

Three new integration vectors and fluorescent proteins for use in the opportunistic human pathogen *Streptococcus pneumoniae*

Lance E. Keller^{a*}, Anne-Stéphanie Rueff^{a*}, Jun Kurushima^a, and Jan-Willem Veening^{a,1}

^a Department of Fundamental Microbiology, Faculty of Biology and Medicine, University of Lausanne, Biophore Building, CH-1015 Lausanne, Switzerland

* These authors contributed equally.

¹ Correspondence: Jan-Willem.Veenning@unil.ch, tel: +41 (0)21 6925625, Twitter: @JWVeening

Abstract

Here we describe the creation of three integration vectors, pPEPX, pPEPY and pPEPZ, for use with the opportunistic human pathogen *Streptococcus pneumoniae*. The constructed vectors, named PEP for Pneumococcal Engineering Platform, employ an IPTG-inducible promoter and BglBrick and BglFusion compatible multiple cloning sites allowing for fast and interchangeable cloning. PEP plasmids replicate in *Escherichia coli* and harbor integration sites that have homology in a large set of pneumococcal strains, including recent clinical isolates. In addition, several options of antibiotic resistance markers are available, even allowing for selection in multidrug resistant clinical isolates. The transformation efficiency of these PEP vectors as well as their ability to be expressed simultaneously were tested. Two of the three PEP vectors share homology of the integration regions with over half of the *S. pneumoniae* genomes examined. Transformation efficiency varied among PEP vectors based on the length of the homology regions, but all were highly transformable and can be integrated simultaneously in strain D39V. Vectors used for pneumococcal cloning are an important tool for researchers for a wide range of uses. The PEP vectors described are of particular use because they have been designed to allow for easy transfer of genes between vectors as well as integrating into transcriptionally silent areas of the chromosome. In addition, we demonstrate the successful production of several new spectrally distinct fluorescent proteins (mTurquoise2, mNeonGreen and mScarlet-I) from the PEP vectors. The PEP vectors and newly described fluorescent proteins will expand the genetic toolbox for pneumococcal researchers and aid future discoveries.

Keywords: *Streptococcus pneumoniae*, pneumococcal engineering platform, synthetic biology, expression vectors, mTurquoise2, mNeonGreen, mScarlet-I

Introduction

Infectious disease is one of the most common causes of death worldwide and is an object of intense scientific study (1). A powerful tool into the research of human pathogens is the ability to manipulate the pathogen in a desired manner to observe the outcome. These manipulations can take many forms; gene deletions, genetic complementation, gene mutations, or controlled expression. To aid in alteration of a target genome, numerous methods have been developed over the years with varying success depending on the targeted pathogen (2). Tool development in model organisms tend to be over represented compared to other species, with bacterial human pathogens having fewer options. While the study of model organisms has led to great advances in understanding the nature of diverse bacterial functions, they are not useful for pathogen specific processes. In an attempt to ameliorate this deficit, we set out to develop several cloning vectors useful for genetic manipulation in *Streptococcus pneumoniae*, even in recent clinically relevant isolates.

The bacterium *S. pneumoniae* is an important opportunistic human pathogen that is responsible for over a million deaths per annum (3). *S. pneumoniae* has been extensively studied in the past decades and several useful tools for genetic manipulation have been created that allow for stable single copy integration at the genome. However, there is still a lack of robust suicide vectors that do not disrupt native expression of important genes that can be combined to insert several genes in the same strain or can be applied to multi-drug resistant clinical isolates. One of

the most commonly used ectopic integration site in *S. pneumoniae* is the *bgaA* locus (4–7), that thereby disrupts this beta-galactosidase, which might have an important function during colonization of human epithelial cells (8–10). The only neutral integration site for which integration plasmids have been developed is the CEP locus, which is a transcriptionally silent site within a truncated IS 1167 element downstream of the well-studied *ami* operon (11, 12). However, this locus is not very well conserved among pneumococcal strains (see below).

Utilizing extensive RNA-seq data for *S. pneumoniae* in multiple conditions (13), transcriptionally silent regions were identified for potential chromosomal locations to use as recombination sites. Using the pneumococcal engineering platform (PEP) vector as a guide (12), novel integration vectors were developed for use at these regions. In this study, we have developed and tested three new PEP vectors, pPEPX, pPEPY, and pPEPZ, for their ability to be transformed into *S. pneumoniae*, compatibility for co-expression, and frequency of occurrence in pneumococcal isolates. In addition, three spectrally distinct and intrinsically bright fluorescent proteins, mTurquoise2, mNeonGreen and mScarlet-I, were added to the fluorescence protein toolbox now available for *S. pneumoniae*.

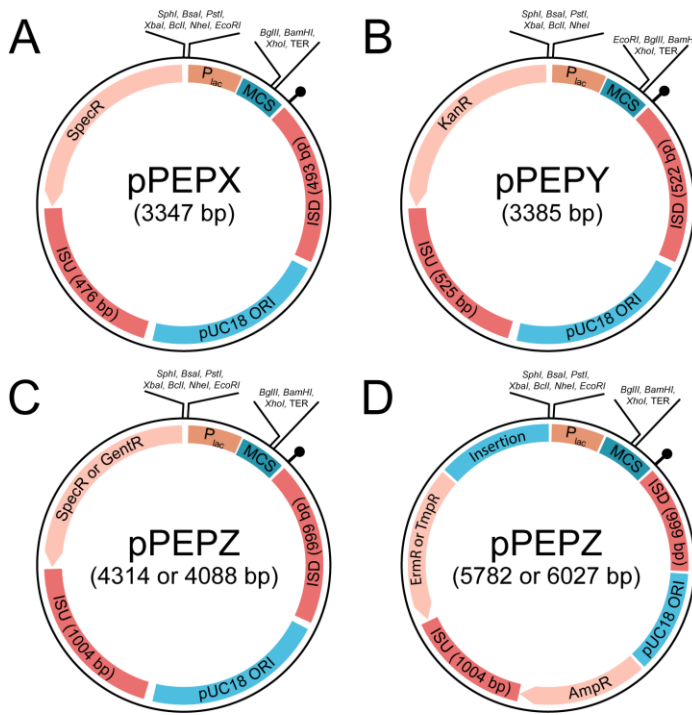


Figure 1: Diagram of general layout for various constructed PEP vectors. All PEP vectors have the same IPTG inducible promoter as well as the same pUC18 origin of replication which is permissive in *E. coli*, but not in *S. pneumoniae*. All vectors have a multiple cloning site featuring the BglBrick cloning method and integration regions that are transcriptionally silent based on RNA-seq data. A) The pPEPX vector has a spectinomycin resistance marker and integration regions of approximately 500 bp in length and integrates at 295.4° around the D39V chromosome near *amiF* (*spv_1667*). B) The pPEPY vector has a kanamycin resistance marker and integration regions of approximately 500 bp in length and integrates at 74.6° around the D39V chromosome near *spv_0422*. C) Two pPEPZ vectors have either a spectinomycin or gentamycin resistance marker and integration regions of approximately 1000 bp in length and integrates at 303.9° around the D39V chromosome near *spv_1736*. D) Two pPEPZ vectors have either an erythromycin or trimethoprim resistance marker along with an ampicillin marker for selection in *E. coli*. A nonfunctional *E. coli* chromosomal sequence insertion between the *S. pneumoniae* selection marker and the promoter is indicated.

Results

Design and construction of pPEPX, pPEPY and pPEPZ

As seen in Figure 1, the three pneumococcal engineering platform plasmids created have compatible cloning sites following the BglBrick and BglFusion principle (12, 27), to facilitate easy exchange between the plasmids. The exception is in the pPEPY vector with the EcoRI site located downstream of the P_{lac} promoter as opposed to upstream as in the other two plasmids. The integration regions vary for each vector, but are in transcriptionally silent areas of the genome as determined by RNA-seq data (13). The pPEPX vector integrates about 295° around the chromosome between the *treR* (SPV_1665) and *amiF* (SPV_1666) genes (Figure 2A, B). The pPEPY vector integrates at 75° around the chromosome between *spv_2165* and *spv_2166*, both encoding hypothetical proteins, (Figure 2A, B). The last constructed vector, pPEPZ, integrates at 304° around the chromosome, upstream of *spv_1736*, which encodes a degenerative hypothetical protein (Figure 2A,

B). Transcriptional terminators flank the MCS in all vectors to ensure that the inserted DNA is not influenced by transcriptional events happening elsewhere on the chromosome. The origin of replication used for all plasmids is a high copy number *E. coli* origin derived from pUC18 that is not recognized by *S. pneumoniae*. All selective markers are expressed in both *E. coli* and *S. pneumoniae*.

Various antibiotic markers were created for pPEPZ to facilitate integration of multiple PEP vectors simultaneously, as well as providing multiple options depending on the researchers need. As pPEPZ has a large pneumococcal host range (see below), we also constructed pPEPZ variants with antibiotic resistance alleles less commonly found in clinical isolates such as trimethoprim and gentamycin (Fig. 2). No PEP vectors contain the tetracycline resistance marker for selection, allowing strains to utilize PEP vectors in combination with the popular pPP1/pJWV/pKB *bgaA*-based integration vectors (5–7). Table 1 summarizes vectors and available resistance markers.

Vector construction was done through various cloning techniques, but common motifs were kept for all PEP vectors. These include the same origin of replication for a high copy number in *E. coli*, similar multiple cloning sites to facilitate rapid exchange between plasmids, an IPTG inducible promoter with tight expression control, and multiple antibiotic selection markers for use in almost any background. Expression without an inducible promoter can be done using pPEPY or standard cloning can be used to exchange the promoter P_{lac} with any desired promoter. A list of available vectors and the selectable markers can be found in Table 1. Note that to make use of the P_{lac} IPTG inducible promoter (Sorg et al. in preparation), the strain also needs to encode the LacI repressor from *E. coli* otherwise the promoter will be constitutively expressed. We noticed significant vector stability when leaving out the gene encoding the *lacI* repressor from the plasmid. Therefore we used a separate, ectopic integration of *lacI* (20). This *lacI* construct can be obtained via Addgene (catalog #85589) and is integrated at the *prsI* locus by selecting for gentamycin (20).

High transformation efficiency of pPEPX, pPEPY and pPEPZ

Transformation efficiency of these three vectors was determined in *S. pneumoniae* strain D39V (Figure 3). Multiple concentrations of DNA were tested and the amounts used were normalized to molar concentrations to remove concentration effects due to size differences in the plasmids. It was found that pPEPZ had higher transformation efficiencies than either pPEPX or pPEPY at all DNA concentrations tested (0.064, 0.32, 1.6, 8.0, 40.0, 200.0, 1000.0 nM), with a maximum transformation efficiency of approximately 10^{-3} (0.1% of transformants per total amount of viable cells) compared to 10^{-4} . This is most likely due to the longer regions of homology present on pPEPZ (1 kb) compared to the homologous integration regions present on pPEPX (500 bps) and pPEPY (500 bps). For all vectors tested there was a two-log increase in transformation efficiency from the lowest concentration (0.064 nM) of DNA added to the highest concentration (1000 nM). A rapid increase in transformation efficiency was observed upon increasing DNA concentrations, but gradually leveled off at a concentration of 40 nM of DNA for all PEP vectors. This data is completely in line with previously described transformation efficiencies in *S. pneumoniae* that demonstrated a similar relation between length of the homology regions and DNA concentration (28, 29). In summary, all described vectors can be efficiently transformed into strain D39V and acquisition of one vector does not exclude integration of other PEP vectors at available integration loci.

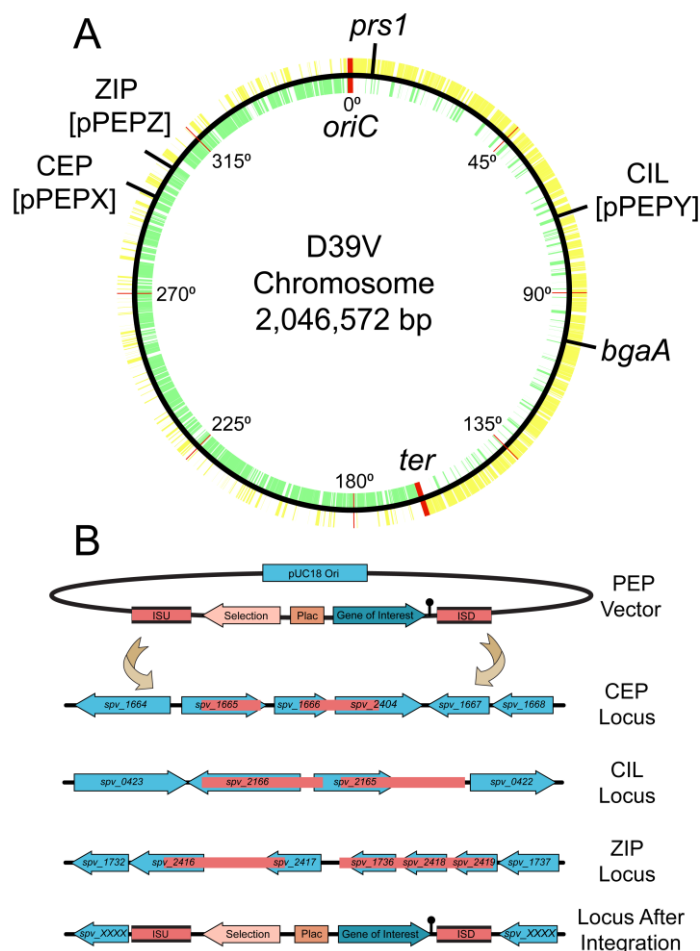


Figure 2: Chromosomal location of integration loci and integration schematic. A) D39V chromosome with all positive strand genes (yellow), negative strand genes (green), and labeled integration loci. Position around the chromosome is demarcated every 45 degrees and origin of replication and terminus are indicated in red. Note that strain D39V has a large inversion around the *ter* region compared to laboratory strain R6 (14). B) Representation of insertion sequence upstream (ISU) and insertion sequence downstream (ISD) in pink for available PEP vectors. Double cross-over event leads to stable integration in CEP (pPEPX), CIL (pPEPY), or ZIP (pPEPZ) locus at indicated chromosomal locations based on most recent D39V genome annotation.

Simultaneous integration and expression of three new fluorescent proteins for *S. pneumoniae* from the pPEPX, pPEPY and pPEPZ loci

To demonstrate the feasibility to clone and construct a strain that expresses heterologous proteins from all three integration platforms simultaneously, we utilized several recent spectrally distinct fluorescent proteins not yet widely used in the pneumococcal research field, namely mTurquoise2 (22), mNeonGreen (23), and mScarlet-I (24). To make sure these proteins do not demonstrate any spectral overlap under our experimental conditions, we fused them to genes encoding proteins with a distinct cellular localization. We cloned the genes encoding *spv_1159-mScarlet-I* in pPEPX, *comEA-mTurquoise2* in pPEPY and *hlpA-mNeonGreen* in pPEPZ and then transformed them into strain D39V in

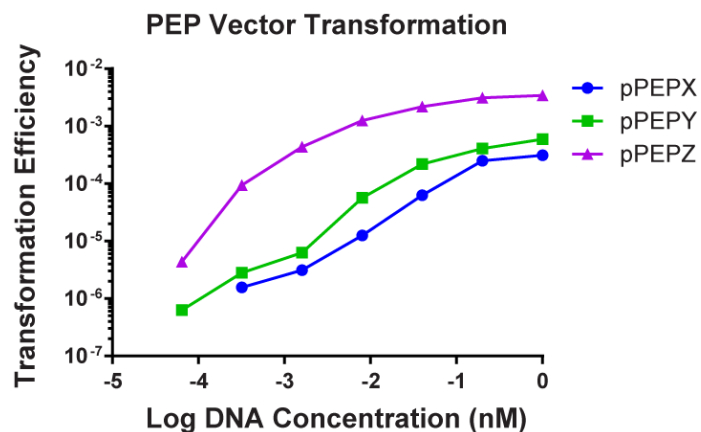


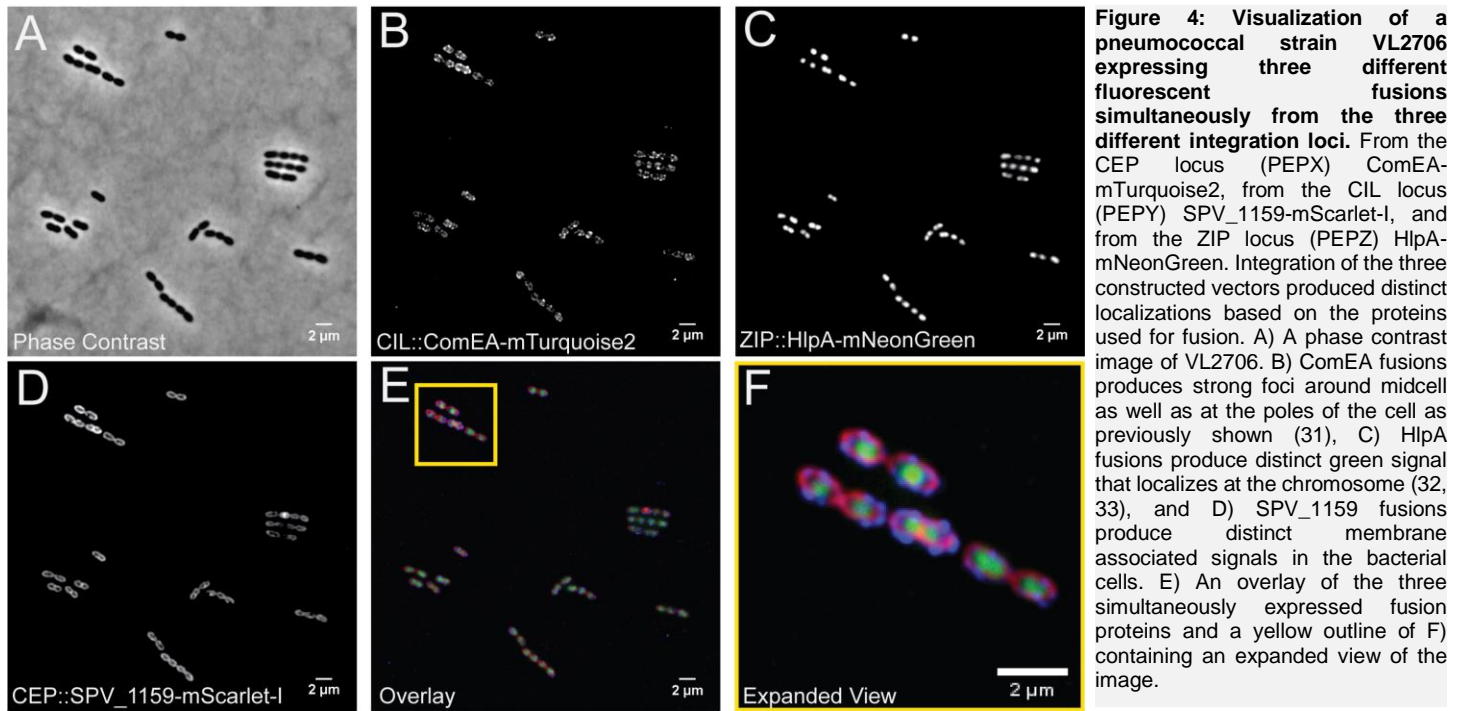
Figure 3: Transformation efficiencies of the various PEP vectors. The three PEP vectors, pPEPX, pPEPY, and pPEPZ were tested for transformation into the *S. pneumoniae* strain D39V. All vectors were transformable into D39V with a dose dependent response based on the amount of DNA added up to a saturation point. At all concentrations of DNA, pPEPZ had the highest transformation efficiency with pPEPX and pPEPY displaying similar rates of transformation and homologous recombination.

serial (see M&M). Note that we directly transformed the ligation mixtures to *S. pneumoniae* without an intermediate *E. coli* clone (30).

Expression of all fusion proteins in constructed strain VL2706 was induced by addition of 1 mM IPTG for 1 hr before visualization by fluorescence microscopy. We selected ComEA as this is a protein involved in DNA uptake and was previously shown to demonstrate a septal localization (31). HlpA (HU) was chosen as it binds to the nucleoid and is a good marker for the chromosome (32, 33). SPV_1159 was selected as it is a predicted small membrane protein with two transmembrane domains with both the N- and C-terminus inside the cytoplasm and was previously shown to demonstrate homogenous membrane localization in our laboratory (Gallay and Veening, unpublished). As shown in Figure 4, cellular morphology of this triple labelled strain is indistinguishable from wild type and we obtained clear, distinguishable fluorescent signals for all three fusion proteins. In addition, all fusions demonstrated their predicted localization. Also, growth curves of parent strain D39V and VL2706 were indistinguishable after 12 hrs of growth (Supp. Figure 1). This demonstrates that stable integration and simultaneous expression of heterologous proteins from pPEPX, pPEPY, and pPEPZ is possible, and provides a new set of fluorescent proteins for use in *S. pneumoniae*.

Broad pneumococcal strain range for pPEPY and pPEPZ

All vectors were designed and based on the genome and transcription data of *S. pneumoniae* strain D39V (13, 14, 34). To determine if the constructed vectors can be used in multiple strains of *S. pneumoniae* as well as closely related species, whole genome sequences were queried for the integrative regions of the PEP vectors. All genomes from NCBI at contig level or above were downloaded for *S. pneumoniae* (2,875 strains), *S. pseudopneumoniae* (38 strains), *S. tigurinus* (19 strains), *S. mutans* (190 strains), *S. equi* (259 strains), *S. pyogenes* (449 strains), and *S. suis* (1,261 strains). Additional genomes for *S. pneumoniae* were obtained from the Maela database (35) (3,085 strains) for a total of 5,960 pneumococcal genomes. Only 5,029 pneumococcal genomes were tested due to duplication of some sequences between the two databases (Figure 5). A positive hit for the strain was determined when both integrative regions with a homology of at least 95% were present and had at least 90% of the total integration region. The pPEPX vector passed these



requirements for 253 (5.03%) of *S. pneumoniae* strains tested. Other streptococcal species tested had no homology to the pPEPX vector that passed our requirements for a positive result. The pPEPY vector had 3,237 (64.37%) positive hits for *S. pneumoniae*, but no positive hits for other species tested. Similar results were found for the pPEPZ vector with 3,229 (64.21%) hits for *S. pneumoniae* and none for other species (Figure 5). Based on our criteria above, all pneumococcal genomes queried can integrate at least one of the PEP vectors available.

Testing for transformation efficiency in numerous strains and species with multiple vectors is tedious and time consuming and testing a small portion of collected strains will not be as beneficial. Although competence for transformation is highly conserved in *S. pneumoniae* and most strains can be transformed in the laboratory, significant differences in transformation efficiencies of different strains are observed (36). Through using sequencing data, it is possible to test thousands of strains for the presence of the integrative regions of the constructed plasmids. While this is not conclusive evidence for the ability of the PEP vectors to be transformable into queried strains, it is indicative that they are transformable. Through using stringent cut off conditions of at least 95%

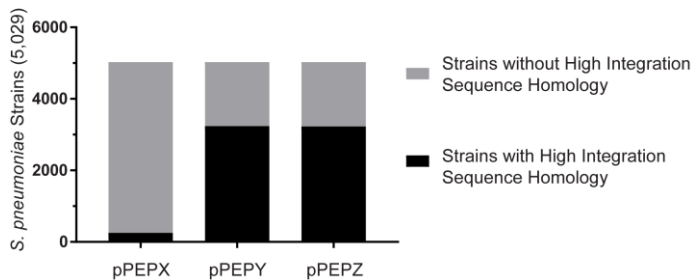


Figure 5: Frequency of *S. pneumoniae* strains with high homology to PEP vector integration regions. Local BLAST database was created containing 5,029 genomes of *S. pneumoniae*. Out of these 5,029 genomes 261, 3,245, and 3,235 strains contained sufficient homology to the pPEPX, pPEPY, and pPEPZ vectors respectively, to potentially allow homologous recombination.

homology to the integration regions and the queried strains having 90% or more of the homology region, the chances of successful genomic integration are high. Supplementary Table 2 list all pneumococcal strains that passed our criteria and indicates possible transformability with the PEP vectors.

Efficient cloning in multidrug resistant clinical isolates using PEP vectors

Antibiotic resistance is on the rise and several clinically relevant strains are now commonly found worldwide, such as PMEN strains (37). To determine the feasibility of using these plasmids in multi-drug resistant clinical isolates we tested the ability of PMEN strains 1, 3, 14, 18, 28 and TIGR4 to integrate PEP plasmids. These are serotype 23F, 9V, 19F, 14, 1, and 4 respectively and have different antibiotic susceptibilities. Table 2 summarizes the strains, susceptibility, transformability based on sequence and experimentally. All strains passed our transformability requirements based on genome sequence for at least one PEP vector. Supplementary figures 2, 3, and 4 demonstrate the homology of the integration regions of the PEP plasmids for several clinically relevant and antibiotic resistant strains. Experimentally determined transformability with the PEP vectors verified the feasibility of using a strains' genome sequence to indicate the ability to integrate a PEP vector. Even when we did not find a high level of homology, we could still transform strains PMEN3 and PMEN14 with plasmids pPEPX and pPEPY. The exception was for the pPEPY vector, which did not yield any transformants with PMEN18 or TIGR4 despite have high sequence homology to the integration region.

Discussion

Here we describe the creation and testing of three integration vectors for use in *Streptococcus pneumoniae*. These vectors can be used for a wide range of applications in the pneumococcal field; from complementation of gene deletions, labeling and expression of genes of interest, and/or expressing non-native proteins. The ability to efficiently transform and express any gene of interest stably from the chromosome as a single copy

from a neutral locus is an essential tool for pneumococcal research. The PEP vectors described here can potentially be used in a large portion of pneumococcal strains and be of benefit to many researchers in the pneumococcal field. All described vectors are made available through Addgene (catalog numbers pending). It is our hope that making these vectors available will facilitate the research of other groups expanding our knowledge base of *S. pneumoniae*.

Materials and Methods

Growth Conditions

All pneumococcal strains are derived from the parent strain D39V (14) and were grown in the semi-synthetic media C + Y formulation 2018 (15) under selection conditions where applicable, see Table 1. Standard pneumococcal transformation (16) protocols using synthetic CSP were used for strain construction with selection on Columbia agar supplemented with 2% sheep blood and appropriate antibiotics. Plasmids were cloned and maintained in *Escherichia coli* strains XL1-Blue or Stbl3 with LB media and selected on LB agar plates with appropriate antibiotics (Table 1). Transformation into *E. coli* was done using rubidium chloride chemically competent cells with standard procedures. Induction of all strains containing an IPTG-inducible promoter were activated with 1 mM of IPTG for 1 hour before visualization by microscopy.

Plasmid Construction

RNA-seq data available at <https://veeninglab.com/pneumoexpress> (13) and gene annotation data at <https://veeninglab.com/pneumobrowse> (14) was used for selection of transcriptionally silent regions.

pPEPX- This vector is a modified smaller version of pPEP1 (12) lacking the chloramphenicol resistance marker and containing the high copy number origin of replication from pUC18, but still integrates at the neutral CEP locus near *amiF* (11). The pUC18 origin of replication was amplified from pLA01 (17) using primers pUC18-F-KpnI and pUC18-R-NotI. The amplified fragment containing pUC18 origin and pPEP1 were digested with *KpnI* and *NotI* (NEB) for 1 hr at 37°C followed by overnight RT ligation with T4 DNA ligase (ThermoScientific). The resulting plasmid was then digested with *KpnI* and *NcoI* (NEB) for 1 hr at 37°C and ligated with hybridized primers *NcoI*-*KpnI*-link-F and *NcoI*-*KpnI*-link-R that were similarly digested. The IPTG inducible promoter was added to the promoterless pPEPX plasmid by digestion of pPEP8PL8-2 (18) with *EcoRI* and *BglIII* and ligating appropriate fragment to complementary digested pPEPX.

pPEPY- Creation of plasmid was completed through Gibson Assembly (NEB) using a minimum of 20 bp overlapping regions of five DNA fragments following manufacturers protocol. Integration regions were amplified from D39V genomic DNA using primer pairs LK042/LK043 and LK046/LK047. Origin of replication and multiple cloning site was amplified from pPEPX vector using primer pairs LK044/LK045 and LK040/LK041 respectively. Kanamycin resistance cassette was amplified from plasmid pDG783 (19) with primer pair LK033/LK034 that included addition of the P_{F6} promoter (12) for kanamycin expression. Three terminators were added to this vector through Gibson assembly of terminators from plasmid

Plasmid Name	Features	Resistance
pPEPX	Promoterless integrative plasmid with terminator	Spec
pPEPX-P _{lac}	Integrative plasmid with IPTG inducible promoter	Spec
pPEPY	Promoterless integrative plasmid with terminator	Kan
pPEPY-P _{lac}	Integrative plasmid with IPTG inducible promoter	Kan
pASR100 (pPEPZ-P _{lac})	Integrative plasmid with IPTG inducible promoter	Spec
pASR102 (pPEPZ-P _{lac})	Integrative plasmid with IPTG inducible promoter	Gent
pASR103 (pPEPZ-P _{lac})	Integrative plasmid with IPTG inducible promoter (Amp selection for <i>E. coli</i>)	Erm/Amp
pASR130 (pPEPZ-P _{lac})	Integrative plasmid with IPTG inducible promoter (Amp selection for <i>E. coli</i>)	Tmp/Amp
pPEPY- P _{lac} -Link- mTurquoise2	Integrative plasmid with IPTG inducible promoter and linker for C-terminal fusion of mTurquoise2	Kan
pPEPY- P _{lac} -Link- mNeonGreen	Integrative plasmid with IPTG inducible promoter and linker for C-terminal fusion of mNeonGreen	Kan
pPEPY- P _{lac} -Link- mScarlet-I	Integrative plasmid with IPTG inducible promoter and linker for C-terminal fusion of mScarlet-I	Kan

Table 1: Spec-spectinomycin, Kan-Kanamycin, Gent-Gentamycin, Erm-Erythromycin, Amp-Ampicillin. Tmp-Trimethoprim

pJWV100 (7) (OVL68/69) and the pre-pPEPY plasmid (OVL66/67), creating pPEPY. The IPTG inducible promoter (P_{lac}) (20) was added through restriction digestion of P_{lac} amplified from FtsK-mNeonGreen (van Raaphorst and Veening, unpublished) with primer pair OVL64/65, creating pPEPY-P_{lac}. Plasmid and DNA was digested with *NheI* and *EcoRI* followed by ligation and transformation into *E. coli* strain XL1-Blue. The pPEPY plasmid integrates at chromosomal integration locus (CIL) thereby interrupting the non-essential pseudogene *spv_2165*.

pPEPZ- Creation of an integration vector of *S. pneumoniae* at the locus *spv_2417* (aka *spd_1735*) was completed through Golden Gate cloning. Primers were designed with Benchling to amplify four different DNA fragments. Primer pairs OVL681/682 and OVL683/684 were used to amplify the downstream and upstream homologous region, respectively, for integration into *S. pneumoniae* by using chromosomal DNA of D39V as

Strain	Serotype	Susceptibility	PEP Transformable based on Sequence	Transformation Efficiency (pPEPX)	Transformation Efficiency (pPEPY)	Transformation Efficiency (pPEPZ*)
PMEN1	23F	Erm, Spec, Kan	pPEPY and pPEPZ	2.56×10 ⁻⁸	2.61×10 ⁻⁷	8.17×10 ⁻⁶
PMEN3	9V	Erm, Spec, Kan	pPEPZ	2.49×10 ⁻⁸	3.31×10 ⁻⁷	2.26×10 ⁻⁸
PMEN14	19F	Gent, Spec, Kan	pPEPZ	8.10×10 ⁻⁷	8.52×10 ⁻⁷	1.43×10 ⁻⁶
PMEN18	14	Gent, Spec, Kan	pPEPY and pPEPZ	3.77×10 ⁻⁷	0	2.16×10 ⁻⁶
PMEN28	1	Erm, Gent, Spec, Kan	N/D	8.79×10 ⁻⁸	7.95×10 ⁻⁸	1.03×10 ⁻⁶
TIGR4	4	Erm, Gent, Spec, Kan, Tmp	pPEPX and pPEPZ	4.98×10 ⁻⁶	0	2.42×10 ⁻⁵

Table 2: Spec-spectinomycin, Kan-Kanamycin, Gent-Gentamycin, Erm-Erythromycin, Tmp-Trimethoprim. *-pPEPZ plasmid chosen based on susceptibility of strain

template. Primers OVL685/686 were used to amplify the pUC18 origin of replication of from pPEPX vector and primers OVL687/688 were used to amplify a fragment containing the promoter P_{lac} , the multiple cloning site, and the spectinomycin marker from the pPEP1- P_{lac} vector (12). *DpnI* treatment was performed for 15 min at 37°C and the four different amplified parts were purified and then digested with *BsmBI* (NEB) enzyme at 55°C for 1.5 hrs. This reaction was purified and ligated at RT for 1.5 hrs then transformed into *E. coli* XL1-Blue. Digestion control followed by sequencing was performed to confirm the construct. The pPEPZ integrates at the PEPZ Integration Position (ZIP) locus upstream of *spv_1736* thereby interrupting the non-essential pseudogene *spv_2417*.

To generate the pPEPZ variants with gentamicin (*gent*) resistance, a Golden Gate strategy was employed. Primer pair OVL1066/1067 was used to amplify pPEPZ- P_{lac} and primers pair OVL1068/1069 was used to amplify the *gent* gene from D39V containing the *lacI* gene at the *prs1* locus as template (20). The two PCR products were purified, digested with *BbsI*-HF (NEB) at 37°C for 1 hr, and ligated at RT for 1 hr. Ligation mixture was transformed in *E. coli* strain Sbt13, sequencing was performed to confirm the construct called pASR102.

pPEPZ containing *bla*- First, beta-lactamase gene (*bla*) was cloned into the pPEPZ- P_{lac} (*spec*) vector in addition to the spectinomycin (*spec*) marker. Primer pairs OVL1190/1191 was used to amplified the *bla* gene from pASR113 as template (a pLA18 derivative, Rueff and Veening unpublished, (17)). The purified PCR fragment and pPEPZ- P_{lac} vector, were digested for 1.5 hrs at 37°C with *NotI* (NEB) to introduce the *bla* gene at the unique *NotI* site of pPEPZ- P_{lac} (*spec*), between the origin of replication of pUC18 and the upstream homologous region. Fragments were ligated at RT for 2 hrs and transformed into *E. coli* strain Sbt13 to generate intermediate plasmid pASR101. Second, this plasmid was used for resistance marker exchange of spectinomycin marker with an erythromycin (*erm*) marker by Golden Gate cloning. For that, pASR101 was amplified by PCR with OVL1066 and OVL1067, and the *erm* gene was amplified with OVL1070 and OVL1071 from strain XL39 (Liu and Veening unpublished) as template. The two PCR products were treated with *DpnI* (NEB) for 30 min at 37°C, followed by purification and then digested with *BbsI*-HF (NEB) enzyme for 1 h at 37°C. Reaction was purified and ligated for 30 min at room temperature followed by transformation in *E. coli* strain Sbt13. Sequencing was performed to confirm the construct called pASR103. Sequencing revealed an insertion between the *erm* marker and the P_{lac} region. This insertion of 777 bp is from *E. coli* chromosomal DNA and does not affect the functioning of the plasmid, as the P_{lac} promoter is still well insulated by the transcriptional terminators, but does increase plasmid size.

Another version of the pPEPZ plasmid has been generated by replacing the *erm* marker with a trimethoprim (*tmp*) antibiotic marker. The plasmid was obtained by Golden Gate cloning as follows: primers OVL1696/1697 were used to amplify pASR103 and then *DpnI* (NEB) treated for 1 hr at 37°C. Primers OVL1698/1699 were used to amplify the *tmp* gene from pneumococcal strain AM39 (21) as template. Thereafter, the two PCR products were purified and then digested with *AarI* (NEB) enzyme at 37°C for 5 hrs. After purification, the two parts were ligated at RT for 2 hrs. The ligation mixture was used to transform *E. coli* strain Sbt13, and subsequent sequencing was performed to confirm the construct called pASR130. This plasmid also contains the *E. coli* chromosomal DNA insertion of 777 bp in between *tmp* and P_{lac} promoter region, since it is based on the pASR103 plasmid.

All plasmids were verified by sequencing and deposited at Addgene (pending accession number).

Three additional pPEPY- P_{lac} plasmids were created for carboxy(C)-terminal fusions of three spectrally distinct fluorescent proteins, mTurquoise2 (22), mNeonGreen (23), and mScarlet-I (24). The plasmid pPEPY- P_{lac} -link-mNeonGreen was created through amplification of mNeonGreen containing an 11 amino acid C-terminal linker with primers link-mNG-F and link-mNG-R from plasmid pRRneon, which contains a *S. pneumoniae* codon optimized version of *mNeonGreen* (Raaphorst and Veening, unpublished). The resulting fragment was digested with *BamHI* and *XhoI* (NEB) for 1 hr at 37°C followed by overnight RT ligation with similarly digested pPEPY- P_{lac} plasmid and T4 DNA ligase (ThermoScientific). The resulting plasmid was used for the template for pPEPY- P_{lac} -link-mTurquoise2 using primers OVL914/915. *mTurquoise2* was amplified with OVL916/917 using Addgene plasmid #54844 (mTurquoise2-pBAD) as template. Subsequent DNA fragments were digested in a single reaction with *BsmBI* (NEB) for 1 hr at 55°C followed by overnight ligation at RT with T4 DNA ligase (ThermoScientific). The *mScarlet* gene was obtained through *S. pneumoniae* codon optimized gene synthesis by IDT followed by cloning into pPEPY- P_{lac} . The *mScarlet-I* variant was made through rPCR of pPEPY- P_{lac} -mScarlet with primers OVL419/420 introducing a T73I point mutation to the protein. Subsequent amplification was digested with *DpnI* (NEB) for 1 hr at 37°C followed by treatment with ClonExpress (Vazyme) following manufacturers protocol.

The *mScarlet-I* variant was amplified from the resulting plasmid with OVL908/909 and combined with linearized pPEPY- P_{lac} -Link (OVL906/907) and treated as for pPEPY- P_{lac} -link-mTurquoise2 creation. All ligation products were subsequently cloned into XL1-Blue and verified by sequencing. Addgene is pending accession number.

Gene fusions for ectopic integration at various PEP loci were created as follows. The fusion containing ComEA and mTurquoise2 was created through *comEA* amplification from D39V gDNA with primers OVL576/615. Amplification was digested with *EcoRI* and *BamHI* (NEB) for 1 hr at 37°C and ligated with similarly digested pPEPY- P_{lac} -Link-mTurquoise2 followed by overnight ligation at RT with T4 DNA ligase (ThermoScientific). The *HlpA* and mNeonGreen fusion in pASR103, pPEPZ plasmid with erythromycin selection in *S. pneumoniae*, was created by *hlpA-mNeonGreen* amplification from strain VL1978 (Veening, unpublished) with primers OVL1569/1570. The resulting fragment and pASR103 were digested with *BglII* and *XhoI* (NEB) for 1 hr at 37°C followed by overnight ligation at RT with T4 DNA ligase. Creation of CEP locus integration of *spv_1159-mScarlet-I* was created through amplification of VL1785 gDNA, which contains CEP integration with a *spv_1159-sfGFP* fusion, (Kurushima and Veening unpublished) with primer pairs OVL037/869 and OVL040/870. The fragment containing *mScarlet-I* was amplified from plasmid pPEPY- P_{lac} -Link-mScarlet-I with primer pair OVL871/872. The resulting three fragments were digested together with *BsmBI* (NEB) for 1 hr at 55°C followed by overnight ligation at RT with T4 DNA ligase. All oligos are listed in Supplementary Table 1.

Strain Construction

Parental strain D39V was transformed with a constitutively expressed *Lacl* under gentamycin selection, downstream of the *prs1* locus (20), leading to strain VL236. This strain was used for all future transformations, with the exception of plasmid pASR102, which was transformed into D39V, due to gentamycin selection of the plasmid. Standard pneumococcal transformation procedure was used to integrate all plasmids independently with appropriate selection (16). For the strain expressing all plasmids (strain VL2706), each ligation product was sequentially cloned with verification of appropriate integration between each transformation.

Database BLAST

Local BLAST was done using NCBI blast-2.7.1+ against whole genomes with a minimum assembly at the contig level. All available *S. pneumoniae* sequences were downloaded from the NCBI database and combined with the Maela database (5,029 strains). Genomes of closely related species *S. pseudopneumoniae* (38 strains) and *S. tigurinus* (19 strains) as well as other streptococcal species *S. mutans* (190 strains), *S. equi* (259 strains), *S. pyogenes* (449 strains), and *S. suis* (1,261 strains) were downloaded from the NCBI database. Integration regions were used as the query sequences against BLAST databases. Results were filtered to include matches with over 95% sequence homology as well as containing at least 90% of the queried DNA. Duplicate matches were removed and results of both integration regions per plasmid were combined. A positive match was only identified if the combined filtered results included two hits per strain, one for upstream integration region and one for downstream integration region. Genetic circular map and comparative alignments were generated using GenomeMatcher and arcWithColor software with modification in graphic representation (25).

Transformation Assay

Wild type *S. pneumoniae* strain D39V was grown to OD₆₀₀ 0.1, washed and suspended in pre-warmed non-acidic C+Y media. Competence was activated with 50 ng/ml of CSP-1 at 37°C for 10 min and 45 μ l was added to a 96 well plate. Five microliters of DNA was added at various concentrations as indicated and incubated for 120 min at 37°C. Samples were serially diluted and plated on non-selective BA and BA with the appropriate antibiotics. Transformation efficiency was calculated by taking the difference between the CFU of the non-selective and the selective BA. For transformation of strains other than D39V, the above protocol was followed with the addition of 50 ng/ml of CSP-2 for initial competence activation due to variations in ComD specificity.

Microscopy

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 430/24 nm (Chroma ET430/24x), 470/40 nm (Chroma ET470/40x), and 545/25 nm (Chroma ET545/25x) for visualization of mTurquoise2, mNeonGreen, or mScarlet-I respectively. For mTurquoise2, light passed through a cube (Leica 11536022) containing a polychroic mirror 455/520/595 nm while

mNeonGreen and mScarlet-I used a cube (Leica 11536022) with a GFP/RFP polychroic mirror (498/564 nm). External emission filters used were from Chroma and ET470/24m, ET520/40m and ET605/70m were used for mTurquoise2, mNeonGreen and mScarlet-I, respectively. Exposure times of 700 ms, 200 ms, and 500 ms with 100% of light from SOLA Light Engine were performed for mTurquoise2, mNeonGreen, and mScarlet-I fusions respectively. Note that we also routinely image mNeonGreen through YFP filters (Chroma ET500/20x for excitation, Chroma ET535/30m for emission with a CFP/YFP/mCherry (455/520/595) polychroic mirror) in case there is signal bleed-through from the mTurquoise2 label in the GFP channel. Images were captured using LasX software (Leica) and exported to ImageJ (26) for final preparation.

Acknowledgments

Work in the Veening lab is supported by Defense Advanced Research Projects Agency (DARPA) cooperative agreement (W911NF-17-2-0132), by the Swiss National Science Foundation (SNSF) (project grant 31003A_172861), a JPIAMR grant (40AR40_185533) from SNSF and ERC consolidator grant 771534-PneumoCaTChER. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

References

1. Bryce J, Boschi-Pinto C, Shibuya K, Black RE. 2005. WHO estimates of the causes of death in children. *Lancet* 365:1147–1152.
2. Khalil AS, Collins JJ. 2010. Synthetic biology: applications and age. *Nat Rev Genet* 11:367–379.
3. Troeger C, Blacker B, Khalil IA, Rao PC, Cao J, Zimsen SRM, Albertson SB, Deshpande A, Farag T, Abebe Z, Adetifa IMO, Adhikari TB, Akibu M, Al Lami FH, Al-Eyadhy A, Alvis-Guzman N, Amare AT, Amoako YA, Antonio CAT, Aremu O, Asfaw ET, Asgedom SW, Atey TM, Attia EF, Avokpaho EFGA, Ayele HT, Ayuk TB, Balakrishnan K, Barac A, Bassat Q, Behzadifar M, Behzadifar M, Bhaumik S, Bhutta ZA, Bijani A, Brauer M, Brown A, Camargos PAM, Castañeda-Orjuela CA, Colombara D, Conti S, Dadi AF, Dandona L, Dandona R, Do HP, Dublinjan E, Edessa D, Elkout H, Endries AY, Fijabi DO, Foreman KJ, Forouzanfar MH, Fullman N, Garcia-Basteiro AL, Gessner BD, Gething PW, Gupta R, Gupta T, Hailu GB, Hassen HY, Hedayati MT, Heidari M, Hibstu DT, Horita N, Ilesanmi OS, Jakovljevic MB, Jamal AA, Kahsay A, Kasaeian A, Kassa DH, Khader YS, Khan EA, Khan MN, Khang Y-H, Kim YJ, Kissoon N, Knibbs LD, Kochhar S, Koul PA, Kumar GA, Lodha R, Magdy Abd El Razek H, Malta DC, Mathew JL, Mengistu DT, Mezgebe HB, Mohammad KA, Mohammed MA, Momeniha F, Murthy S, Nguyen CT, Nielsen KR, Ningrum DNA, Nirayo YL, Oren E, Ortiz JR, PA M, Postma MJ, Qorbani M, Quansah R, Rai RK, Rana SM, Ranabhat CL, Ray SE, Rezai MS, Ruhago GM, Safiri S, Salomon JA, Sartorius B, Savic M, Sawhney M, She J, Sheikh A, Shiferaw MS, Shigematsu M, Singh JA, Somayaji R, Stanaway JD, Sufiyan MB, Taffere GR, Temsah M-H, Thompson MJ, Tobe-Gai R, Topor-Madry R, Tran BX, Tran TT, Tuem KB, Ukwaja KN, Vollset SE, Walson JL, Weldegebreal F, Werdecker A, West TE, Yonemoto N, Zaki MES, Zhou L, Zodpey S, Vos T, Naghavi M, Lim SS, Mokdad AH, Murray CJL, Hay SI, Reiner RC. 2018. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Infect Dis* 18:1191–1210.
4. Robertson GT, Ng W-L, Foley J, Gilmour R, Winkler ME. 2002. Global transcriptional analysis of clpP mutations of type 2 *Streptococcus pneumoniae* and their effects on physiology and virulence. *J Bacteriol* 184:3508–3520.
5. Halfmann A, Hakenbeck R, Brückner R. 2007. A new integrative reporter plasmid for *Streptococcus pneumoniae*. *FEMS Microbiol Lett* 268:217–224.
6. Eberhardt A, Wu LJ, Errington J, Vollmer W, Veening J-W. 2009. Cellular localization of choline-utilization proteins in *Streptococcus pneumoniae* using novel fluorescent reporter systems. *Mol Microbiol* 74:395–408.
7. Overkamp W, Beilharz K, Detert Oude Weme R, Solopova A, Karsens H, Kovács ÁT, Kok J, Kuipers OP, Veening J-W. 2013. Benchmarking Various Green Fluorescent Protein Variants in *Bacillus subtilis*, *Streptococcus pneumoniae*, and *Lactococcus lactis* for Live Cell Imaging. *Appl Environ Microbiol* 79:6481–6490.
8. Wong A, Grau MA, Singh AK, Woodiga SA, King SJ. 2018. The role of neuraminidase-producing bacteria in exposing cryptic carbohydrate receptors for *Streptococcus gordonii* adherence. *Infect Immun* IAI-00068.
9. Dalia AB, Standish AJ, Weiser JN. 2010. Three surface exoglycosidases from *Streptococcus pneumoniae*, NanA, BgaA, and StrH, promote resistance to opsonophagocytic killing by human neutrophils. *Infect Immun* 78:2108–2116.
10. King SJ, Hippe KR, Weiser JN. 2006. Deglycosylation of human glycoconjugates by the sequential activities of exoglycosidases expressed by *Streptococcus pneumoniae*. *Mol Microbiol* 59:961–974.
11. Guiral S, Henard V, Laaberki M-H, Granadel C, Prudhomme M, Martin B, Claverys J-P. 2006. Construction and evaluation of a chromosomal expression platform (CEP) for ectopic, maltose-driven gene expression in *Streptococcus pneumoniae*. *Microbiology* 152:343–349.
12. Sorg RA, Kuipers OP, Veening J-W. 2014. Gene Expression Platform for Synthetic Biology in the Human Pathogen *Streptococcus pneumoniae*. *ACS Synth Biol*.
13. Aprianto R, Slager J, Holsappel S, Veening J-W. 2018. High-resolution analysis of the pneumococcal transcriptome under a wide range of infection-relevant conditions. *Nucleic Acids Res* 46:9990–10006.
14. Slager J, Aprianto R, Veening J-W. 2018. Deep genome annotation of the opportunistic human pathogen *Streptococcus pneumoniae* D39. *Nucleic Acids Res* 46:9971–9989.
15. Domenech A, Slager J, Veening J-W. 2018. Antibiotic-induced cell chaining triggers pneumococcal competence by reshaping quorum sensing to autocrine-like signaling. *Cell Rep* 25:2390–2400.
16. Martin B, Garcia P, Castanié M, Claverys J. 1995. The *recA* gene of *Streptococcus pneumoniae* is part of a competence-induced operon and controls lysogenic induction. *Mol Microbiol* 15:367–379.
17. Slager J, Kjos M, Attaiech L, Veening J-W. 2014. Antibiotic-Induced Replication Stress Triggers Bacterial Competence by Increasing Gene Dosage near the Origin. *Cell* 157:395–406.
18. Sorg RA. 2016. Engineering approaches to investigate pneumococcal gene expression regulation and antibiotic resistance development. *Rijksuniversiteit Groningen*.
19. Fischer KE, Bremer E. 2012. Activity of the osmotically regulated σ^{B} promoter from $Bacillus subtilis$ is controlled at a distance. *J Bacteriol*.
20. Liu X, Gallay C, Kjos M, Domenech A, Slager J, van Kessel SP, Knoops K, Sorg RA, Zhang J, Veening J. 2017. High-throughput CRISPRi phenotyping identifies new essential genes in $Streptococcus pneumoniae$. *Mol Syst Biol* 13.
21. Minnen A, Attaiech L, Thon M, Gruber S, Veening J. 2011. SMC is recruited to *oriC* by ParB and promotes chromosome segregation in *Streptococcus pneumoniae*. *Mol Microbiol* 81:676–688.
22. Goedhart J, von Stetten D, Noirclerc-Savoye M, Lemmoussin M, Joosen L, Hink MA, van Weeren L, Gadella Jr TWJ, Royant A. 2012. Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nat Commun* 3:751.
23. Shaner NC, Lambert GG, Chammass A, Ni Y, Cranfill PJ, Baird MA, Sell BR, Allen JR, Day RN, Israelsson M. 2013. A bright monomeric green fluorescent protein derived from *Branchiostoma lanceolatum*. *Nat Methods* 10:407–409.

24. Bindels DS, Haarbosch L, van Weeren L, Postma M, Wiese KE, Mastop M, Aumonier S, Gotthard G, Royant A, Hink MA, Gadella Jr TWJ. 2016. mScarlet: a bright monomeric red fluorescent protein for cellular imaging. *Nat Methods* 14:53.
25. Ohtsubo Y, Ikeda-Ohtsubo W, Nagata Y, Tsuda M. 2008. GenomeMatcher: a graphical user interface for DNA sequence comparison. *BMC Bioinformatics* 9:376.
26. Rasband WS. 2014. ImageJ. US National Institutes of Health; Bethesda, Maryland, USA: 1997–2012. There is no Corresp Rec this Ref.
27. Anderson Jc, Dueber JE, Leguia M, Wu GC, Goler JA, Arkin AP, Keasling JD. 2010. BglBricks: A flexible standard for biological part assembly. *J Biol Eng* 4:1.
28. Lau PCY, Sung CK, Lee JH, Morrison DA, Cvitkovitch DG. 2002. PCR ligation mutagenesis in transformable streptococci: application and efficiency. *J Microbiol Methods* 49:193–205.
29. Lee MS, Seok C, Morrison DA. 1998. Insertion-Duplication Mutagenesis in *Streptococcus pneumoniae*: Targeting Fragment Length Is a Critical Parameter in Use as a Random Insertion Tool. *Appl Environ Microbiol* 64:4796 LP-4802.
30. Beilharz K, van Raaphorst R, Kjos M, Veening J-W. 2015. Red fluorescent proteins for gene expression and protein localization studies in *Streptococcus pneumoniae* and efficient transformation with Gibson assembled DNA. *Appl Environ Microbiol* .
31. Bergé MJ, Kamgoué A, Martin B, Polard P, Campo N, Claverys J-P. 2013. Midcell recruitment of the DNA uptake and virulence nuclease, EndA, for pneumococcal transformation. *PLoS Pathog* 9:e1003596.
32. Kjos M, Aprianto R, Fernandes VE, Andrew PW, van Strijp JAG, Nijland R, Veening J-W. 2015. Bright fluorescent *Streptococcus pneumoniae* for live-cell imaging of host-pathogen interactions. *J Bacteriol* 197:807–818.
33. Kjos M, Veening J. 2014. Tracking of chromosome dynamics in live *S. pneumoniae* reveals that transcription promotes chromosome segregation. *Mol Microbiol* 91:1088–1105.
34. Aprianto R, Slager J, Holsappel S, Veening J-W. 2016. Time-resolved dual RNA-seq reveals extensive rewiring of lung epithelial and pneumococcal transcriptomes during early infection. *Genome Biol* 17:198.
35. Chewapreecha C, Harris SR, Croucher NJ, Turner C, Marttinen P, Cheng L, Pessia A, Aanensen DM, Mather AE, Page AJ. 2014. Dense genomic sampling identifies highways of pneumococcal recombination. *Nat Genet* 46:305–309.
36. Evans BA, Rozen DE. 2013. Significant variation in transformation frequency in *Streptococcus pneumoniae*. *ISME J* 7:791–799.
37. McGee L, McDougal L, Zhou J, Spratt BG, Tenover FC, George R, Hakenbeck R, Hryniewicz W, Lefevre JC, Tomasz A. 2001. Nomenclature of major antimicrobial-resistant clones of *Streptococcus pneumoniae* defined by the pneumococcal molecular epidemiology network. *J Clin Microbiol* 39:2565–2571.