations in its genome (Jiang et al. 2013). Additionally, CRISPRi studies, utilizing nuclease-inactive versions of Cas9, provided high-throughput interrogation of genome-wide gene functions and genetic interactions, facilitating the study of pneumococcal essential genes and pathogenesis (Liu *et al.*, 2017, 2021). In this thesis we are interested in using the CRISPR-Cas9 technology and apply it to perform genome editing, with the ultimate goal of genome minimization and uncover novel pneumococcal biology.

The quest of the minimal genome

One consequence of progress in the field of synthetic biology is an emerging view of cells as assemblages of parts that can be put together to produce an organism with a desired phenotype (Ferber 2004). A central undertaking of the field is the quest for the ‘minimal genome’ that includes only the genes essential for life. In 2016, Craig Venter and his team designed and built a minimal genome of the bacterium *Mycoplasma mycoides* that consists

only of 531-kbp and encodes 473 gene products, a genome smaller than any autonomously replicating cell found in nature (Hutchison et al. 2016). A wide variety of experimental approaches, as well as a number of computational approaches were employed to define the core essential gene set needed to sustain life. However, the complete set of essential genes that is universally required by all organisms is difficult to determine experimentally and hard to predict. Determination of a minimal genome will lead to the construction of the more suitable cellular 'chassis' and help to create tailor-made minimal cells for industrial applications encoded by genomes composed of nothing else but essential genes.

Bacillus subtilis and *Escherichia coli* are both well characterized model bacteria used as hosts for a plethora of biotechnological applications. Over the past years, *B. subtilis* and *E. coli* were subjected to genome reductions that often led even to unexpected beneficial traits, while not having negative impacts on fitness. Significant progress has been made in the genome reduction of both model bacteria in recent years. Taken together, several studies showed that large regions of the *E. coli* and *B. subtilis* genomes are dispensable under standard laboratory growth conditions. These two bacteria have genomes sizes of ~4.2 Mbp and ~4.6Mbp respectively (Kunst et al. 1997; Blattner et al. 1997). After genome reduction, a genome size of 2.98 Mbp and 2.68 Mbp, respectively, have been reported (Hirokawa et al. 2013 and Reuß et al. 2017). A common approach for the restructuring of a microbial genome with a scope to create custom-designed microorganisms is sequence-specific deletions. Various methods have been developed, with the desired disruptions commonly made by inserting a drug resistance gene that provides direct selection of the recombinants. However, the inserted selection markers, while powerful, they also have drawbacks, preventing further modifications of the genome, which limits the repeated use of these procedures. Also, many important categories of gene mutation, such as missense substitutions and in-frame deletions, usually present no selectable phenotype (Sung et al. 2001). Our approach is to use the CRISPR technology, which has revolutionized conventional genetic engineering methods and unprecedentedly facilitated strain engineering.

CRISPR-Cas9 system for genome editing in *S. pneumoniae*

To reach our goal, we have designed and developed a CRISPR-Cas9-based system for advanced and markerless genome engineering. To achieve this, we first needed to establish an inducible CRISPR-Cas9 system in pneumococcus. Firstly, we designed and set up a CRISPR-cas9 system which performs with very high efficiency DSB in pneumococcus (Chapter 2). We explored two options to harness the CRISPR-Cas9 system after the DSB; NHEJ and HR.

NHEJ does not facilitate repair from DSB by Cas9

Our first approach was to establish the NHEJ system in pneumococcus and exploit it, alongside with CRISPR-Cas9, to remove genes or large genomic regions. Two sgRNAs, targeting on both sides of the deletion target, would guide Cas9 to introduce the DSBs. Subsequently, the NHEJ mechanism would facilitate the repair by re-ligating the two ends, leaving out the deletion target. This way we could have markerless deletions with only requirement, every time, two specific sgRNAs. In Chapter 2, we transplanted four different NHEJ systems from other bacteria to *S. pneumoniae.* Despite our efforts and several different approaches, though, we did not manage to setup functional NHEJ and demonstrate DSB repair.

The physiological role of bacterial NHEJ is to offer protection to species that live in stressful environments, when only a single copy of the genome is available, such as after sporulation or during stationary phase (Moeller et al. 2007; Pitcher et al. 2007). Therefore, NHEJ can confer resistance to Ionizing Radiation (IR), to harmful effects of desiccation or to genotoxic defense mechanisms of macrophages (Weller et al. 2002; Rengarajan, Bloom, and Rubin 2005; Pitcher et al. 2007). However, the NHEJ repair pathway is not conserved across bacterial species, suggesting that it is nonessential. Indeed, deletion of the NHEJ complex from species that have this pathway does not result in any obvious growth defects under normal growth conditions (Bowater and Doherty 2006). While multicellular eukaryotes tend to directly re-ligate DSB using NHEJ, bacteria primarily rely on homologous recombination. Additionally, eukaryotic organisms utilize a large number of factors to repair breaks by NHEJ. In contrast, the minimal bacterial NHEJ complex is a two-component system that possesses all of the break-recognition, end-processing, and ligation activities required to facilitate the DSB repair (Pitcher, Brissett, and Doherty 2007). It is possible that more interactions with other systems are required. Furthermore, it is suggested that the bacterial NHEJ can be more complex than the minimal Ku-LigD system, thought to operate (Bhattarai, Gupta, and Glickman 2014).

Although it has been demonstrated that in nature there is a negative association between NHEJ and type II-A CRISPR-Cas, with one single case of co-occurrence of both systems (Bernheim et al. 2017), the interaction between the NHEJ system and Cas9 is at the heart of the CRISPR-Cas-based genetic engineering technologies, which rely on the repair of DNA breaks generated by Cas nucleases. For our ultimate goal, the NHEJ system could be valuable for one step genome minimization. In Chapter 2 we discuss potential strategies to tackle the emerging problems, keeping in mind that in Eukaryotes, such an interplay is possible. Finally, we could also consider transplanting the more complicated NHEJ system of yeast (Daley et al. 2005).

Harnessing CRSIPR-Cas9 and HR for advance genome engineering

After excluding the NHEJ as a viable option for genome engineering in the pneumococcus, we explored the idea of using CRISPR-Cas9 editing with HR templates to repair the DSB, which was proven to be successful previously (Jiang et al. 2013). Gene and chromosomal deletions were selected as a demonstration of gene editing. With this scope, in Chapter 3, we developed a replicative plasmid with a broad host-range temperature-sensitive origin of replication, carrying a concise CRISPR-Cas9-based system for advanced and markerless genome engineering (Synefiaridou and Veening 2021). Using CRISPR-Cas9-Mediated counterselection, we successfully deleted genes and large chromosomal regions in a precise and sequential way. The plasmid that we designed has the temperature sensitive origin of replication *pG*$^{+}$*host*, facilitating the elimination of the plasmid from the strains, upon the desired deletion. This feature allowed us to successfully perform consecutive deletions, since we are ultimately interested to remove multiple genes and chromosomal regions from the genome. Additionally, we demonstrated the advantages by using our system, compared to just performing natural transformation without counterselection, which would be an alternative for clean deletions. Finally, using our plasmid, we created a new a strain with three important virulence factors removed, proving that our CRISPR-Ca9 system has shown robustness and flexibility in genetic manipulation of the bacterial genome. This final strain was confirmed with WGS, eliminating also a major concern of the CRISPR-Cas9 system, the off target DSB.

Promising models for pneumococcal pathogenesis

To elucidate pneumococcal virulence, different experimental murine models have been developed (Gerber et al. 2001; Chiavolini, Pozzi, and Ricci 2008; Mook-Kanamori et al. 2012). These murine models, though, have limitations, including ethical issues, high costs and time consuming experiments, making mice not suitable for large-scale screening (Chiavolini, Pozzi, and Ricci 2008).

A future plan is that the newly constructed strain will be injected in the host model *Galleria mellonella.* The larvae of the wax moth *Galleria mellonella* is an informative *S. pneumoniae* infection model for investigating *in vivo* aspects of bacterial–host interactions, such as the role of antimicrobial peptide activity and resistance (Evans and Rozen 2012). In short, *Galleria* are injected with different number of bacteria from different strains. Treated *Galleria* are incubated at 37^{o}C and the number of dead larvae is recorded. Larvae that do not move in response to touch are considered to be dead. Infection dynamics can be determined by obtaining bacterial counts from larvae over a time course. Furthermore, differences in virulence between serotypes could be distinguished in this host. Additionally, infection with strains differing in known virulence factors can demonstrate differences in virulence (Evans and Rozen 2012).

Therefore, by using different combinations of virulence factors and this straightforward host model, we can identify the minimal requirements for pneumococcus to become virulent and cause disease. Preliminary data, not included in this thesis, confirm that D39V pneumococcal cells injected in Gallerias can kill them within 24 hours, while the control with PBS does not, excluding that it is the injection that is harmful.

Additionally, it has also been demonstrated that the nematode *Caenorhabditis elegans* is a facile and inexpensive model host for several Gram-positive human bacterial pathogens, including *S. pneumoniae* (Garsin et al. 2001). In closely related to pneumococcus, *S. pyogenes*, it has been shown that the killing of *C. elegans* is solely mediated by hydrogen peroxide. Specifically, the killing required live streptococci and its capacity depends on the amount of hydrogen peroxide produced, while it can be inhibited by catalase (Jansen et al. 2002; Bolm et al. 2004). Preliminary data, not included in this thesis, do not confirm these findings, since we could not achieve killing by *S. pneumoniae*.

Alternatively, another promising model to study pneumococcal virulence, is the zebrafish embryo meningitis model (Jim et al. 2016). Zebrafish has emerged as an important model of vertebrate development and human disease, due to its small size, rapid generation time, powerful genetic systems, and genomic resources. It has a well-developed adaptive and innate cellular immune systems, which makes the zebrafish an ideal model for the study of infectious diseases (Renshaw and Trede 2012). The transparent phenotype of zebrafish embryos offers a unique advantage to this model, enabling the study of host-pathogen interaction in real time and the wide range of available fluorescent tools (H. Meijer and P. Spaink 2011). Other advantages include high fecundity of the zebrafish, external development of the embryo and availability of gene-editing tools and tools to manipulate gene expression (Nasevicius and Ekker 2001). A pneumococcal meningitis model in zebrafish embryos that allows for real-time investigation of early host-microbe interaction has been developed and characterized (Jim et al. 2016). Therefore, this model can also be used in our future virulence studies.

CRISPRi-seq is a powerful tool to investigate fratricide

In Chapter 4, we utilized as a tool another version of CRISPR-Cas9, namely CRISPR interference (CRISPRi) combined with next generation sequencing (CRISPRi-seq). We applied this new technology to identify pneumococcal genes important for competence and investigate the molecular mechanism underlying immunity against fratricide. As discussed in the Introduction, fratricide is the phenomenon by which competent pneumococci lyse non-competent pneumococci by the secretion of the amidase CbpD. As CbpD is under control of the alternative competence-dependent sigma factor ComX, this only occurs during

competence. Competent cells are protected by the action of CbpD by expression of the immunity factor ComM, which is under control of ComE. How ComM provides immunity from fratricide was unknown. The results from our CRISPRi-seq screen suggested that ten operons were essential during competence. The screen included the expected hits, like *comCDE,* essential for competence and *comM,* since in the absence of immunity protein ComM the fratricines rapidly lyse the cells. The rest of the hits, interestingly, were all operons related to teichoic acid (TA) synthesis. Since TAs are not required for competence development, they probably have another function, suggesting that they are involved in protecting the cells during 'attack' of competence-induced fratricins (i.e., CbpD, LytA).

Evaluating the hits with further experiments, we either excluded their role in fratricide (i.e., TsaE and SPV_1742), or revealed their important role, like in the case of LytR, in incorporating TAs into the cell wall and subsequently protecting from fratricins. We showed that both ComM and LytR act together to protect the cells from fratricide. Moreover, our data suggest that ComM and LytR are essential for WTA production, and that ComM is required for LytR activity, although the exact mechanism by which ComM works to aid LytR remains unknown. We hypothesize that ComM shuttles TA precursors from the TA flipase (TacF) to LytR, or that ComM inhibits the activity of TacL (the protein responsible to anchor TAs to the membrane glycolipids), thus increasing the flux of TA precursor available for LytR (Chapter 4).

CRISPR-Cas alternatives for genome engineering in *S. pneumoniae*

CRISPR–Cas technologies have enabled programmable gene editing in eukaryotes and prokaryotes. To date, Cas9 is the leading nuclease driving these advances. In pneumococcal research, only Cas9 and the catalytically impaired, dCas9 (D10A and H840A), have been harnessed for genome editing and transcriptional regulation, respectively. Another version of Cas9, a catalytically impaired Cas9 (H840A), is a nickase endonuclease that nicks the PAM-containing strand. This nickase is used in base editing, a form of genome editing that enables direct, irreversible conversion of one base pair to another at a target genomic locus without requiring double-stranded DNA breaks (DSBs), homology- directed repair (HDR) processes, or donor DNA templates (Komor et al. 2016, 2017; Nishida et al. 2016). Compared with standard genome editing methods to introduce point mutations, base editing can proceed more efficiently (Komor et al. 2016).

An option to introduce precise and targeted SNPs in the pneumococcal genome is the programmable base editing of A•T → G•C without DNA cleavage. This can be achieved by an RNA adenosine deaminase, fused to a Cas9 nickase. The deamination of adenine, yields inosine and is treated as guanine by polymerases adenine base editors (ABEs) and is enabling the direct, programmable introduction of four transition mutations without double-stranded

DNA cleavage (Gaudelli et al. 2017). By using this base editor, we will be able to easily introduce stop codons in genes, muting them for further studies, without disturbing the expression of other genes or risking other phenotypic effects.

Another option, for introducing SNPs in the pneumococcal genome, is prime editing, a 'search-and-replace' genome editing technology that directly writes new genetic information into a specified DNA site. For this, it is using the Cas9 nickase. This catalytically impaired Cas9 is fused to an engineered reverse transcriptase and is programmed with a prime editing guide RNA (pegRNA) that both specifies the target site and encodes the desired edit. Currently, it is feasible to perform all the eight transversion mutations (C→A, C→G, G→C, G→T, A→C, A→T, T→A, and T→G) (Anzalone et al. 2019). Interestingly, in other organisms, like humans, prime editing might help researchers tackle nearly 90% of the more than 75,000 disease-associated DNA variants listed in ClinVar, a public database developed by the US National Institutes of Health (Anzalone et al. 2019).

Apart from Cas9, which is a Class 2 CRISPR system, Class 1 systems also hold great potential for editing technologies, despite consisting of multi-subunit complexes and open new avenues for CRISPR-based applications in bacteria. A new system utilizing a processive nuclease Cas3, together with a minimal Type I-C Cascade-based system has been proposed for targeted genome engineering in *Pseudomonas aeruginosa* (Csörgő et al. 2020). DNA cleavage guided by a single CRISPR RNA generated large deletions with high efficiency. Specifically, a minimal Cascade–Cas3 system, using only a single crRNA with modified repeat sequences, generated bidirectional deletions originating from the programmed site, which was exploited to reduce the *P. aeruginosa* genome. Additionally, large deletion boundaries were efficiently specified by a homology-directed repair template during editing with Cascade–Cas3, but not Cas9 (Csörgő et al. 2020). This system seems very promising, and we are highly interested in establishing it in *S. pneumoniae,* to utilize it as an alternative for genome minimization, which has been the main goal of this thesis.

Thesis Summary and Outlook

The research presented in this thesis aimed at paving the way for CRISPR-Cas applications in *S. pneumoniae.* In Chapter 2, we demonstrate a newly established CRISPR-Cas9-sgRNA system, integrated in the pneumococcal genome, which can successfully introduce DSBs. Next, we attempted to also establish a NHEJ system, by transplanting well-described NHEJ machineries from other organisms. Although we nearly exhausted the possible approaches, our efforts were not fruitful. Our suggested strategy to resolve this, is to transfer our experimental setup to *Bacillus subtilis* and, in parallel, clone the machinery in pneumococcus from the CIL locus, to the ZIP locus, which at the time was not established.

In Chapter 3, we transferred the CRISPR-Cas9-sgRNA system to a not integrated pneumococcal plasmid with a temperature sensitive origin of replication. After we demonstrated that the plasmid can be propagated in *S. pneumoniae* and consequently be cured from the cells when grown in the non-permissive temperature, our goal was to delete well-known virulence factors. By harnessing this and the native homology repair system, we successfully performed precise, multiplex and markerless deletions of genes and large genomic regions with great efficiency. Our plasmid is promising to be applicable to other Gram-positive bacteria as well, with minor modifications, like exchanging the specific to pneumococci zinc-inducible promoter of Cas9, by an alternative inducible promoter for the species of interest. Eight years after researchers first demonstrated that Cas9 proteins could be loaded with RNA to cleave DNA targets , the proverbial CRISPR toolbox continues to expand (Hanna and Doench 2020). Recent discoveries offer us new opportunities for advanced genome editing. Programmable base editing, either with an RNA adenosine deaminase or with a prime editing guide RNA (pegRNA) can enable us to easily introduce STOP codons or SNPs, facilitating pneumococcal research. Additionally, in our pursuit for a minimal pneumococcal genome, which has been the initial inspiration of this thesis, we are now not limited to Cas9. We are highly interested in setting up in pneumococcus a minimal Cascade–Cas3-crRNA system to perform large genomic deletions.

In Chapter 4, we used dCas9, this time not for genome editing, but to study the cellular processes of competence and fratricide, demonstrating at the same time the versatility of the CRISPR-Cas technologies. Performing a genome-wide CRISPRi-seq targeting 1499 TSS, we evaluated the hits that were underrepresented when both the library and competence were induced, expecting that the bacteria undergoing transcriptional downregulation of genes important for immunity against fratricins would be outcompeted or lysed by fratricins produced by bacteria in which the sgRNA targets a neutral gene. After rejecting false positives, or confirming the hits, we started investigating in detail these genes, with the majority of them being involved in TA synthesis pathways. Specifically, we demonstrate that the levels, or relative amounts, of LTA and WTA are altered when competence is triggered, proposing a model in which competence activation during infection is crucial to transfer important virulence factors to the outside of the pneumococcal cell to ensure better adherence or immune evasion. More functional assays to support our findings are currently undergoing. This chapter showed the possibilities of CRISPR intereference as otherwise essential genes (such as many TA genes) can be probed for additional roles (in this case their role in protecting from autolysins). We anticipate that CRISPR-cas9 based genome-wide screens will be more and more applied in microbial research, similarly to the way transposon-insertion sequencing studies have transformed genetic screening.

References

Anzalone, Andrew V., Peyton B. Randolph, Jessie R. Davis, Alexander A. Sousa, Luke W. Koblan, Jonathan M. Levy, Peter J. Chen, et al. 2019. “Search-and-Replace Genome Editing without Double-Strand Breaks or Donor DNA.” *Nature* 576 (7785): 149–57. https://doi.org/10.1038/s41586-019-1711-4.

Barrangou, Rodolphe, Christophe Fremaux, Hélène Deveau, Melissa Richards, Patrick Boyaval, Sylvain Moineau, Dennis A. Romero, and Philippe Horvath. 2007. “CRISPR Provides Acquired Resistance against Viruses in Prokaryotes.” *Science (New York, N.Y.)* 315 (5819): 1709–12. https://doi.org/10.1126/science.1138140.

Bernheim, Aude, Alicia Calvo-Villamañán, Clovis Basier, Lun Cui, Eduardo P. C. Rocha, Marie Touchon, and David Bikard. 2017. “Inhibition of NHEJ Repair by Type II-A CRISPR-Cas Systems in Bacteria.” *Nature Communications* 8 (1): 2094. https://doi.org/10.1038/s41467-017-02350-1.

Bhattarai, Hitesh, Richa Gupta, and Michael S. Glickman. 2014. “DNA Ligase C1 Mediates the LigD-Independent Nonhomologous End-Joining Pathway of *Mycobacterium smegmatis*.” *Journal of Bacteriology* 196 (19): 3366–76. https://doi.org/10.1128/JB.01832-14.

Bibikova, Marina, Kelly Beumer, Jonathan K. Trautman, and Dana Carroll. 2003. “Enhancing Gene Targeting with Designed Zinc Finger Nucleases.” *Science* 300 (5620): 764. https://doi.org/10.1126/science.1079512.

Bikard, David, Asma Hatoum-Aslan, Daniel Mucida, and Luciano a. Marraffini. 2012. “CRISPR Interference Can Prevent Natural Transformation and Virulence Acquisition during in Vivo Bacterial Infection.” *Cell Host and Microbe* 12 (2): 177–86. https://doi.org/10.1016/j.chom.2012.06.003.

Blattner, F R, G Plunkett, C A Bloch, N T Perna, V Burland, M Riley, J Collado-Vides, et al. 1997. “The Complete Genome Sequence of Escherichia Coli K-12.” *Science (New York, N.Y.)* 277 (5331): 1453–62. http://www.ncbi.nlm.nih.gov/pubmed/9278503.

Bolm, Maike, W. T M Jansen, Ralf Schnabel, and Gursharan S. Chhatwal. 2004. “Hydrogen Peroxide-Mediated Killing of *Caenorhabditis elegans*: A Common Feature of Different Streptococcal Species.” *Infection and Immunity* 72 (2): 1192–94. https://doi.org/10.1128/IAI.72.2.1192-1194.2004.

Bowater, Richard, and Aidan J Doherty. 2006. “Making Ends Meet: Repairing Breaks in Bacterial DNA by Non-Homologous End-Joining.” *PLoS Genetics* 2 (2): e8. https://doi.org/10.1371/journal.pgen.0020008.

Chiavolini, Damiana, Gianni Pozzi, and Susanna Ricci. 2008. “Animal Models of *Streptococcus pneumoniae* Disease.” *Clinical Microbiology Reviews* 21 (4): 666–85.

https://doi.org/10.1128/CMR.00012-08.

Csörgő, Bálint, Lina M. León, Ilea J. Chau-Ly, Alejandro Vasquez-Rifo, Joel D. Berry, Caroline Mahendra, Emily D. Crawford, Jennifer D. Lewis, and Joseph Bondy-Denomy. 2020. “A Compact Cascade–Cas3 System for Targeted Genome Engineering.” *Nature Methods* 17 (12): 1183–90. https://doi.org/10.1038/s41592-020-00980-w.

Daley, James M, Phillip L Palmbos, Dongliang Wu, and Thomas E Wilson. 2005. “Nonhomologous End Joining in Yeast.” *Annu Rev Genet* 39: 431–51. https://doi.org/10.1146/annurev.genet.39.073003.113340.

Evans, B A, and D E Rozen. 2012. “A Streptococcus Pneumoniae Infection Model in Larvae of the Wax Moth *Galleria mellonella*.” *European Journal of Clinical Microbiology & Infectious Diseases : Official Publication of the European Society of Clinical Microbiology* 31 (10): 2653–60. https://doi.org/10.1007/s10096-012-1609-7.

Ferber, Dan. 2004. “Synthetic Biology. Microbes Made to Order.” *Science (New York, N.Y.)* 303 (5655): 158–61. https://doi.org/10.1126/science.303.5655.158.

Garneau, Josiane E., Marie Ève Dupuis, Manuela Villion, Dennis A. Romero, Rodolphe Barrangou, Patrick Boyaval, Christophe Fremaux, Philippe Horvath, Alfonso H. Magadán, and Sylvain Moineau. 2010. “The CRISPR/Cas Bacterial Immune System Cleaves Bacteriophage and Plasmid DNA.” *Nature* 468 (7320): 67–71. https://doi.org/10.1038/nature09523.

Garsin, D A, C D Sifri, E Mylonakis, X Qin, K V Singh, B E Murray, S B Calderwood, and F M Ausubel. 2001. “A Simple Model Host for Identifying Gram-Positive Virulence Factors.” *Proc.Natl.Acad.Sci.U.S.A* 98 (0027–8424): 10892–97. https://doi.org/10.1073/pnas.191378698.

Gaudelli, Nicole M., Alexis C. Komor, Holly A. Rees, Michael S. Packer, Ahmed H. Badran, David I. Bryson, and David R. Liu. 2017. “Programmable Base Editing of A•T to G•C in Genomic DNA without DNA Cleavage.” *Nature* 551 (7681): 464–71. https://doi.org/10.1038/nature24644.

Gerber, J., G. Raivich, A. Wellmer, C. Noeske, T. Kunst, A. Werner, W. Brück, and R. Nau. 2001. “A Mouse Model of *Streptococcus pneumoniae* Meningitis Mimicking Several Features of Human Disease.” *Acta Neuropathologica* 101 (5): 499–508. https://doi.org/10.1007/s004010000326.

H. Meijer, Annemarie, and Herman P. Spaink. 2011. “Host-Pathogen Interactions Made Transparent with the Zebrafish Model.” *Current Drug Targets* 12 (7): 1000–1017. https://doi.org/10.2174/138945011795677809.

Hanna, Ruth E., and John G. Doench. 2020. “Design and Analysis of CRISPR–Cas Experiments.” *Nature Biotechnology*. Nature Research. https://doi.org/10.1038/s41587-020-0490-7.

Hirokawa, Yasutaka, Hiroaki Kawano, Kimie Tanaka-Masuda, Noriko Nakamura, Akira Nakagawa, Mikito Ito, Hideo Mori, Taku Oshima, and Naotake Ogasawara. 2013. “Genetic Manipulations Restored the Growth Fitness of Reduced-Genome *Escherichia coli*.” *Journal of Bioscience and Bioengineering* 116 (1): 52–58. https://doi.org/10.1016/j.jbiosc.2013.01.010.

Hutchison, Clyde A., R.-Y. Ray Yuan Chuang, Vladimir N. Noskov, Nacyra Assad-Garcia, Thomas J. Deerinck, Mark H. Ellisman, John Gill, et al. 2016. “Design and Synthesis of a Minimal Bacterial Genome.” *Science* 351 (6280): aad6253–aad6253. https://doi.org/10.1126/science.aad6253.

Ishino, Y., H. Shinagawa, K. Makino, M. Amemura, and A. Nakatura. 1987. “Nucleotide Sequence of the Iap Gene, Responsible for Alkaline Phosphatase Isoenzyme Conversion in Escherichia Coli, and Identification of the Gene Product.” *Journal of Bacteriology* 169 (12): 5429–33. https://doi.org/10.1128/jb.169.12.5429-5433.1987.

Jansen, W T M, M Bolm, R Balling, G S Chhatwal, and R Schnabel. 2002. “Hydrogen Peroxide-Mediated Killing of Caenorhabditis Elegans by *Streptococcus pyogenes.*” *Infect. Immun.* 70 (9): 5202–7. https://doi.org/10.1128/IAI.70.9.5202.

Jiang, Wenyan, David Bikard, David Cox, Feng Zhang, and Luciano A Marraffini. 2013. “RNA-Guided Editing of Bacterial Genomes Using CRISPR-Cas Systems.” *Nature Biotechnology* 31 (3): 233–39. https://doi.org/10.1038/nbt.2508.

Jim, Kin Ki, Joo Yeon Engelen-Lee, Astrid M. van der Sar, Wilbert Bitter, Matthijs C. Brouwer, Arie van der Ende, Jan Willem Veening, Diederik van de Beek, and Christina M.J.E. Vandenbroucke-Grauls. 2016. “Infection of Zebrafish Embryos with Live Fluorescent *Streptococcus pneumoniae* as a Real-Time Pneumococcal Meningitis Model.” *Journal of Neuroinflammation* 13 (1): 188. https://doi.org/10.1186/s12974-016-0655-y.

Jinek, Martin, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A Doudna, and Emmanuelle Charpentier. 2012. “A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity.” *Science (New York, N.Y.)* 337 (6096): 816–21. https://doi.org/10.1126/science.1225829.

Kadioglu, Aras, Jeffrey N Weiser*, James C Paton, and Peter W Andrew. 2008. “The Role of Streptococcus pneumoniae* Virulence Factors in Host Respiratory Colonization and Disease.” *Nature Reviews. Microbiology* 6 (4): 288–301. https://doi.org/10.1038/nrmicro1871.

Keller, Lance E., Anne Stéphanie Rueff, Jun Kurushima, and Jan Willem Veening. 2019. “Three New Integration Vectors and Fluorescent Proteins for Use in the Opportunistic Human Pathogen *Streptococcus pneumoniae*.” *Genes* 10 (5). https://doi.org/10.3390/genes10050394.

Komor, Alexis C., Yongjoo B. Kim, Michael S. Packer, John A. Zuris, and David R. Liu. 2016.

"Programmable Editing of a Target Base in Genomic DNA without Double-Stranded DNA Cleavage." *Nature* 533 (April): 420–24. https://doi.org/10.1038/nature17946.

Komor, Alexis C., Kevin T. Zhao, Michael S. Packer, Nicole M. Gaudelli, Amanda L. Waterbury, Luke W. Koblan, Y. Bill Kim, Ahmed H. Badran, and David R. Liu. 2017. "Improved Base Excision Repair Inhibition and Bacteriophage Mu Gam Protein Yields C:G-to-T:A Base Editors with Higher Efficiency and Product Purity." *Science Advances* 3 (8): 1–10. https://doi.org/10.1126/sciadv.aao4774.

Kunst, F, N Ogasawara, I Moszer, A M Albertini, G Alloni, V Azevedo, M G Bertero, et al. 1997. "The Complete Genome Sequence of the Gram-Positive Bacterium *Bacillus subtilis." Nature* 390 (6657): 249–56. https://doi.org/10.1038/36786.

Liu, Xue, Clement Gallay, Morten Kjos, Arnau Domenech, Jelle Slager, Sebastiaan P Kessel, Kèvin Knoops, Robin A Sorg, Jing-Ren Zhang, and Jan-Willem Veening. 2017. " High-throughput CRISPRi Phenotyping Identifies New Essential Genes in *Streptococcus pneumoniae* ." *Molecular Systems Biology* 13 (5): 931. https://doi.org/10.15252/msb.20167449.

Liu, Xue, Jacqueline M. Kimmey, Laura Matarazzo, Vincent de Bakker, Laurye Van Maele, Jean Claude Sirard, Victor Nizet, and Jan Willem Veening. 2021. "Exploration of Bacterial Bottlenecks and *Streptococcus pneumoniae* Pathogenesis by CRISPRi-Seq." *Cell Host and Microbe* 29 (1): 107-120.e6. https://doi.org/10.1016/j.chom.2020.10.001.

Massidda, Orietta, Linda Nováková, and Waldemar Vollmer. 2013. "From Models to Pathogens: How Much Have We Learned about *Streptococcus pneumoniae* Cell Division?" *Environmental Microbiology* 15 (12): 3133–57. https://doi.org/10.1111/1462-2920.12189.

Moeller, Ralf, Erko Stackebrandt, Günther Reitz, Thomas Berger, Petra Rettberg, Aidan J Doherty, Gerda Horneck, and Wayne L Nicholson. 2007. "Role of DNA Repair by Nonhomologous-End Joining in *Bacillus subtilis* Spore Resistance to Extreme Dryness, Mono- and Polychromatic UV, and Ionizing Radiation." *Journal of Bacteriology* 189 (8): 3306–11. https://doi.org/10.1128/JB.00018-07.

Mook-Kanamori, Barry, Madelijn Geldhoff, Dirk Troost, Tom van der Poll, and Diederik van de Beek. 2012. "Characterization of a Pneumococcal Meningitis Mouse Model." *BMC Infectious Diseases* 12. https://doi.org/10.1186/1471-2334-12-71.

Nasevicius, A., and S. C. Ekker. 2001. "The Zebrafish as a Novel System for Functional Genomics and Therapeutic Development Applications." *Current Opinion in Molecular Therapeutics* 3 (3): 224–28.

Nishida, Keiji, Takayuki Arazoe, Nozomu Yachie, Satomi Banno, Mika Kakimoto, Mayura Tabata, Masao Mochizuki, et al. 2016. "Targeted Nucleotide Editing Using Hybrid Prokaryotic and Vertebrate Adaptive Immune Systems." *Science* 353 (6305).

https://doi.org/10.1126/science.aaf8729.

Pitcher, Robert S., Andrew J. Green, Anna Brzostek, Malgorzata Korycka-Machala, Jaroslaw Dziadek, and Aidan J. Doherty. 2007. “NHEJ Protects Mycobacteria in Stationary Phase against the Harmful Effects of Desiccation.” *DNA Repair* 6 (9): 1271–76. https://doi.org/10.1016/j.dnarep.2007.02.009.

Pitcher, Robert S, Nigel C Brissett, and Aidan J Doherty. 2007. “Nonhomologous End-Joining in Bacteria: A Microbial Perspective.” *Annual Review of Microbiology* 61: 259–82. https://doi.org/10.1146/annurev.micro.61.080706.093354.

Rengarajan, Jyothi, Barry R. Bloom, and Eric J. Rubin. 2005. “Genome-Wide Requirements for *Mycobacterium tuberculosis* Adaptation and Survival in Macrophages.” *Proceedings of the National Academy of Sciences of the United States of America* 102 (23): 8327–32. https://doi.org/10.1073/pnas.0503272102.

Renshaw, Stephen A., and Nikolaus S. Trede. 2012. “A Model 450 Million Years in the Making: Zebrafish and Vertebrate Immunity.” *DMM Disease Models and Mechanisms*. Dis Model Mech. https://doi.org/10.1242/dmm.007138.

Reuß, Daniel R., Josef Altenbuchner, Ulrike Mäder, Hermann Rath, Till Ischebeck, Praveen Kumar Sappa, Andrea Thürmer, et al. 2017. “Large-Scale Reduction of the *Bacillus subtilis* Genome: Consequences for the Transcriptional Network, Resource Allocation, and Metabolism.” *Genome Research* 27 (2): 289–99. https://doi.org/10.1101/gr.215293.116.

Sung, C K, H Li, J P Claverys, and D A Morrison. 2001. “An RpsL Cassette , Janus , for Gene Replacement through Negative Selection in *Streptococcus pneumoniae* An RpsL Cassette , Janus , for Gene Replacement through Negative Selection in *Streptococcus pneumoniae*.” *Applied and Environmental Microbiology* 67 (11): 5190–96. https://doi.org/10.1128/AEM.67.11.5190.

Synefiaridou, Dimitra, and Jan-Willem Veening. 2021. “ Harnessing CRISPR-Cas9 for Genome Editing in *Streptococcus pneumoniae* D39V .” *Applied and Environmental Microbiology*, January. https://doi.org/10.1128/aem.02762-20.

Weller, Geoffrey R., Boris Kysela, Rajat Roy, Louise M. Tonkin, Elizabeth Scanlan, Marina Della, Susanne Krogh Devine, et al. 2002. “Identification of a DNA Nonhomologous End-Joining Complex in Bacteria.” *Science* 297 (5587): 1686–89. https://doi.org/10.1126/science.1074584.

Acknowledgments

When in 2013, I moved to Groningen for my master studies, as a part of the program, I had to find a lab for my master project. I knew I was interested in Molecular Genetics and The Department of MolGen was my first choice, while a classmate insisted that I get in touch, specifically, with Jan-Willem Veening, the young PI of the department, because he was “the best”. So in the beginning of 2014, I joined the *Pneumos* as a master student and in September of 2015 I returned for my PhD, which I resumed in early 2017 in Lausanne. Above everyone, I would like to thank **Jan-Willem**, who has been a charismatic *group leader.* Throughout the years, I always felt supported and inspired, during the failures, the successes, the writing, or any other challenge I had to face.

Apart from his scientific guidance, though, JW is also the person responsible for assembling this group of great scientists and people, making my PhD experience so much more pleasant. I would like to thank the Pneumos from Groningen. **Katrin,** for being my master supervisor and a good friend. Our Sunday Yoga was one of my favorite rituals. **Jelle,** who is probably the most charismatic human I have ever met, I particularly remember the Pneumo board game weekend he organized in Alkmaar. **Morten**, one of the kindest person and greatest scientists I know. **Rieza** for being an amazing office mate and for all our discussions. And of course **Robin**, for being effortlessly fantastic on everything, and a great friend. These people combining being great at Science and personalities was a big part of my appeal to join the lab for my PhD.

In the same category belong also my collogues with whom we continued this journey from Groningen to Lausanne. I would like to thank **Arnau,** for collaborating for Chapter 4 and for being a valuable member of the lab. Organizing borrels with him as PV, is one of my favorite memories. **Clement,** for being this endless source of energy, always cool and fun. **Renske,** for being effortlessly amazing on everything she is doing. I always found this very inspiring. Also, together with **Doran** for sharing a house for a few months, making our transition in Lausanne so much easier. **Lance,** for being fun and always there for some advice. Being bench mates with him was great, not only because he never minded me taking up more space and stealing his tips. **Stefano,** my other bench mate, again for always trusting that the tips I was stealing from him were clean and for being fun and kind. And, of course, **Xue.** Her intelligence and her humor was what impressed me when we first met, but her kindness is what keeps impresses me until today. Setting up the new lab and our new lives in Lausanne, with these people, has been a life-changing experience for me.

And when I thought that our lab was perfect and we could never recreate this success in Lausanne, new people joined, and I was proven wrong. I would like to thank **Liselot** for

demonstrating that it is possible for a person not to have any flaws, being always intelligent, social and funny. **Monica,** for bringing her Colombian vibe and for always being kind. **Jun,** for inspiring me with his hard work, kindness and generosity. **Vincent,** for always being helpful and genuinely nice. **Afonso,** for being a good friend and for the movie marathons. Also for being only person I (almost) didn't hate running (with). **Julien,** for always smiling and teaching me Vaudois tips, and for doing my homework. **Anne-Stephanie,** for our great conversations at the lab. Watching her being a great scientist and a great mum has been an inspiration for me. **Bevika,** who arrived recently, but I can already tell she is turning into a very good friend. And the new arrivals; **Vik,** for keeping the coolness levels of our lab so high and for always being so kind, and **Axel** who is always eager to help. **Jessica,** I am looking forward for you to join the team again. I would also like to thank the amazing students that joined the lab during the years; **Vish, Alice, Pascal, Nadia, Maria, Aya** and **Nina.** Last but not least, I would like to thank **Paddy**, for being my lab sister and one of my best friends. Her being in my life is contributing to making me the person I am becoming.

Leaving **MolGen** has been very hard for me but finding **DMF** made me also very happy. There I met some of the people that made my work and social life so much nicer, and I deeply appreciate them for this. Among so many I would like to thank especially **Rita, Ania, Sylvia, Hammam, Flo, Roberto, Diogo, Konstantin,** and **Sudip.**

I would also like to thank my **Friends**, that became family. From my bachelor in Thessaloniki, my master in Groningen and the ones I met here in Lausanne.

Finally, I would like to thank my **Family.** My grandmothers, my aunts and uncles, my cousins. Their support and love keep me going. My sisters **Anastasia** and **Konstantina** for being my other two thirds. And my loving **Parents**, Christos and Katerina. For everything. Ευχαριστώ.