Unbiased homeologous recombination during pneumococcal transformation allows for multiple chromosomal integration events

Jun Kurushima[1], Nathalie Campo[2], Renske van Raaphorst[1], Guillaume Cerckel[1], Patrice Polard[2], Jan-Willem Veening[1*]

[1]Department of Fundamental Microbiology, Faculty of Biology and Medicine, University of Lausanne, Biophore Building, CH-1015 Lausanne, Switzerland
[2]Laboratoire de Microbiologie et Génétique Moléculaires (LMGM), Centre de Biologie Intégrative (CBI), Toulouse, France

[*]Correspondence to Jan-Willem Veening: Jan-Willem.Veening@unil.ch, tel: +41 (0)21 6925625, Twitter: @JWVeening

Abstract

The rapid spread of antimicrobial resistance and vaccine escape in the opportunistic human pathogen *Streptococcus pneumoniae* can be largely attributed to competence-induced transformation. To better understand why competence-induced transformation is so effective, we studied the dynamics of this process at the single-cell level. We show that within isogenic populations, all cells become naturally competent and bind exogenous DNA. In addition, we find that transformation is highly efficient and that the chromosomal location of the integration site or whether the transformed gene is encoded on the leading or lagging strand has limited influence on recombination efficiency. Indeed, we have observed multiple recombination events in single recipients in real-time. However, because of saturation of the DNA uptake and integration machinery and because a single stranded donor DNA replaces the original allele, we find that transformation efficiency has an upper threshold of approximately 50% of the population. Counterintuitively, in the presence of multiple transforming DNAs, the fraction of untransformed cells increases to more than 50%. This results in a fail-safe strategy for the population as half of the population generally keeps an intact copy of the original genome. Together, this work advances our understanding of pneumococcal genome plasticity.

Keywords: *Streptococcus pneumoniae*, natural transformation, competence development, homologous recombination, homeologous recombination, single cell analysis, antibiotic resistance, horizontal gene transfer

Introduction

The opportunistic human pathogen *Streptococcus pneumoniae* (the pneumococcus) kills over a million individuals each year, despite the introduction of several vaccines targeting its capsule[1-3]. Because of its ability to take up DNA from its environment by competence activation, genes associated with capsule biosynthesis are rapidly transferred from one strain to the other thereby contributing to vaccine escape[4]. In addition, antibiotic resistance remains a cause of concern and competence-dependent recombination plays an important role in the spread of drug resistance[5]. For example, one of the main genetic sources for penicillin resistance in pneumococcus is DNA acquired from non-pathogenic *Streptococcus* from the viridans group such as *S. mitis* that also lives in the human nasal and oral cavities[6,7]. Consistently, antibiotic resistant pneumococci and vaccine-escape variants remain an important cause of invasive infections in spite of the introduction of the conjugate vaccines[8-10].

Although pneumococcal competence is one of the best studied bacterial regulatory system[4,11-16], and pneumococcal transformation was already discovered in the early twentieth century[17,18], we have a poor understanding on how competence-dependent transformation drives pneumococcal population dynamics, serotype displacement and the spread of antibiotic resistance. Importantly, horizontal gene transfer via natural transformation is not only conserved in Streptococci but is present in many human pathogens where it promotes the spread of virulence determinants and antibiotic resistance[12,19,20]. For this reason, it is crucial to understand what the main bottlenecks are during the take-up and recombination of exogenous DNA that leads to transformed new genotypes.

In contrast to many other competent pathogens such as *Haemophilus influenzae* and *Neisseria meningitidis* in which competence is constitutively expressed, competence development in *S. pneumoniae* is only activated under specific conditions[21,22]. Pneumococcal competence is under control of a two-component quorum sensing system (**Fig. 1**). ComC is cleaved and exported by the peptidase-containing ATP-binding cassette transporter ComAB[23-25]. Cleaved ComC autoinducer is commonly referred to as CSP, for Competence Stimulating Peptide[24,26,27]. CSP is recognized by the membrane-bound histidine kinase ComD[27]. Once a certain threshold level of CSP has been reached, as the culture reaches higher densities, or when other environmental factors increase local CSP concentrations[28,29], ComD will autophosphorylate and transfer the phosphoryl group to the response regulator ComE[30]. Phosphorylated ComE then dimerizes[31,32] and binds to a specific DNA sequence[30,33-35]. The *comCDE* and *comAB* operons are under direct control of ComE, setting up a positive feedback loop. The genes under control of ComE are called the early *com* genes (**Fig. 1**). Importantly, phosphorylated ComE also activates expression of the gene encoding the alternative sigma factor ComX. ComX activates transcription of the so-called late *com* genes, which includes the genes required for DNA uptake and integration[35-39] (**Fig. 1**). While regulation of competence is highly diverse between naturally transformable bacteria, the actual DNA uptake and integration machinery is largely conserved[12,40].

During pneumococcal competence, exogenous double stranded DNA (dsDNA) is bound by a type IV-like pilus[41] and subsequently sequestered to the DNA uptake machinery (**Fig. 1**). Note that in contrast to some other competent bacteria, pneumococcus binds and takes up DNA of any sequence, including non-kin DNA[42]. Next, the dsDNA is processed into single stranded DNA (ssDNA) by the EndA nuclease and internalized through a membrane pore consisting of ComEC. Once inside, the ssDNA is bound by a competence-specific ssDNA binding protein, SsbB, and stabilized by DprA and RecA[43,44] (**Fig. 1**). This complex undergoes homology scanning and forms a temporal hetero-duplex during strand invasion which can lead to homologous recombination[45]. The exact details on the kinetics of this

process, as well as how the heteroduplex is resolved in most cells remains elusive. The competent transformation state in *S. pneumoniae* is transient as DprA interacts with phosphorylated ComE to inhibit its activity[46]. In addition, several key Com proteins are rapidly turned over after their synthesis, leading to a window of DNA uptake of approximately 15 min[47-49].

As most work on pneumococcal competence and transformation has been performed using bulk assays, it is unclear what the actual bottlenecks are during competence development and why one cell will be transformed whereas another one will not. Here, we have set up single cell transformation assays that allow us to quantify successful recombination events in real-time. This study provides direct evidence for several decades-old models underpinning bacterial transformation, and offers new insights that help explain why competence-induced transformation is so effective in changing global pneumococcal population structures.

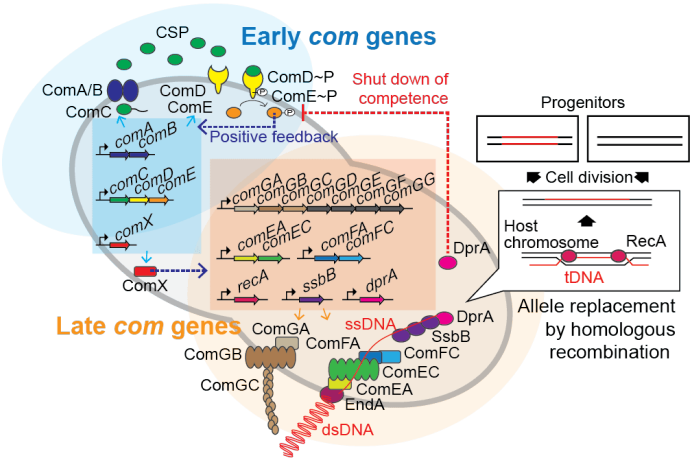

Figure 1. Regulation of pneumococcal competence and transformation. Schematic overview showing representative competence-related genes involved in pneumococcal transformation. Competence development is initiated by activation of the early *com* genes (shown in blue area). ComAB exports ComC and processes it into the competence stimulating peptide (CSP). The two-component system ComDE recognizes CSP and positively regulates the early *com* genes. Subsequently, the alternative sigma factor ComX, activates late *com* gene expression (shown in orange area). ComGA, GB and GC are assembled to form the DNA-binding pilus. EndA is the endonuclease that cleaves dsDNA into ssDNA. ComEA, ComEC, ComFA and ComFC form the ssDNA uptake channel. Internalized foreign ssDNA is protected by SsbB and DprA. DprA ensures the loading of RecA on single strand tDNA to form a presynaptic filament and the resulting DNA scanning complex is capable of homologous (or homeologous) recombination with the recipient chromosome.

Results

All pneumococci become competent and bind exogenous DNA

To quantify pneumococcal transformation efficiency and determine at which step potential bottlenecks arise, we systematically analyzed every stage during the process: 1) competence development, 2) production of the DNA uptake machinery, 3) binding of exogenous DNA, and 4) recombination and expression of the newly acquired genetic information (Fig. 1). While competence development in *B. subtilis* is limited to approximately 10% of the population[50,51], up to 100% of cells within pneumococcal populations have been reported to become competent when induced with exogenously added synthetic CSP or when grown on semi-solid surfaces[28,29,52-55].

To quantify competence development in clonal pneumococcal populations in a systematic fashion, we constructed a set of reporters. First, we assessed the timing of both naturally induced and artificially induced competence (by the addition of synthetic CSP) at the population level utilizing a firefly luciferase reporter under the control of the late competence *ssbB* promoter (strain DLA3). Cells were grown in C+Y medium at 37°C (see Methods) and growth and luciferase activity were measured every 10 min. As expected, under these experimental conditions, the population rapidly activates *ssbB* in the presence of added CSP, while in the absence of externally added CSP, the *ssbB* promoter peaks after approximately 100 min (Fig. 2a). To determine which fraction of the cells switch on the competence pathway, we fused the *ssbB* promoter to a fast folding yellow fluorescent protein (msfYFP) and integrated this construct at the native *ssbB* locus (strain VL2219). As shown in Figure 2b-c, ~97% of the population was positive for *ssbB* expression 20 minutes after addition of synthetic CSP as determined by fluorescence microscopy followed by automated image analysis (see Methods for details). Importantly, spontaneous competence without the addition of synthetic CSP was reached in 92% of the population showing that almost all pneumococci, regardless of their cell length and cell cycle status become naturally competent (Supplementary Fig. S1).

To test whether competent cells actually produce the machinery required for DNA uptake, we constructed translational msfYFP fusions to three essential components of the transformation machinery: ComGA (ATPase driving the DNA uptake pilus), ComEA (DNA receptor) and ComFA (ATPase driving DNA import) as the only copy integrated at their native locus. After 20 min of incubation with synthetic CSP, cells were collected for fluorescence microscopy. In line with the fraction of cells that become competent, msfYFP-ComEA, ComFA-msfYFP and ComGA-msfYFP were also expressed in the majority of the cells (~92%, ~99% and ~99%, respectively) (Fig. 2d, S2). A double-labeled strain (strain OVL2536: P*ssbB*-mScarlet-I, ComGA-msfYFP) demonstrated that all competent cells indeed produce the DNA uptake machinery (Fig. 2e).

Finally, to assess whether the proteins required for recombination and chromosomal integration of exogenous DNA also were expressed in the majority of the population, we constructed translational fusions to RecA and the recombination mediator protein DprA. Similar to the DNA-uptake proteins, RecA and DprA were induced in most competent cells (Figs. 2d and S2).

During pneumococcal competence, the capture of extracellular DNA by the ComGC pilus is an essential step for transformation[56]. To examine which proportion of cells is capable of binding DNA during competence, we labeled extracellular DNA (285 bp *S. pneumoniae* DNA fragment, see Methods) fluorescently with the Cy3 dye. After induction of competence with synthetic CSP of cells mutated for EndA (to prevent degradation of the exogenous DNA), ~90% of the population bound extracellular DNA as visualized by fluorescence microscopy (Fig. 2f). Even without additional CSP, spontaneous competence also led to most cells (89.6%) binding exogenous DNA (Fig. 2g). As observed before in an unencapsulated R6 strain[57], we note that also in the encapsulated serotype 2 D39V strain, DNA mainly bound to the mid-cell positions of the cell, corresponding to the localization of the DNA uptake machinery (Figs. 2f and S2). Collectively, these data demonstrate that pneumococcal competence development, the subsequent production of the DNA uptake and integration machinery, as well as DNA binding is highly efficient and occurs in nearly every cell of the population regardless of their cell cycle state.

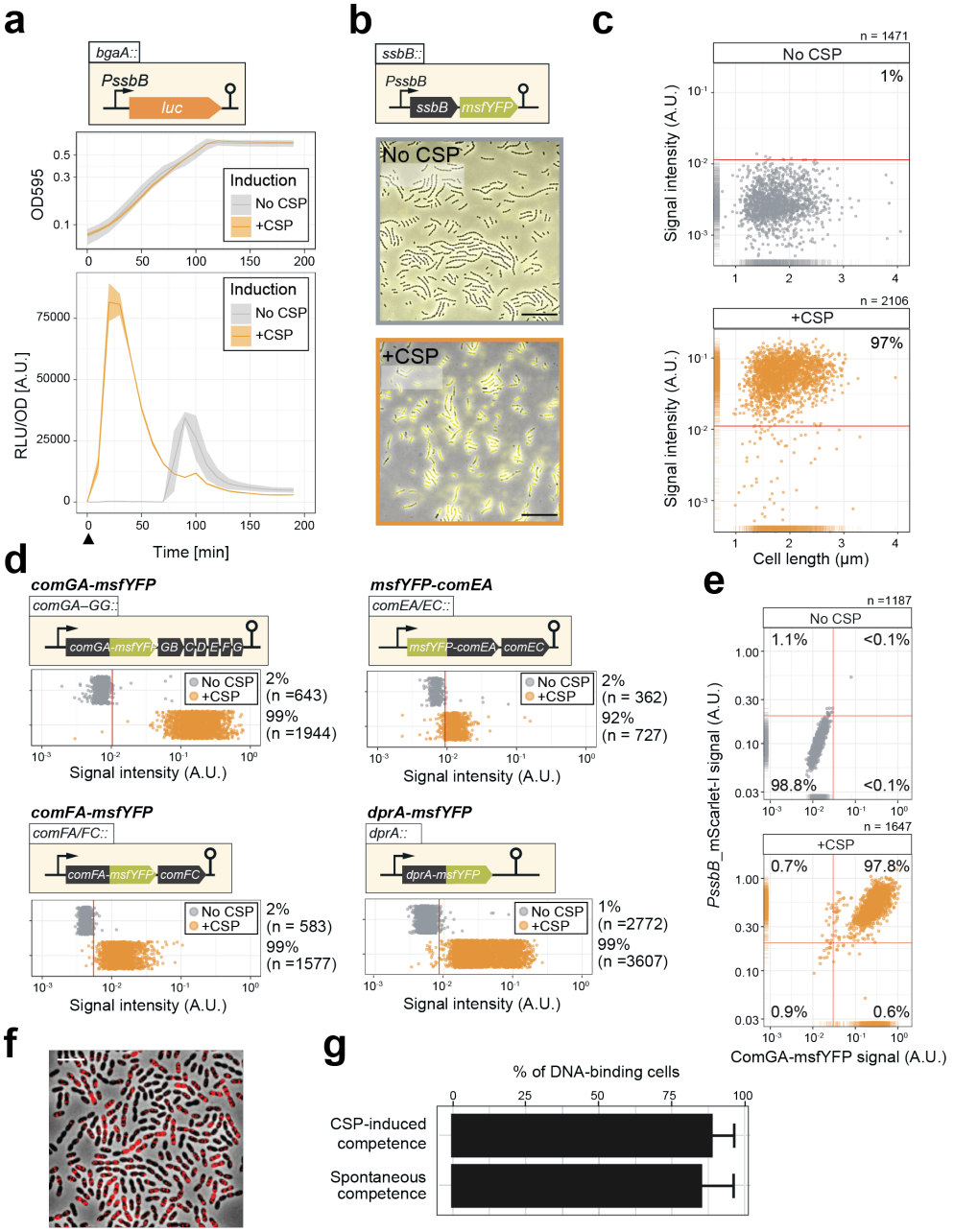

Figure 2. Single cell analysis of competence activation and DNA binding

a. Kinetics of bacterial growth and competence development. Growth curves (top) and OD-normalized bioluminescence activity (bottom) of strain DLA3 (P*ssbB-luc*) in the presence (orange) or absence (grey) of CSP. Arrow indicates the moment after addition of CSP (0 min). Lines and confidence bands represent means of three replicates and SD, respectively. **b.** Single cell imaging of fluorescence competence reporter cells. VL2219 (P*ssbB- msfYFP*) was treated with (top, grey frame) or without (bottom, orange frame) CSP for 20 min and analyzed by fluorescence microscopy. Images are overlays of phase contrast and YFP signal. Scale bar: 20 µm. **c.** Quantification of the imaging. Scatter plots of single cell YFP signal intensity (y axis) against cell length (x axis), based on microscopy images. Red line indicates the threshold used to score YFP positive cells. Proportion of positive cells (%) is shown. **d.** Quantification of cells expressing the transformation machinery. Fluorescence signal intensity for indicated strain harboring *comGA-msfYFP* (VL2536), *msfYFP-comEA* (VL2537), *comFA-msfYFP* (VL2538) or *dprA-msfYFP* (VL3355) treated with (orange) or without (grey) CSP for 20 min. Red line indicates threshold for YFP positive cells. Proportion of positive cells (%) is shown. **e.** Correlation between competence activation and ComGA production. VL2536 (*comGA-msfYFP*, P*ssbB_mScarlet-I*) was incubated with or without CSP. Scatter plot of single cell YFP signal intensity (translational fusion of ComGA) against mScarlet-I signal (transcriptional fusion to *ssbB*). Red line indicates threshold used. Proportion of positive cells (%) is shown on each plot. **f** and **g.** DNA binding analysis using Cy3-labeled DNA added to induced- or spontaneous competent cells. **f.** Representative image of Cy3-labeled DNA-bound TD290 (*ssbB_luc*, Δ*endA*) cells. Scale bar: 2 µm. **g.** Quantification of microscopy images of Cy3-labeled DNA bound D39V cells. Bacteria were treated with CSP for 20 min to induce competence in C+Y (pH 7.9). Efficiency of spontaneous competent cells (no CSP) is also represented. Total of 6027 cells (without added CSP) and 3082 cells (with CSP) were collected over three independent experiments.

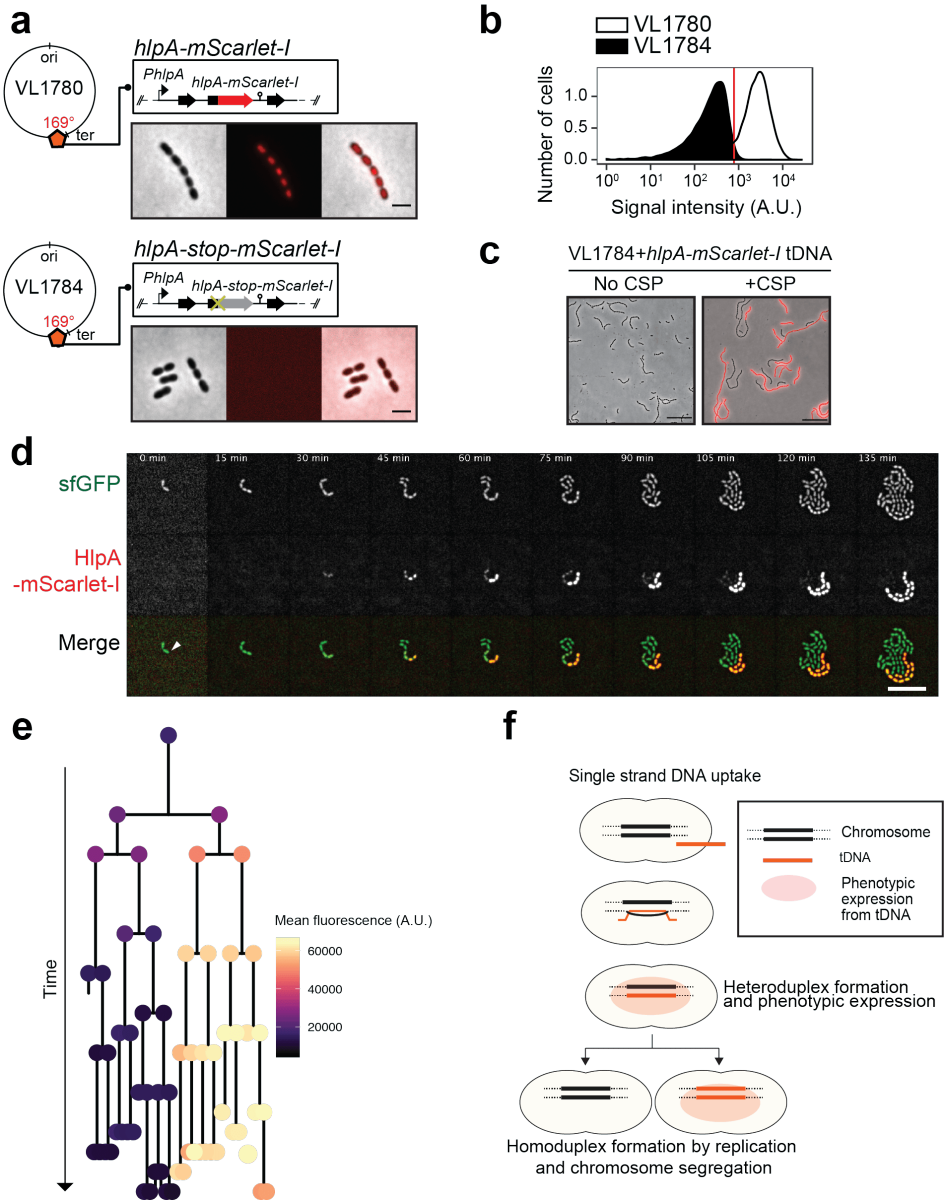

Figure 3. Development of a fluorescence-based real-time reporter for genetic transformation in *S. pneumoniae*
a. Schematic representation of the reporter system. *hlpA-mScarlet-I* was inserted downstream of the native *hlpA* locus as a second copy (strain VL1780), resulting in red-fluorescently marked nucleoids as shown by fluorescence microscopy (Left: phase contrast, middle: red fluorescence, right: overlay, scale bar: 4 µm). A single nucleotide mutation generating a stop codon was introduced in the linker sequence between *hlpA* and *mScarlet-I*, resulting in non-fluorescent strain VL1784 (*hlpA-stop-mScarlet-I*). **b.** Flow-cytometry measurement of *hlpA-mScarlet-I* signal of VL1780 (*hlpA-mScarlet-I*, white) and VL1784 (*hlpA-stop-mScarlet-I*, grey). **c.** CSP-treated (right) or untreated (left) VL1784 was provided with tDNA (*hlpA-mScarlet-I*) and analyzed by fluorescence microscopy after 4 h of incubation. Scale bar, 20 µm. **d.** Time-lapse visualization of transformation with *hlpA-mScarlet-I* in VL1832 (VL1784+constitutively expressing cytoplasmic sfGFP). VL1832 was treated with CSP for 10 min, tDNA added (*hlpA-mScarlet-I*) for 10 min, and then spotted on C+Y agarose pad to start time-lapse imaging with a 5 min interval. Signal of constitutively expressed cytoplasmic sfGFP (top, green in the overlay) was used for cell segmentation in image analysis. Successfully transformed cells were detected by expression of HlpA-mScarlet-I (middle, red in the overlay). Scale bar, 10 µm. Also see supplementary movie 1. **e.** Cell lineage tree with superimposed fluorescence intensity was built based on the time-lapse image shown in E. The quantified mean mScarlet-I signal intensity of each cell during its cell cycle was plotted as a color coded dot onto the lineage tree with each dot corresponding to the moment of 'birth'. Note that the tree represents the lineage from only one of the two progenitor cells (indicated by white arrow in panel d). **f.** Working model for DNA integration and chromosomal segregation of the transformed allele. tDNA is internalized as ssDNA, and recombines to replace one strand on the host chromosome forming a hetero-duplex after recombination. Following chromosomal replication and segregation, the two daughter cells have distinct homo-duplexes (either the original allele or the tDNA allele). Note that initial hetero-duplex formation permits for phenotypic expression from the newly acquired allele if the noncoding strand is replaced by tDNA.

Real-time single-cell analysis of homeologous recombination during transformation

Having established that there are no significant bottlenecks during the process of both induced and natural competence development and DNA uptake, we next set out to develop a system that allows for the direct visualization of successful recombination. Traditionally, transformation efficiencies are evaluated using antibiotic selection methods. However, these selection methods have limitations because they depend on the counting of colony forming units, which can lead to the overestimation of transformation efficiencies, due to inefficient separation of transformed from non-transformed daughter cells and nongenetic inheritance of antibiotic resistance[28,58-60] (Fig. S3). In order to overcome these concerns and analyze successful recombination events during transformation at the single cell level, we developed a fluorescence-based reporter system inspired by a system previously used to observe DNA uptake in *S. pneumoniae*[57]. To do so, we utilized a fluorescent donor strain in which the gene encoding the abundant histone-like protein HlpA (aka HU) was fused in frame with the gene encoding the red fluorescent protein mScarlet-I integrated at the native *hlpA* locus at 169° on the circular chromosome[61] (strain VL1780) (Fig. 3a). A recipient, non-fluorescent strain was constructed (strain VL1784) in which *hlpA* was separated from *mScarlet-I* by a stop codon mutation (G>T) (Figs. 3a-c and S4a, *hlpA-stop-mScarlet-I*). Upon take up, integration and expression of exogenous transforming DNA (tDNA) containing the donor construct (intact *hlpA-mScarlet-I*), successfully transformed recipient cells will produce functional HlpA-mScarlet-I that can be quantified by fluorescence microscopy or flow cytometry (Figs. 3b, 3c and S4b). As this is a recombination event between highly similar but not identical DNA (except for the SNP causing a stop codon), this is called a homeologous recombination event[62,63]. Note that this reporter system does not affect growth regardless of the presence of the stop codon and that flow cytometry analysis slightly overestimates the real transformation efficiencies due to cell chaining (Figs. 5 and S6, see below).

As mScarlet-I is a fast folding red fluorescent protein[64], this reporter system should allow for the real-time detection of homeologous

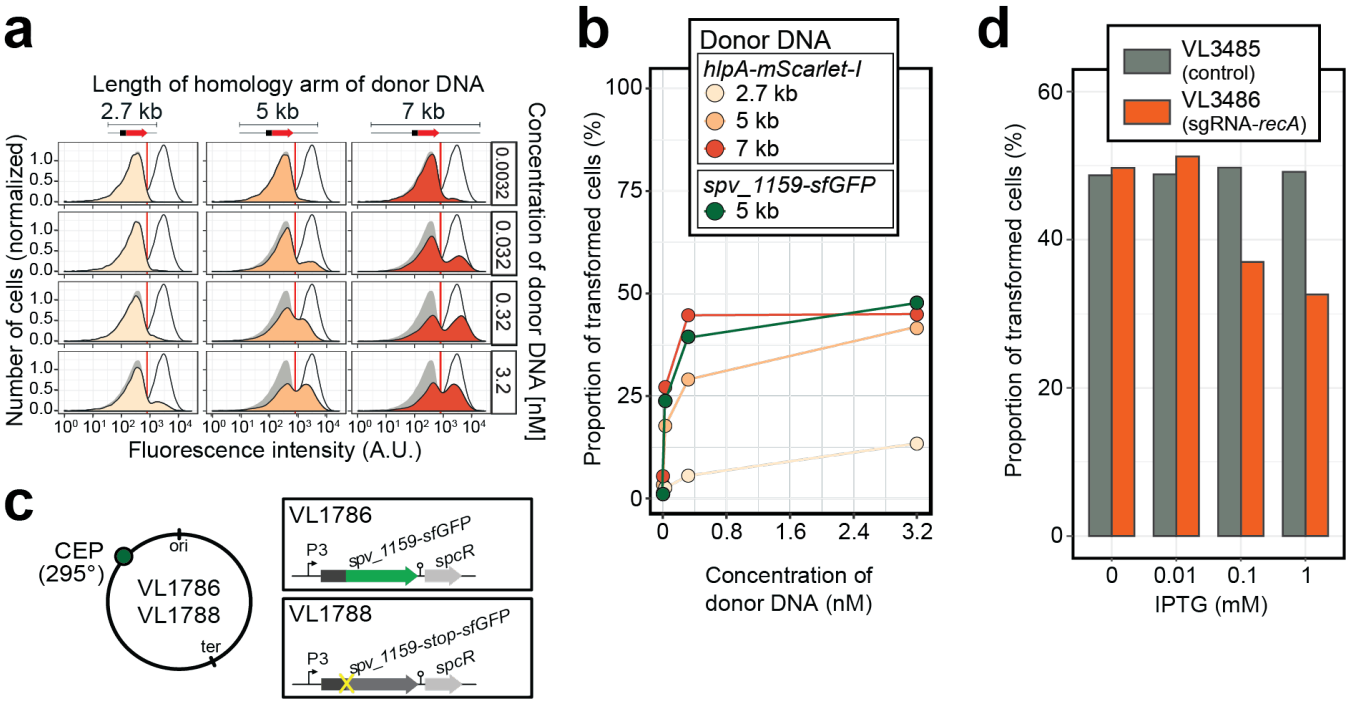

Figure 4. Single cell quantification of recombination reveals an upper level of transformation efficiency
a. Quantification of transformation frequency by flow-cytometry. CSP-treated VL1784 was transformed with *hlpA-mScarlet-I* tDNA of various lengths (2.7 kb, 5 kb or 7 kb) at differing tDNA concentrations (0.0032 nM, 0.032 nM, 0.32 nM or 3.2 nM). The single nucleotide variant to repair the point mutation is located just in the middle of the each fragment (Fig. S4). After 4h incubation post tDNA addition, cell chains were separated (see Fig. S5) and analyzed by flow-cytometry. Negative control without any donor DNA (filled grey histogram) and positive control (VL1780, open histogram) is shown in all panels. Red vertical line indicates the threshold used to score mScarlet-I positive cells. **b.** Correlation between transformation frequency and donor DNA concentration. Transformation frequency was plotted against final concentration of donor DNA. Frequency was calculated by dividing the number of cells with a FL intensity above the threshold by the total number of cell based on flow-cytometry data, as shown in panel A. **c.** Alternative transformation reporter present on a different chromosomal position. The *spv_1159-sfGFP* reporter fusion was cloned into the CEP locus (295°; VL1786, VL1788). A point mutation resulting in a stop codon was introduced in the linker sequence separating *spv_1159* and *sfGFP* in VL1788. **d.** Reduction of transformation efficiency by CRISPRi-based *recA* depletion. CRISPRi-based depletion strains VL3485 (P_{lac}*_dcas9*, no sgRNA control) and VL3486 (P_{lac}*_dcas9*, sgRNA targeting *recA*) were introduced in the *hlpA-stop-mScarlet-I* reporter strain. Strains were pre-grown with or without IPTG (0, 0.01, 0.1 or 1 mM) in acidic C+Y (pH 6.8), and then incubated with CSP (100 ng/µl) in fresh C+Y (pH 7.8) provided with donor tDNA (5 kb length, 0.32 nM). After 4 h of phenotypic expression, transformed cells were analyzed by flow-cytometry.

recombination during transformation. To test this, we provided competent recipient cells that besides the *hlpA-stop-mScarlet-I* allele also constitutively expressed sfGFP (strain VL1832) with intact *hlpA-mScarlet-I* as donor tDNA in the presence of CSP and then performed time-lapse microscopy (see Methods for details). As shown in Fig. 3d and Movies S1 and S2, recipient cells do not display any red fluorescence in the beginning and then gradually start to express red fluorescence. When quantifying the fluorescence signals and superimposing this on a cell lineage tree constructed using a set of new scripts written in BactMAP [65] (see Methods), it becomes apparent that the initial recipient cell already expresses HlpA-mScarlet-I right after the addition of tDNA before the first cell division as red fluorescent signals above background levels can be detected (Fig. 3e). Notably, only half of the recipients' descendants appear to strongly express HlpA-mScarlet-I (Fig. 3e, right lineage). Contrary, after three more divisions the non-transformed lineage no longer expresses red fluorescence (Fig. 3e, left lineage). These results are in line with a recent study in *Vibrio cholerae* that showed a period of non-genetic inheritance in daughter cells during transformation[60]. Similar observations were made when using a different transformation reporter system (Fig. S5 and movie S3, see below). In line with current models of transformation[66-72], these observations are consistent with a model in which recombination occurs by direct integration of the ssDNA donor and forms a hetero-duplex. Therefore, at least one round of DNA replication and division is required to generate two different homo-duplex chromosomes in progeny cells (Fig. 3f). The fact that we initially also observe fluorescence in the un-transformed lineage shows that phenotypic expression derived from the acquired allele can occur prior to forming a homo-duplex. This suggests that in this case the transformed ssDNA likely replaced the antisense, noncoding strand so functional *hlpA-mScarlet-I* could be immediately transcribed after integration via RecA-directed homeologous recombination (mismatched pairing between exchanged DNA strands that are tolerated during the process of homologous recombination).

Single cell quantification of homeologous recombination during transformation

The constructed system now allows us to quantify successful homeologous recombination events at the single cell level, without the bias introduced by traditional plating assays. Previously, it was shown that the concentration of donor DNA as well as the length of the homology regions strongly influences transformation efficiency[61,73]. To examine recombination bottlenecks in our single cell setup, we treated our reporter recipient strain VL1784 with CSP and used intact *hlpA-mScarlet-I* donor tDNA with various lengths of homology surrounding the stop codon (fragments of 2.7 kb, 5 kb or 7 kb) at a range of different concentrations (0.0032 nM, 0.032 nM, 0.32 nM or 3.2 nM). Then, after 4 h incubation in liquid medium to allow for complete homo-duplex

allele formation and dilution of non-genetically inherited HlpA-mScarlet-I, cells were separated from chains by vigorously shaking on a bead beater devise (see Fig. S6). Finally, transformation efficiencies were quantified by flow-cytometry (Fig. 4a). In line with studies using classical plating methods to assess transformation efficiencies[61,73], higher transformation frequencies were observed at higher donor DNA concentrations and with longer homology regions (Fig. 4b). Interestingly, the frequency of transformation plateaued at ~50% regardless of the concentration of donor DNA and sequence homology length (Fig. 4B). This is in contrast to reported transformation frequencies using traditional plating assays where transformation frequencies of higher than 75%[58] and up to 100%[74] have been reported. This discrepancy can be explained by the lack of separation of transformed from non-transformed cells within the counted colony (Fig. S3). To exclude the possibility that the observed limitation in transformation frequency is due to an unique feature of the *hlpA-stop-mScarlet-I* reporter, we constructed an alternative reporter cassette in which we translationally fused the superfolder green fluorescent protein (sfGFP) and SPV_1159, a nonessential small membrane protein under control of the strong constitutive P3 promoter [61,75] cloned into the transcriptionally silent CEP locus at 295° on the circular chromosome (Fig. 4C, strain VL1786). Based on this construct, a recipient strain was constructed containing a stop codon mutation in the linker between *spv_1159* and *sfGFP* (strain VL1788). Indeed, this *spv_1159-sfGFP*-based transformation reporter demonstrated similar transformation characteristics as the *hlpA-mScarlet-I* reporter in time-lapse microscopy and flow-cytometry analysis (Supplementary Fig. S7 and Supplementary movie S3). The transformation frequency of the *spv_1159-sfGFP* reporter was also dependent on donor DNA concentration and never exceeded ~50% (Fig. 4b).

These data show that there is a limit on the maximum efficiency of transformation, despite the fact that most cells become competent and bind extracellular DNA (Fig. 2) and support a model in which in general only one of the recipient allele strands is replaced by the donor DNA[76] (Fig. 3f). Importantly, these experiments indicate that during competence-dependent transformation, given the donor DNA is of sufficient (homology) length and concentration (see Discussion), in principle all targeted loci can be replaced at least on one strand.

Recombination with tDNA is RecA-dependent and independent of mismatch repair

Previous work showed that pneumococcal genetic transformation involves the DNA mismatch repair (MMR) system, which is mediated by HexA[77], and it was suggested that certain alleles upon transformation might be particularly prone to repair[76]. To test whether *hexA* plays a role in our reporter system, we quantified transformation efficiencies in a *hexA* mutant background. This showed no significant recombination differences compared to the wild-type background (Supplementary figure S8). Another potent bacterial defense system against incoming exogenous DNA is restriction-modification (RM) and RM systems have been shown to protect chromosome integrity from invasive DNA such as plasmids or naked foreign DNA[78,79]. To test whether RM influences efficiency of our single cell transformation reporter setup, we deleted the two RM systems present in strain D39V, *dpnI* and *hsdRMS*. As shown in Fig. S7, neither mutant demonstrated altered transformation efficiency. To test if our transformation reporter system depends on the competence-induced homologous recombinase, RecA, we depleted RecA expression level using CRISPR interference[80]. In control strain VL3485 (P_{lac}_*dcas9*, without sgRNA), induction of dCas9 by IPTG did not affect the transformation efficiency with *hlpA-mScarlet-I* tDNA. However, when RecA expression was depleted by induction with IPTG in strain VL3486 (P_{lac}_*dcas9*+sgRNA-recA), the transformation frequency was decreased in an IPTG-dose dependent manner. Note that although RecA is known to be critical for optimal growth in *S. pneumoniae*[81], the CRISPRi depletion levels during competence did not affect bacterial growth (Fig. S9). Collectively, this data shows that our fluorescence-based transformation assay faithfully reflects RecA-dependent homeologous recombination events.

Cell-cycle independent homeologous recombination

It was previously suggested that the genomic location and the cell cycle state might influence transformation efficiency as each heteroduplex needs to be resolved to a homoduplex by DNA replication and loci located close to *oriC* will have multiple copies[57,60,72,82]. To test whether the genomic location of the recombination site and the read orientation of the newly acquired functional allele influences transformation efficiency, we introduced the *spv_1159-stop-sfGFP* reporter at 4 different positions on the chromosome: on the right arm of the chromosome at 101° of the circular chromosome (*bgaA* locus), near the terminus at 169° (*hlpA* locus), on the left arm of the chromosome at 295° (*cep* locus) and near *oriC* at 359° (*comCDE* locus). In addition, *spv_1159-stop-sfGFP* was introduced on both the positive and negative strand on the left arm of the chromosome (*cep* locus at 295°) and on the right arm of the chromosome (*bgaA* locus at 101°). As shown in Figs. 5, and S10, transformation efficiencies were similar across all tested loci and genetic orientations, with a maximal recombination efficiency of approximately 50%.

By performing time-lapse microscopy and tracking cell fates across several generations, we can, in principle, tell whether there was a preference for integration at either the leading or lagging strand (Fig. S11a). By placing the direction of transcription of the reporter on the leading strand, RNAP will thus use the noncoding strand as template. In this situation, only if the noncoding strand is replaced by the donor DNA, fluorescence will be apparent during the first cell cycle upon transformation. If the donor DNA is integrated at the coding strand, it will take one more replication cycle before the heteroduplex is resolved and the noncoding strand contains the intact reporter and fluorescence will be observed later than in the first case (Fig. S11b). Indeed, we can observe all different scenarios with transformants rapidly expressing HlpA-mScarlet-I (possible non-coding strand or double stranded recombinants) and cells that only express HlpA-mScarlet-I after the first cell division (possible coding strand recombinants) (Fig. S11c). Together with the 'bulk' (FACS) single-cell transformation experiments described above, the time-lapse data strongly suggest that there is no preference for integration at either the leading or lagging strand and that this is an unbiased event. These findings correspond with work done by in the 1960s and 1970s that showed that either strand of the incoming dsDNA is degraded randomly by EndA and either strand has a similar chance of being integrated [83,84]. Recent work in *V. cholerae* demonstrated that 7% of transformation events occurred at both strands of the integration site, and it was speculated that this was because of integration of multiple donor ssDNA's replacing both the leading and lagging strand of the recipient[60]. By recording 76 single cell transformation events using time-lapse microscopy, we found 6 cases in which both daughter cells (7.8%) expressed fluorescence, suggestive of dsDNA transformation events. Together, this data shows that heteroduplexes with exogenous DNA are made across all available loci regardless of reading strand or distance to *oriC*.

Direct observations of multiple recombination events in single recipients

The previous experiments demonstrated that, under ideal conditions with long flanking homology regions and high DNA concentrations, all available recombination sites are transformed on at least one of the

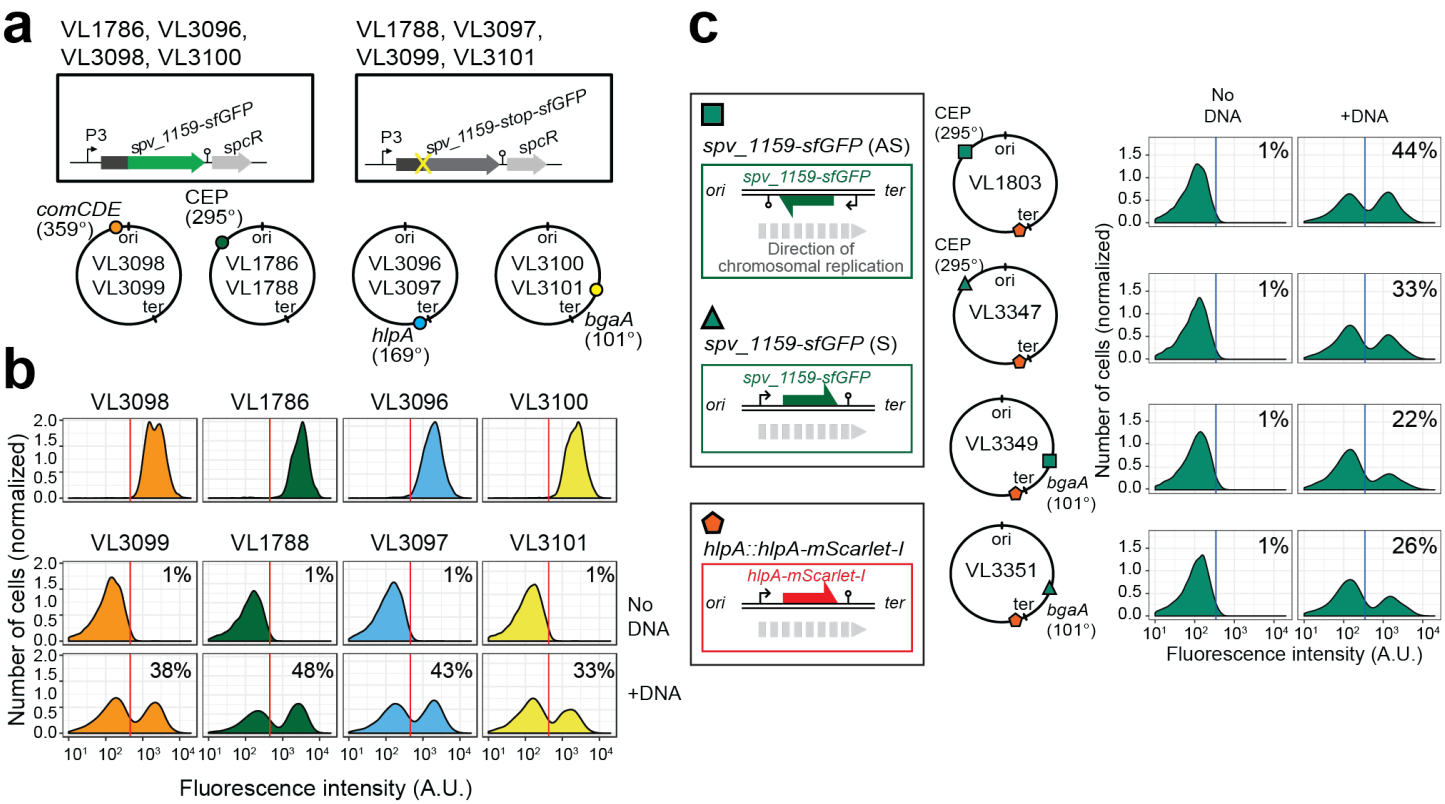

Figure 5. Effect of chromosomal position and strand on recombination potential
a. The spv_1159-sfGFP reporter was cloned into various loci; CEP (295°; VL1786, VL1788), *hlpA* (169°; VL3096, VL3097), *comCDE* (359°; VL3098, VL3099) or *bgaA* (101°; VL3100, VL3101). A point mutation that generates a stop codon was introduced in the linker sequence between *spv_1159* and *sfGFP* for each strain (VL1788, VL3097, VL3099 or VL3101). **b.** Flow cytometry analysis on transformations with intact *spv_1159-sfGFP* tDNA. Strain VL1788, VL3097, VL3099 or VL3101 was transformed with intact *spv_1159-sfGFP* tDNA with 5 kb homology arm at the final concentration of 3.2 nM. 4 h post transformation, cells were separated by beat beating and analyzed by flow-cytometry. The red vertical line indicates the threshold of positive cells in spv_1159-sfGFP signal expression. **c.** Genetic orientation effect on transformation efficiencies. All dual reporter strains VL1803, VL3347, VL3349 or VL3351 harbor both *hlpA-stop-mScarlet-I* and *spv_1159-stop-sfGFP* reporters, but *spv_1159-sfGFP* was cloned at distinct chromosomal positions and different reading directions. *spv_1159-stop-sfGFP* was cloned at the CEP (295°) or *bgaA* (101°) locus resulting in strains VL1803/VL3347 or VL3349/3351, respectively. The coding strand of spv_1159-sfGFP was cloned either in the same direction as the DNA replication fork (read direction on the leading strand) (green triangle, strains VL3347/3351) or in the opposite direction (read direction on the lagging strand) (green square, strains VL1803/3349). The strains were treated with CSP and transformed with corresponding *spv_1159-sfGFP* tDNA (5 kb, 3.2 nM) alone. Transformants were analyzed by flow-cytometry. Vertical blue lines represent the threshold for green fluorescence intensity.

strands. In order to investigate the possibility of observing multiple recombination events, we constructed a dual reporter strain (strain VL1803), which harbors both *hlpA-stop-mScarlet-I* and *spv_1159-stop-sfGFP* at distinct chromosomal locations (Fig. 6A). Transformation efficiencies of this reporter strain with each single donor DNA at the saturated concentration typically reached 50% for both *hlpA-mScarlet-I* and *spv_1159-sfGFP* as quantified by microscopy (Fig. 6C-D). When both donor DNA's were provided, double transformants were observed (15.6 ± 4.4%) as well as single *hlpA-mScarlet-I* transformants (20.2 ± 8.9%) and single *spv_1159-sfGFP* transformants (15.2 ± 4.9%). Time-lapse imaging of competent recipient VL1803 cells with both donor DNAs clearly demonstrated that single recipients could successfully recombine both fragments (Fig. 6B, 6C, Supplementary movies 4 and 5). We note that, on average, the fraction of non-transformed cells are close to 50% (48.9 ± 9.5%), implying that each recombination event is not independent from the next or that there is an upper limit to the number of successful recombinations, otherwise we would expect the fraction of non-transformed cells to decrease with multiple donor DNAs (Fig. S12). An alternative model could be that each recombination event is independent from the next but due to recombination events outside the stop codon SNP, which cannot be quantified in our setup, a reduced transformation efficiency is recorded (see Discussion).

To further explore whether transformation efficiency indeed has a plateau, we constructed a triple reporter strain (VL3127) that harbors *ftsZ-stop-mTurquoise2*, *spv_1159-stop-msfYFP* and *hlpA-stop-mScarlet-I* at three different genomic locations (Fig. 7A). Beside the fact that the fluorescent proteins used are spectrally distinct, every fluorescent reporter also has a specific cellular localization, facilitating automated image analyses of successful recombination. The triple reporter strain was transformed with donor tDNA fragments *ftsZ-mTurquoise2*, *spv_1159-msfYFP* and *hlpA-mScarlet-I*. After 4 h of incubation for fluorescent protein maturation and chromosomal segregation, cells were assessed by fluorescence microscopy. As shown in Fig. 7B and Movie S5, multiple transformed cells with double or triple acquired fluorescence signals were readily observed. Next, we performed single cell transformation assays with strain VL3127 providing one tDNA or all three tDNAs and automatically quantified recombination efficiencies using Oufti and BactMAP-based image analysis[65,85] (Fig. 7C). In line with our previous observations, each single transformation with a saturated concentration of donor tDNA resulted in a recombination efficiency not higher than 50% (Fig. 7C).

Interestingly, every possible recombination event happened within the population: cells were observed in which just a single recombination event took place (the most occurring type of transformation), two recombination events (2.2 ± 0.9%, 4.1 ± 2.7% or 2.1 ± 1.8% for each possible combination) or even three recombination events (1.5 ± 1.1% of all cells). Nevertheless, more than half of the population (58.7 ± 13.4%) did not show any fluorescence when simultaneously transformed with three tDNAs. These observations support a model in which each transformation is in principle independent from the next (Fig. S12).

Non-homologous DNA competes with homologous DNA to reduce transformation efficiency

The data so far show that in principle every locus in *S. pneumoniae* can be efficiently transformed to a maximum of 50% of the cells when providing tDNA at high DNA concentrations and with long homology arms. However, when providing multiple tDNAs the untransformed fraction even increases and becomes greater than 50%. Recently, it was shown using whole genome sequencing of transformation events occurring during contact-dependent DNA uptake, that a single recipient could have at least 29 different recombination events[86]. Together, this suggests that many recombination events are likely going unnoticed in our single cell transformation assay and that these recombination events become limiting, as we can only detect successful recombination when the stop codon in our fluorescent reporter is replaced for a functional allele. If this is true, adding non-specific DNA would compete with donor tDNA resulting in reduced transformation efficiencies. To test this, we utilized homology-unrelated *E. coli*-derived DNA fragments of 5 kb with a similar GC content to *S. pneumoniae* as competing donor DNA. Indeed, as shown in Fig. 7E, co-transformation of strain VL1803 (*hlpA-stop-mScarlet-I*, *spv_1159-stop-sfGFP*) with *E. coli* DNA significantly reduced the transformation efficiency. When 0.32 μM of *hlpA-mScarlet-I* tDNA alone was used as 7 kb donor DNA, approximately 43 % of cells were transformed. However, when 0.32 μM of *hlpA-mScarlet-I* tDNA was given in the presence of saturating amounts of *E. coli* DNA (3.2 μM), only 3% of transformants were observed. Together, this data suggests that the fact that we never reach transformation efficiencies higher than 50% of the population even in the presence of multiple tDNAs, is because of saturation of the DNA uptake and integration machinery. The saturation can be caused by non-successful recombination events with donor DNA or successful recombination events with the donor DNA but outside of the stop SNP that cannot be detected in the single cell assay (Fig. S12).

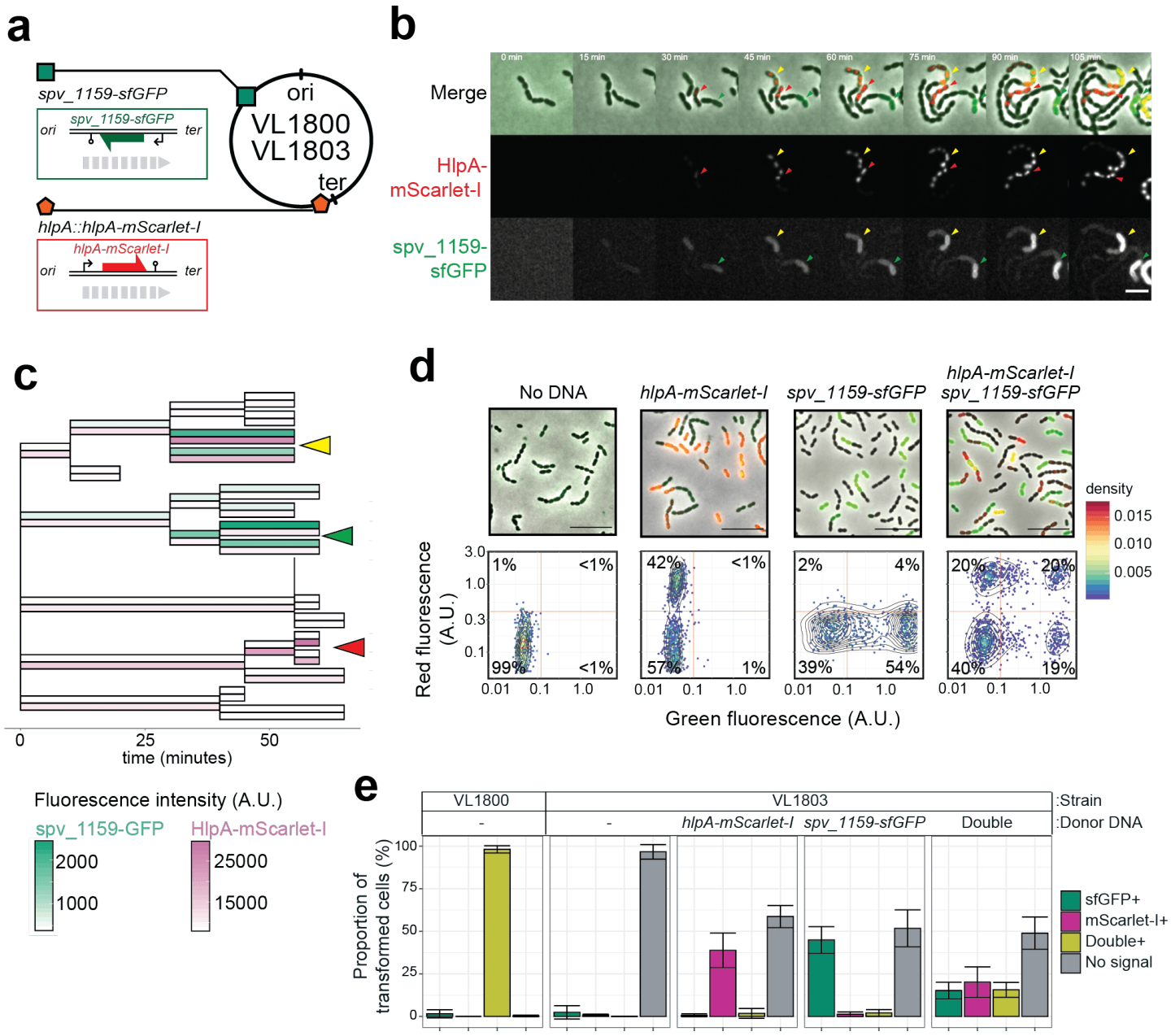

Figure 6. Dual transformation at distinct chromosomal positions

a. Graphical representation of dual reporter strains. Dual transformation reporter strain VL1804 harbors the two transformation reporter constructs *hlpA-stop-mScarlet-I* and *spv_1159-stop-sfGFP*, at the *hlpA* and *CEP* loci, respectively. Donor tDNA (5 kb) *hlpA-mScarlet-I* and *spv_1159-sfGFP* were amplified from strain VL1800 and used for transformation. **b.** Time-lapse visualization of double transformation. Dual reporter strain VL1803 (*hlpA-stop-mScarlet-I, spv_1159-stop-sfGFP*) was treated with CSP for 10 min, provided with 3.2 nM of both *hlpA-mScarlet-I* and *spv_1159-sfGFP* tDNAs (5 kb) for 10 min, and then spotted on a C+Y agarose pad to start time-lapse imaging at 5 min intervals. Successfully transformed cells were detected by expression of HlpA-mScarlet-I (middle, red in merge) and spv_1159-sfGFP (bottom panels, green in merge). Red and Green arrows indicate single transformed cells with *hlpA-mScarlet-I* and *spv_1159-sfGFP* tDNA, respectively. Yellow arrows indicates doubly transformed cells. Scale bar: 4 µm. See supplementary movie S4. **c.** Cell lineage tree with superimposed fluorescence intensity was built based on the time-lapse image shown in B. Means of mScarlet-I (pink) and sfGFP (green) signal intensity of each cell was calculated and displayed with a color bar. Yellow, green and red arrows indicate double transformed, single *spv_1159-sfGFP*-transformed and single *hlpA-mScarlet-I*-transformed lineages, respectively. **d.** Snap shots and quantitative image analysis of transformed populations. Strain VL1803 was transformed with single (*hlpA-mScarlet-I* or *spv_1159-sfGFP*) or double (*hlpA-mScarlet-I*/*spv_1159-sfGFP*) tDNA(s) (5 kb) at final concentration of 3.2 nM. After 4 h of incubation, still images were obtained and the fluorescence intensities were quantified and plotted. Scale bar: 10 µm **e.** Proportion of transformed phenotypes. Stacked bars represent the fraction of single transformed (red or green), double transformed (yellow) and non-transformed (grey) cells. Population of each transformed phenotype was quantified from microscopy images. Bars represent mean ± SD of three independent replicates.

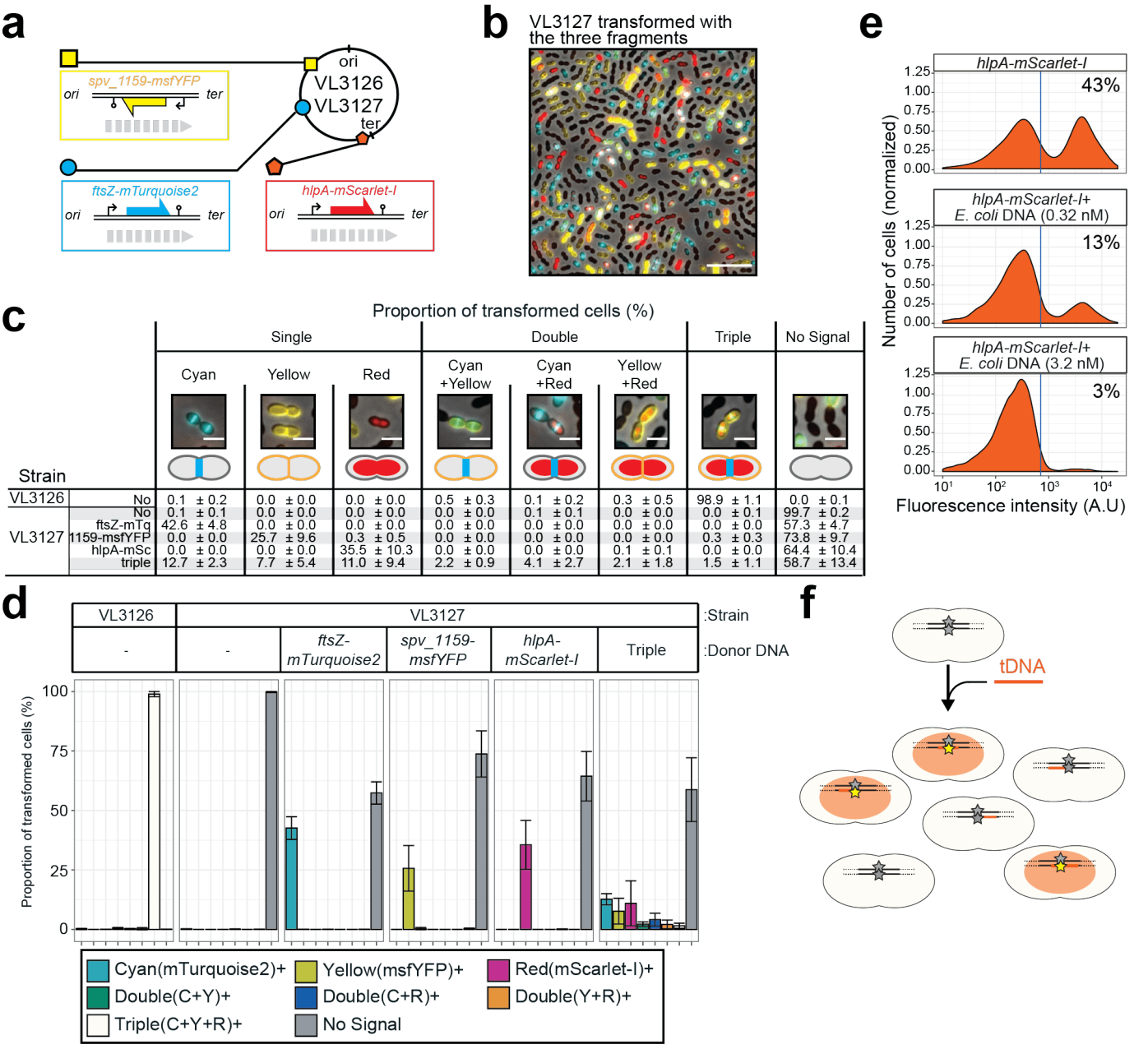

Figure 7. Direct observation of recombination of three separate tDNAs during a single transformation event

a. Schematic representation of the triple labeled strain VL3126 harboring three reporter cassettes: *hlpA-mScarlet-I*, *spv_1159-msfYFP* and *ftsZ-mTurquoise2* at the *hlpA*, *CEP* and *ftsZ* loci, respectively. Strain VL3127 contains stop codon mutations in the linker between each of the fluorescent fusion proteins. Gray arrows indicate the direction of the DNA replication fork relative to the reporter cassette. **b.** Microscope image of strain VL3127 treated with CSP and transformed with the tDNAs of *hlpA-mScarlet-I*, *spv_1159-sfGFP* and *ftsZ-mTurquioise2* (3.2 nM each) amplified from VL3126. Merge image of phase contrast, cyan (FtsZ-mTurquoise2), yellow (spv_1159-msYFP) and red (HlpA-mScarlet-I) fluorescence is shown. Scale bar: 10 µm. **c** and **d.** Proportion of transformed phenotypes. Population of each transformed phenotype was quantified from microscope images. Representative images for each phenotype are shown. Scale bar: 2 µm. **d.** Stacked bars represent proportion of single transformed (cyan, yellow or red), double transformed [green (cyan+yellow), blue (cyan+red), orange (yellow+red)], triple transformed (white) and non-transformed (grey) cells. Bars represent mean ± SD of three independent replicates. **e.** Competition effect of unrelated DNA on transformation frequency. CSP-treated VL1803 was transformed with 7 kb *hlpA-mScarlet-I* tDNA at the final concentration of 0.32 nM in the absence or the presence of an unrelated DNA fragment (0.32 nM or 3.2 nM) amplified from *E. coli*. After incubation of 4 h post transformation, cells were separated and analyzed by flow-cytometry. Red vertical line indicates the threshold of positive cells in mScarlet-I signal expression. The proportion of positive cells (%) is depicted in the plots. **f**. Fragmented tDNA recombination model. The fluorescence-based reporters used in this study rely on replacement of the stop codon SNP (grey star) by intact (amino acid coding) SNP (yellow star) that is located in the middle of the tDNA fragment (orange line). All prepared tDNA molecules have obviously intact SNP, but, integration into host chromosome may take place outside the SNP, which is never distinguished from true untransformed cells by the fluorescence-based system and effectively acting as competing DNA for tDNA's that transform the SNP.

Discussion

The species of *Streptococcus pneumoniae* is vastly diverse with a core genome of approximately 500-1100 orthologues genes and a pan-genome of 5000-7000 orthologs[87]. In addition, many genes are mosaic such as several genes encoding for penicillin-binding proteins in penicillin-resistant clinical strains[88]. One of the main reasons for the high level of genome plasticity and rapidly changing population dynamics is because of the highly conserved competence-based transformation system present in nearly all pneumococcal genomes[89]. Indeed, rapid spread of antibiotic resistance alleles and capsule loci have been observed among human populations under selective pressure[90]. Here, we investigated the molecular basis for competence-dependent transformation at the single-cell level and show that the uptake, integration and expression of tDNA is highly efficient and is largely independent from the recipient's cell cycle stage or of the chromosomal position of the target locus. This was made possible by the setup of a sensitive real-time detection system to quantify successful homeologous recombination events. Indeed, using this system, we could visualize and quantify the recombination of three different tDNAs in single recipient cells demonstrating the efficiency of the pneumococcal transformation process.

Genome sequencing has indicated that up to 29 recombination events may have taken place in a single round of transformation in the same cell when selecting for the transfer of an antibiotic resistance allele[86]. Our work now provides direct evidence that this is not an anomaly and that multiple recombination events are possible during a single transformation episode, even in the absence of selection. Besides shedding light on the efficiency by which transformation can happen in *S. pneumoniae*, by imaging transformation at the single cell level, we provide direct evidence that typically only one recipient strand is replaced during competence-dependent transformation, and that there is no bias towards replacement of the leading or lagging strand. As observed in *V. cholerae*, in approximately 7% of transformants, both strands can be replaced, which is likely caused by DNA repair leading to removal of the recipient strand on the heteroduplex or by integration of multiple tDNAs[60]. This is in line with predictions made using unlinked antibiotic resistance alleles[82]. In addition, our single cell observations demonstrate that the replaced noncoding strand by recombination within the heteroduplex is immediately transcribed by RNAP and can lead to lineages of cells with non-genetic inherited phenotypes (Fig. 3).

We show that any site regardless of its chromosomal position or orientation with regards to DNA replication can be efficiently transformed, although not with the exact same efficiencies (Fig. S10). Possible explanations for local difference in recombination efficiency could be the levels of DNA compaction or transcription activity. Interestingly, we find that the percentage of untransformed cells is lower when three tDNAs are provided instead of two tDNAs (~58% vs ~49% of untransformed cells, respectively: Figs. 6 and 7). Together with the observation that the presence of non-homologous DNA reduced our observed transformation efficiency (Fig. 7E), suggests that, in principle, every recombination event is independent of the next, but that many unsuccessful recombination events and successful recombination events outside the stop codon of our reporter are taking place and that this limits the efficiency of site-specific recombination (Fig. S12, Fig. 7f).

The overall biological implication of the limitation on competence-dependent transformation is that this mechanism ensures that in most cases one copy of the original recipient DNA remains unaltered. This might represent a fail-safe mechanism so that in case a deleterious tDNA is incorporated, at least one daughter cell will survive. Being able to sample from a large pan-genome might contribute to the vast genome plasticity and genome diversity as observed in natural pneumococcal populations. Future single-cell work will allow the investigation of the localization of the enzymes involved in transformation, how strand exchange during transformation occurs and what the dynamics of the molecular machines are during DNA uptake, integration and expression of tDNA.

Materials and Methods

Bacterial strains and growth condition

All pneumococcal strains used in this study are derivatives of serotype 2 *S. pneumoniae* D39V[18,91] unless specified otherwise. See Table S1 for a list of the strains used and the Supplemental information for details on the construction of the strains. *S. pneumoniae* was grown in C+Y (pH 6.8) medium at 37°C. C+Y was adapted from Adams and Roe (Adams and Roe 1945) and contained the following compounds: adenosine (68.2 mM), uridine (74.6 mM), L-asparagine (302 mM), L-cysteine (84.6 mM), L-glutamine (137 mM), L-tryptophan (26.8 mM), casein hydrolysate (4.56 g L^{-1}), BSA (729 mg L^{-1}), biotin (2.24 mM), nicotinic acid (4.44 mM), pyridoxine (3.10 mM), calcium pantothenate (4.59 mM), thiamin (1.73 mM), riboflavin (0.678 mM), choline (43.7 mM), $CaCl_2$ (103 mM), K2HPO4 (44.5 mM), $MgCl_2$ (2.24 mM), $FeSO_4$ (1.64 mM), $CuSO_4$ (1.82 mM), $ZnSO_4$ (1.58 mM), $MnCl_2$ (1.29 mM), glucose (10.1 mM), sodium pyruvate (2.48 mM), saccharose (861 mM), sodium acetate (22.2 mM) and yeast extract (2.28 g L^{-1}).

Luminescence assays of competence development

To monitor competence development, strains containing a transcriptional fusion of the firefly *luc* gene with the late competence gene *ssbB* were used. Cells were pre-cultured in C+Y (pH 6.8) at 37ºC to an OD595 nm of 0.2. Right before inoculation, cells were collected by centrifugation (6,000 xg for 3 minutes) and resuspended in fresh C+Y at pH 7.9, which is permissive for natural spontaneous competence. Luciferase assays were performed in 96-wells plates with a Tecan Infinite 200 PRO illuminometer (TECAN) at 37ºC as described before [53]. Luciferin was added at a concentration of 0.45 mg/mL to monitor competence by means of luciferase activity. Optical density (OD595nm) and luminescence (relative luminescence units [RLU]) were measured every 10 minutes.

Phase contrast and fluorescence microscopy

Microscopy acquisition was performed using a Leica DMi8 microscope with a sCMOS DFC9000 (Leica) camera and a SOLA light engine (Lumencor) and a 100x/1.40 oil-immersion objective. Images were primarily processed using LAS X (Leica).For snap shot imaging, cells were concentrated 10x by centrifugation (6,000 xg, 3 min) and 0.5 µl of cells were spotted on 1% agarose/PBS. For time-lapse microscopy, a semi-solid growth surface was prepared with C+Y (pH 7.9) containing 1% agarose in Gene Frame (Thermo Fischer)[92]. As C+Y medium has some background fluorescence, the C+Y agar pad was pre-exposed on a UV illuminator for 1 min to bleach the background fluorescence.

Phase contrast images were acquired using transmission light with 100 ms exposure for snap shot and 50 ms exposure for time-lapse. Fluorescence was usually acquired with 700 ms exposure for snap shot, and 200–500 ms exposure (17–30% of power from Sola light engine) for time-lapse using filter settings described below. Time-lapses images were recorded by taking images every 5 or 10 minutes.

Leica DMi8 filters set used are as followed: mTurquoise2 (Ex: 430/24 nm Chroma, BS: LP 455 Leica 11536022, Em: 470/24 nm Chroma ET470/24 nm or Ex: 430/29 nm Chroma, BS: 455 (450–490) Chroma 69008, Em: 470/26), sfGFP (Ex: 470/40 nm Chroma ET470/40x, BS: LP 498 Leica 11536022, Em: 520/40 nm Chroma ET520/40m), msfYFP (Ex: 500/20 nm Chroma ET500/20x, BS: LP 520 Leica 11536022, Em: 535/30 nm Chroma ET535/30m or Ex:495/25 nm Chroma ET495/25x, BS520 (510–560) Chroma 69008, Em: 533/30 nm) and mScarlet-I (Chroma 49017, Ex: 560/40 nm, BS: LP 590 nm, Em: LP 590 nm or Ex: 575/35 nm, BS: 595 (590–670) nm Chroma 69008, Em: 635/70 nm).

Quantitative image analysis

For quantitative image analysis of single cells, obtained microscopic images were processed by FIJI software[93]. Single cell segmentation and fluorescence signal intensity measurement were performed by Oufti[85]. The generated celllist files were analyzed in R (https://www.r-project.org/), using BactMAP [65] for statistical analysis and visualization. After celllist file were imported into R, cells were filtered between 0.7–1.2 µm in width length to exclude false events derived from noise or miss-segmentation. Threshold of fluorescence of signal intensity was defined based on negative or positive control for each experiment setting. >500 cells were analyzed at least for each replicate. To exclude the possibility of overlap in detection of fluorescence (particularly mTurquoise2/msfYFP and msfYFP/mScarlet-I) in multi-fragments transformation, we ensured that single transformation experiments did not show any singal in the other channels and this was confirmed by looking at the protein localization patterns.

For generating cell lineage trees from time-lapse imaging, the stacked time-lapse images were processed by FIJI[93] and stabilization between time frames was performed by Huygens (Scientific volume imaging). Single cell segmentation and fluorescence intensity acquisition were performed by SuperSegger[94]. The resultant data set was analyzed using BactMAP[65].

DNA binding assays

Analysis of DNA binding was performed in an *endA* mutant background (strain D39V *ssbB::luc (cam) endA::kan*), to favor accumulation of transforming DNA at the surface of competent cells, as previously shown[57]. After gentle thawing stock cultures, aliquots were inoculated at OD550 0.006 in C+Y medium, supplemented with 20mM HCl to prevent spontaneous competence development, and grown at 37°C to an OD550 of 0.3. These precultures were inoculated (1/50) in C+Y medium (pH 7.8) and incubated at 37°C. In these conditions, competence developed spontaneously and reached its maximal level in the population after 55-60 minutes. At 35 minutes, 1 ml samples were collected and induced, or not, with synthetic CSP (50 ng ml−1). At 50 minutes, these samples were incubated for 5 minutes with 10 ng of a 285 bp DNA fragment labelled with a Cy3 fluorophore at its 5' extremities[57]. Cells were pelleted (3,000 xg, 3 min), washed twice in 500µl C+Y, and resuspended in 20 to 50 µL C+Y medium before microscopy. Two µl of this suspension was spotted on a microscope slide containing a slab of 1.2% C+Y agarose as described previously[92].

Phase contrast and fluorescence microscopy were performed with an automated inverted epifluorescence microscope Nikon Ti-E/B, a phase contrast objective (CFI Plan Apo Lambda DM 100X, NA1.45), a Semrock filter set for Cy3 (Ex: 531BP40; DM: 562; Em: 593BP40), a LED light source (Spectra X Light Engine, Lumencor), and a sCMOS camera (Neo sCMOS, Andor). Images were captured and processed using the Nis-Elements AR software (Nikon). Cy3 fluorescence images were false colored red and overlaid on phase contrast images. Overlaid images were further analyzed to quantify the number of cells bound with Cy3-labelled DNA. Single cells were first detected using the threshold command from Nis-Elements and cells bound or not to DNA were manually classified using the taxonomy tool.

Evaluation of transformation frequency using the fluorescence reporter

To quantify the efficiency of transformation with tDNA fragments, reporter cells were pre-cultured in C+Y (pH 6.8) at 37ºC to an OD595 nm of 0.2. Right before inoculation, cells were collected by centrifugation (6,000 xg for 3 minutes) and resuspended in fresh C+Y at pH 7.9, adjusted to OD=0.1. Competence was induced by incubation in the presence of CSP (100 ng/µl) at 37°C for 10 min and then donor tDNA was provided the indicated concentration. After an additional 4 h of incubation at 37°C for complete cell division to form homo-duplex and maturation of fluorescence proteins, cells were placed on ice to stop cell growth and were directly analyzed by fluorescence microscopy or flow-cytometry.

Donor tDNA was designed in such a way that the single nucleotide mutation is positioned in the middle of the entire fragment so that the left and right homology arms are of equal length. Preparation of the tDNA was performed by PCR using primer pairs indicated in Supplementary table S2, using the corresponding parent strain as template. For competition experiments using unrelated tDNA in Fig. 7, DNA fragment that has no homology to pneumococcal genome but equal size (5 kb) and GC contents (~40% of GC) to *hlpA-mScarlet-I/spv_1159-sfGFP* fragments was amplified from *E. coli* DH5alpha[95].

Flow-cytometry analysis

Cells were collected by centrifugation (6,000 xg for 3 minutes) and resuspended in filtered (0.22 µm) PBS adjusted to a cell density of approximately $1.0x10^5$~$1.0x10^6$ cells/mL. As *S. pneumoniae* D39V cells tend to form chains particularly in competence-activated cells, we separated cells by bead beating (BioSpec) without any glass beads. At least >$1.0x10^4$ events were analyzed by a Novocyte flow cytometer (ACEA bioscience) harboring 488 nmand 561 nm lasers. Fluorescence filters used were: FITC (Ex: 488nm, Em: 530/45nm) for sfGFP and PE.Texas.Red (Ex: 561 nm, Em: 615/20 nm) for mScarlet-I. Obtained raw data was imported and analyzed in R. Non-bacterial particles were excluded by gating the FSC and SSC values. A threshold was determined so that positive events counted in a negative control strain were <1% and validated with both negative (no DNA control) and positive (parent strain without the point mutation) control for each experimental setting.

Author Contributions and Notes

J.K. and J.W.V. designed research, J.K., N.C., R.v.R. and G.C. performed research, J.K. and N.C. analyzed data; and J.K. P.P. and J.W.V. wrote the paper.

The authors declare no conflict of interest.

This article contains supporting information online.

Acknowledgments

We thank Jelle Slager and Arnau Domenech for critically reading this manuscript and all members of the Veening lab for stimulating discussions. Work in the Veening lab is supported by the Swiss National Science Foundation (SNSF) (project grant 31003A_172861), a JPIAMR grant (40AR40_185533) from SNSF and ERC consolidator grant 771534-PneumoCaTChER. Work in the Polard lab is supported by the Centre National de la Recherche Scientifique, Université Paul Sabatier, and the Agence Nationale de la Recherche (grants ANR-10-BLAN-1331 and EXStasis-17-CE13-0031-01). Jun Kurushima was supported by The Naito Foundation.

References

1. Croucher, N. J., Løchen, A. & Bentley, S. D. Pneumococcal Vaccines: Host Interactions, Population Dynamics, and Design Principles. *Annu. Rev. Microbiol.* **72,** 521–549 (2018).
2. O'Brien, K. L. *et al.* Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. *Lancet* **374,** 893–902 (2009).
3. Prina, E., Ranzani, O. T. & Torres, A. Community-acquired pneumonia. *Lancet* **386,** 1097–1108 (2015).
4. Salvadori, G., Junges, R., Morrison, D. A. & Petersen, F. C. Competence in *Streptococcus pneumoniae* and Close Commensal Relatives: Mechanisms and Implications. *Front Cell Infect Microbiol* **9,** 94 (2019).
5. Lo, S. W. *et al.* A mosaic tetracycline resistance gene tet(S/M) detected in an MDR pneumococcal CC230 lineage that underwent capsular switching in South Africa. *J. Antimicrob. Chemother.* **19,** 759 (2019).
6. Bryskier, A. Viridans group streptococci: a reservoir of resistant bacteria in oral cavities. *Clin. Microbiol. Infect.* **8,** 65–69 (2002).

7. Janoir, C., Podglajen, I., Kitzis, M. D., Poyart, C. & Gutmann, L. In vitro exchange of fluoroquinolone resistance determinants between *Streptococcus pneumoniae* and viridans streptococci and genomic organization of the parE-parC region in S. mitis. *Journal of Infectious Diseases* **180,** 555–558 (1999).
8. Fenoll, A. *et al.* Serotypes and genotypes of S. *pneumoniae* isolates from adult invasive disease in Spain: A 5-year prospective surveillance after pediatric PCV13 licensure. The ODIN study. *Vaccine* **36,** 7993–8000 (2018).
9. Levy, C. *et al.* Diversity of Serotype Replacement After Pneumococcal Conjugate Vaccine Implementation in Europe. *J. Pediatr.* **213,** 252–253.e3 (2019).
10. Ouldali, N. *et al.* Incidence of paediatric pneumococcal meningitis and emergence of new serotypes: a time-series analysis of a 16-year French national survey. *Lancet Infect Dis* **18,** 983–991 (2018).
11. Gómez-Mejia, A., Gámez, G. & Hammerschmidt, S. *Streptococcus pneumoniae* two-component regulatory systems: The interplay of the pneumococcus with its environment. *Int. J. Med. Microbiol.* **308,** 722–737 (2018).
12. Johnston, C., Martin, B., Fichant, G., Polard, P. & Claverys, J.-P. Bacterial transformation: distribution, shared mechanisms and divergent control. *Nature Reviews Microbiology* **12,** 181–196 (2014).
13. Lin, M. & Kussell, E. Correlated Mutations and Homologous Recombination Within Bacterial Populations. *Genetics* **205,** 891–917 (2017).
14. Shanker, E. & Federle, M. J. Quorum Sensing Regulation of Competence and Bacteriocins in *Streptococcus pneumoniae* and mutans. *Genes (Basel)* **8,** 15 (2017).
15. Straume, D., Stamsås, G. A. & Håvarstein, L. S. Natural transformation and genome evolution in *Streptococcus pneumoniae*. *Infect. Genet. Evol.* **33,** 371–380 (2015).
16. Veening, J.-W. & Blokesch, M. Interbacterial predation as a strategy for DNA acquisition in naturally competent bacteria. *Nature Reviews Microbiology* **15,** 621–629 (2017).
17. Griffith, F. The Significance of Pneumococcal Types. *J Hyg (Lond)* **27,** 113–159 (1928).
18. Avery, O. T., Macleod, C. M. & McCarty, M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation by a deoxyribonucleic acid fraction isolated from pneumococcus type III. *J. Exp. Med.* **79,** 137–158 (1944).
19. Dubnau, D. & Blokesch, M. Mechanisms of DNA Uptake by Naturally Competent Bacteria. *Annu. Rev. Genet.* **53,** 217–237 (2019).
20. Brockhurst, M. A. *et al.* The Ecology and Evolution of Pangenomes. *Curr. Biol.* **29,** R1094–R1103 (2019).
21. Blokesch, M. Natural competence for transformation. *Curr. Biol.* **26,** R1126–R1130 (2016).
22. Claverys, J.-P., Prudhomme, M. & Martin, B. Induction of competence regulons as a general response to stress in gram-positive bacteria. *Annu. Rev. Microbiol.* **60,** 451–475 (2006).
23. Chandler, M. S. & Morrison, D. A. Identification of two proteins encoded by com, a competence control locus of *Streptococcus pneumoniae*. *J. Bacteriol.* **170,** 3136–3141 (1988).
24. Håvarstein, L. S., Coomaraswamy, G. & Morrison, D. A. An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus pneumoniae*. *Proc Natl Acad Sci U S A* **92,** 11140–11144 (1995).
25. Hui, F. M., Zhou, L. & Morrison, D. A. Competence for genetic transformation in *Streptococcus pneumoniae*: organization of a regulatory locus with homology to two lactococcin A secretion genes. *Gene* **153,** 25–31 (1995).
26. Alloing, G., Granadel, C., Morrison, D. A. & Claverys, J. P. Competence pheromone, oligopeptide permease, and induction of competence in *Streptococcus pneumoniae*. *Mol. Microbiol.* **21,** 471–478 (1996).
27. Håvarstein, L. S., Gaustad, P., Nes, I. F. & Morrison, D. A. Identification of the streptococcal competence-pheromone receptor. *Mol. Microbiol.* **21,** 863–869 (1996).
28. Domenech, A., Slager, J. & Veening, J.-W. Antibiotic-Induced Cell Chaining Triggers Pneumococcal Competence by Reshaping Quorum Sensing to Autocrine-Like Signaling. *Cell Rep* **25,** 2390–2400.e3 (2018).
29. Moreno-Gámez, S. *et al.* Quorum sensing integrates environmental cues, cell density and cell history to control bacterial competence. *Nat Commun* **8,** 854 (2017).
30. Martin, B. *et al.* ComE/ComE~P interplay dictates activation or extinction status of pneumococcal X-state (competence). *Mol. Microbiol.* **87,** 394–411 (2013).
31. Boudes, M. *et al.* Structural insights into the dimerization of the response regulator ComE from *Streptococcus pneumoniae*. *Nucleic Acids Res.* **42,** 5302–5313 (2014).
32. Sanchez, D., Boudes, M., van Tilbeurgh, H., Durand, D. & Quevillon-Cheruel, S. Modeling the ComD/ComE/comcde interaction network using small angle X-ray scattering. *FEBS J.* **282,** 1538–1553 (2015).
33. Pestova, E. V., Håvarstein, L. S. & Morrison, D. A. Regulation of competence for genetic transformation in *Streptococcus pneumoniae* by an auto-induced peptide pheromone and a two-component regulatory system. *Mol. Microbiol.* **21,** 853–862 (1996).
34. Ween, O., Gaustad, P. & Håvarstein, L. S. Identification of DNA binding sites for ComE, a key regulator of natural competence in *Streptococcus pneumoniae*. *Mol. Microbiol.* **33,** 817–827 (1999).
35. Slager, J., Aprianto, R. & Veening, J.-W. Refining the pneumococcal competence regulon by RNA-sequencing. *J. Bacteriol.* JB.00780–18 (2019). doi:10.1128/JB.00780-18
36. Campbell, E. A., Choi, S. Y. & Masure, H. R. A competence regulon in *Streptococcus pneumoniae* revealed by genomic analysis. *Mol. Microbiol.* **27,** 929–939 (1998).
37. Dagkessamanskaia, A. *et al.* Interconnection of competence, stress and CiaR regulons in *Streptococcus pneumoniae*: competence triggers stationary phase autolysis of ciaR mutant cells. *Mol. Microbiol.* **51,** 1071–1086 (2004).
38. Pestova, E. V. & Morrison, D. A. Isolation and characterization of three *Streptococcus pneumoniae* transformation-specific loci by use of a lacZ reporter insertion vector. *J. Bacteriol.* **180,** 2701–2710 (1998).
39. Luo, P., Li, H. & Morrison, D. A. ComX is a unique link between multiple quorum sensing outputs and competence in *Streptococcus pneumoniae*. *Mol. Microbiol.* **50,** 623–633 (2003).
40. Chen, I. & Dubnau, D. DNA uptake during bacterial transformation. *Nature Reviews Microbiology* **2,** 241–249 (2004).
41. Laurenceau, R. *et al.* A type IV pilus mediates DNA binding during natural transformation in *Streptococcus pneumoniae*. *PLoS Pathog* **9,** e1003473 (2013).
42. Mell, J. C. & Redfield, R. J. Natural competence and the evolution of DNA uptake specificity. *J. Bacteriol.* **196,** 1471–1483 (2014).
43. Attaiech, L. *et al.* Role of the single-stranded DNA-binding protein SsbB in pneumococcal transformation: maintenance of a reservoir for genetic plasticity. *PLoS Genet* **7,** e1002156 (2011).
44. Berge, M., Mortier-Barrière, I., Martin, B. & Claverys, J.-P. Transformation of *Streptococcus pneumoniae* relies on DprA- and RecA-dependent protection of incoming DNA single strands. *Mol. Microbiol.* **50,** 527–536 (2003).
45. Mortier-Barrière, I. *et al.* A key presynaptic role in transformation for a widespread bacterial protein: DprA conveys incoming ssDNA to RecA. *Cell* **130,** 824–836 (2007).

46. Mirouze, N. *et al.* Direct involvement of DprA, the transformation-dedicated RecA loader, in the shut-off of pneumococcal competence. *Proc Natl Acad Sci U S A* **110,** E1035–44 (2013).
47. Tomasz, A. Model for the mechanism controlling the expression of competent state in Pneumococcus cultures. *J. Bacteriol.* **91,** 1050–1061 (1966).
48. Weng, L., Piotrowski, A. & Morrison, D. A. Exit from competence for genetic transformation in *Streptococcus pneumoniae* is regulated at multiple levels. *PLoS ONE* **8,** e64197 (2013).
49. Liu, Y. *et al.* HtrA-mediated selective degradation of DNA uptake apparatus accelerates termination of pneumococcal transformation. *Mol. Microbiol.* **112,** 1308–1325 (2019).
50. Smits, W. K. *et al.* Stripping Bacillus: ComK auto-stimulation is responsible for the bistable response in competence development. *Mol. Microbiol.* **56,** 604–614 (2005).
51. Maamar, H. & Dubnau, D. Bistability in the Bacillus subtilis K-state (competence) system requires a positive feedback loop. *Mol. Microbiol.* **56,** 615–624 (2005).
52. Litt, M., Marmur, J., Ephrussi-Taylor, H. & Doty, P. The dependence of pneumococcal transformation on the molecular weight of deoxyribose nucleic acid. *Proc Natl Acad Sci U S A* **44,** 144–152 (1958).
53. Slager, J., Kjos, M., Attaiech, L. & Veening, J.-W. Antibiotic-induced replication stress triggers bacterial competence by increasing gene dosage near the origin. *Cell* **157,** 395–406 (2014).
54. Martin, B. *et al.* Expression and maintenance of ComD-ComE, the two-component signal-transduction system that controls competence of *Streptococcus pneumoniae*. *Mol. Microbiol.* **75,** 1513–1528 (2010).
55. Bergé, M. J. *et al.* A programmed cell division delay preserves genome integrity during natural genetic transformation in *Streptococcus pneumoniae*. *Nat Commun* **8,** 1621–13 (2017).
56. Berge, M., Moscoso, M., Prudhomme, M., Martin, B. & Claverys, J.-P. Uptake of transforming DNA in Gram-positive bacteria: a view from *Streptococcus pneumoniae*. *Mol. Microbiol.* **45,** 411–421 (2002).
57. Bergé, M. J. *et al.* Midcell recruitment of the DNA uptake and virulence nuclease, EndA, for pneumococcal transformation. *PLoS Pathog* **9,** e1003596 (2013).
58. Ephrussi-Taylor, H. The mechanism of deoxyribonucleic acid-induced transformations. *Recent Progress in Microbiology* 5148 (1958).
59. Ephrussi-Taylor, H. Appearance of Streptomycin Resistance Following the Uptake of Transforming Deoxyribonucleic Acid in Pneumococcus. *Nature* **196,** 748–752 (1962) .
60. Dalia, A. B. & Dalia, T. N. Spatiotemporal Analysis of DNA Integration during Natural Transformation Reveals a Mode of Nongenetic Inheritance in Bacteria. *Cell* **179,** 1499–1511.e10 (2019).
61. Keller, L. E., Rueff, A.-S., Kurushima, J. & Veening, J.-W. Three New Integration Vectors and Fluorescent Proteins for Use in the Opportunistic Human Pathogen *Streptococcus pneumoniae*. *Genes (Basel)* **10,** 394 (2019).
62. Petit, M. A., Dimpfl, J., Radman, M. & Echols, H. Control of large chromosomal duplications in Escherichia coli by the mismatch repair system. *GeneticS11***9,** 327–332 (1991).
63. Humbert, O., Prudhomme, M., Hakenbeck, R., Dowson, C. G. & Claverys, J. P. Homeologous recombination and mismatch repair during transformation in *Streptococcus pneumoniae*: saturation of the Hex mismatch repair system. *Proc Natl Acad Sci U S A* **92,** 9052–9056 (1995).
64. Bindels, D. S. *et al.* mScarlet: a bright monomeric red fluorescent protein for cellular imaging. *Nat. Methods* **14,** 53–56 (2017).
65. van Raaphorst, R., Kjos, M. & Veening, J.-W. BactMAP: An R package for integrating, analyzing and visualizing bacterial microscopy data. *Mol. Microbiol.* **11,** 2699 (2019).
66. Gabor, M. & Hotchkiss, R. D. Manifestation of linear organization in molecules of pneumococcal transforming DNA. *Proc Natl Acad Sci U S A* **56,** 1441–1448 (1966).
67. Lacks, S. Molecular fate of DNA in genetic transformation of Pneumococcus. *Journal of Molecular Biology* **5,** 119–131 (1962).
68. Piechowska, M. & Fox, M. S. Fate of transforming deoxyribonucleate in Bacillus subtilis. *J. Bacteriol.* **108,** 680–689 (1971).
69. Davidoff-Abelson, R. & Dubnau, D. Fate of transforming DNA after uptake by competent Bacillus subtilis: failure of donor DNA to replicate in a recombination-deficient recipient. *Proc Natl Acad Sci U S A* **68,** 1070–1074 (1971).
70. Fox, M. S. & ALLEN, M. K. On the Mechanism of Deoxyribonucleate Integration in Pneumococcal Transformation. *Proc Natl Acad Sci U S A* **52,** 412–419 (1964).
71. Méjean, V. & Claverys, J. P. Use of a cloned DNA fragment to analyze the fate of donor DNA in transformation of *Streptococcus pneumoniae*. *J. Bacteriol.* **158,** 1175–1178 (1984).
72. Ephrussi-Taylor, H. & Gray, T. C. Genetic studies of recombining DNA in pneumococcal transformation. *J. Gen. Physiol.* **49,** 211–231 (1966).
73. Lee, M. S., Seok, C. & Morrison, D. A. Insertion-duplication mutagenesis in *Streptococcus pneumoniae*: targeting fragment length is a critical parameter in use as a random insertion tool. *Appl Environ Microbiol* **64,** 4796–4802 (1998).
74. Marie, L. *et al.* Bacterial RadA is a DnaB-type helicase interacting with RecA to promote bidirectional D-loop extension. *Nat Commun* **8,** 15638–14 (2017).
75. Sorg, R. A., Kuipers, O. P. & Veening, J.-W. Gene expression platform for synthetic biology in the human pathogen *Streptococcus pneumoniae*. *ACS Synth Biol* **4,** 228–239 (2015).
76. Ephrussi-Taylor, H. Genetic recombination in DNA-induced transformation of Pneumococcus. IV. The pattern of transmission and phenotypic expression of high and low-efficiency donor sites in the amiA locus. *Genetics* **54,** 211–222 (1966).
77. Claverys, J. P. & Lacks, S. A. Heteroduplex deoxyribonucleic acid base mismatch repair in bacteria. *Microbiol Rev* **50,** 133–165 (1986).
78. Johnston, C., Polard, P. & Claverys, J.-P. The DpnI/DpnII pneumococcal system, defense against foreign attack without compromising genetic exchange. *Mob Genet Elements* **3,** e25582 (2013).
79. Manso, A. S. *et al.* A random six-phase switch regulates pneumococcal virulence via global epigenetic changes. *Nat Commun* **5,** 5055 (2014).
80. Liu, X. *et al.* High-throughput CRISPRi phenotyping identifies new essential genes in *Streptococcus pneumoniae*. *Mol. Syst. Biol.* **13,** 931 (2017).
81. Mortier-Barrière, I., de Saizieu, A., Claverys, J. P. & Martin, B. Competence-specific induction of recA is required for full recombination proficiency during transformation in *Streptococcus pneumoniae*. *Mol. Microbiol.* **27,** 159–170 (1998).
82. Porter, R. D. & Guild, W. R. Number of transformable units per cell in Diplococcus *pneumoniae*. *J. Bacteriol.* **97,** 1033–1035 (1969).
83. Puyet, A., Greenberg, B. & Lacks, S. A. Genetic and structural characterization of endA. A membrane-bound nuclease required for transformation of *Streptococcus pneumoniae*. *Journal of Molecular Biology* **213,** 727–738 (1990).
84. Louarn, J. M. & Sicard, A. M. Identical transformability of both strands of recipient DNA in Diplococcus *pneumoniae*. *Biochem. Biophys. Res. Commun.* **36,** 101–109 (1969).

85. Paintdakhi, A. *et al.* Oufti: an integrated software package for high-accuracy, high-throughput quantitative microscopy analysis. *Mol. Microbiol.* **99,** 767–777 (2016).
86. Cowley, L. A. *et al.* Evolution via recombination: Cell-to-cell contact facilitates larger recombination events in *Streptococcus pneumoniae*. *PLoS Genet* **14,** e1007410 (2018).
87. Hiller, N. L. & Sá-Leão, R. Puzzling Over the Pneumococcal Pangenome. *Front Microbiol* **9,** 2580 (2018).
88. Hakenbeck, R., Brückner, R., Denapaite, D. & Maurer, P. Molecular mechanisms of β-lactam resistance in *Streptococcus pneumoniae*. *Future Microbiol* **7,** 395–410 (2012).
89. Nicholas J Croucher, R. M. C. W. P. T. S. D. B. C. F. *et al.* Horizontal DNA Transfer Mechanisms of Bacteria as Weapons of Intragenomic Conflict. *PLoS Biol* **14,** e1002394–42 (2016).
90. Chewapreecha, C. *et al.* Dense genomic sampling identifies highways of pneumococcal recombination. *Nat. Genet.* **46,** 305–309 (2014).
91. Slager, J., Aprianto, R. & Veening, J.-W. Deep genome annotation of the opportunistic human pathogen *Streptococcus pneumoniae* D39. *Nucleic Acids Res.* **46,** 9971–9989 (2018).
92. de Jong, I. G., Beilharz, K., Kuipers, O. P. & Veening, J.-W. Live Cell Imaging of Bacillus subtilis and *Streptococcus pneumoniae* using Automated Time-lapse Microscopy. *J Vis Exp* (2011). doi:10.3791/3145
93. Schindelin, J. *et al*. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9,** 676–682 (2012).
94. Stylianidou, S., Brennan, C., Nissen, S. B., Kuwada, N. J. & Wiggins, P. A. SuperSegger: robust image segmentation, analysis and lineage tracking of bacterial cells. *Mol. Microbiol.* **102,** 690–700 (2016).
95. Hanahan, D., Jessee, J. & Bloom, F. R. Plasmid transformation of Escherichia coli and other bacteria. *Meth. Enzymol.* **204,** 63–113 (1991).