ORIGINAL ARTICLE

Multiplex gene transfer by genetic transformation between isolated *S. pneumoniae* cells confined in microfluidic droplets

Trinh Lam ¹, Mark Maienschein-Cline², David T. Eddington¹, and Donald A. Morrison^{3,*}

¹Department of Bioengineering, University of Illinois at Chicago, Chicago, IL, USA, ²Research Informatics Core, Research Resources Center, University of Illinois at Chicago, Chicago, IL, USA, and ³Department of Biological Sciences, University of Illinois at Chicago, Chicago, IL, USA

* Corresponding author. E-mail: damorris@uic.edu

Abstract

Gene exchange via genetic transformation makes major contributions to antibiotic resistance of the human pathogen, *Streptococcus pneumoniae* (pneumococcus). The transfers begin when a pneumococcal cell, in a transient specialized physiological state called competence, attacks and lyses another cell, takes up fragments of the liberated DNA, and integrates divergent genes into its genome. Recently, it has been demonstrated that the pneumococcal cells can be enclosed in femtoliter-scale droplets for study of the transformation mechanism, offering the ability to characterize individual cell–cell interactions and overcome the limitations of current methods involving bulk mixed cultures. To determine the relevance and reliability of this new method for study of bacterial genetic transformation, we compared recombination events occurring in 44 recombinants recovered after competence-mediated gene exchange between pairs of cells confined in femtoliter-scale droplets vs. those occurring in exchanges in parallel bulk culture mixtures. The pattern of recombination events in both contexts exhibited the hallmarks of the macro-recombination exchanges previously observed within the more complex natural contexts of biofilms and long-term evolution in the human host.

Key words: droplets; microfluidics; *Streptococcus pneumoniae*; transformation; horizontal gene transfer; whole genome sequencing

INSIGHT BOX

Streptococcus pneumoniae adapts and evolves through recombinational exchanges of genes between strains carrying ~2000-gene samples of a much larger pan-genome of the species. To model gene transfer by the mechanism of genetic transformation at the single-cell level, dispersed cells were enclosed in femtoliter-scale microfluidics droplets. Whole-genome sequencing of strains resulting from isolated interactions within the droplets provides a picture of genetic transformation as an interaction between two living cells, in which lysis of a target cell by contact with a single competent cell releases sufficient DNA to support multiple simultaneous recombination events, including some replacements as long as 80 kbp.

INTRODUCTION

The globally important bacterial human pathogen *S. pneumoniae* evolves rapidly in response to newly developed therapies and vaccine prophylaxis, primarily by inter-strain transfers of genetic material conferring antibiotic resistance and antigenic variation. Some of these transfers occur through the conjugative activity of self-mobilizing multigenic elements (known as insertional conjugative elements) such as Tn1545 [1]. Others occur via homologous recombination events mediated by the highly conserved mechanism of natural genetic transformation [2, 3].

Pneumococcal natural genetic transformation begins when a pneumococcal strain develops a transient specialized physiological state called competence. In the natural interaction, pneumococcus achieves competence by relying on a mechanism by which cells communicate at the intercellular level [4]. Competent pneumococci secrete competence stimulating peptide (CSP) via an ABC transporter (ComAB) and sense its level to initiate expression of multiple genes termed early and late competence (*com*) genes. When development of the competent state is activated, seven proteins, CbpD, CibA, CibB, CibC, ComM, LytA, and LytC act to lyse nearby non-competent cells in a process that has been proposed to regain intimate contact with the competent cells. Binding of released double-stranded DNA, uptake of single-stranded fragments by the competent cells, and homologous recombination complete the transformation process [5–9].

The non-specific genome-wide transfer mediated by genetic transformation can, in principle, support a wide evolutionary potential for adaptation via exchanges within the large pan-genome possessed by *S. pneumoniae* strains in the aggregate [10–13]. However, its role in transfer of large blocks of genes, such as the antigen-determining capsule polysaccharide synthesis cassettes, has been uncertain. The phenomenon of genetic transformation studied over more than six decades in the context of interactions of a transiently competent cell with a purified ‘donor’ DNA fragment has been understood as mediating transfer of genetic fragments of modest size, comprising typically a few genes per recombination event [14, 15].

More recently, pedigrees of numerous pneumococcal clinical isolates, reconstructed via massive whole-genome sequencing studies, have revealed historical natural exchanges encompassing a wider range of sizes, extending from below 1 kbp up to sizes above 50 kbp [16–18]. Two aspects of these exchange events led to discussion of whether some natural transfers are poorly explained by the natural transformation process as characterized *in vitro*. The unexpectedly large size of some of these inferred transfer events *in vivo* and the high frequency of multiple apparently coincident events motivated speculation that such ‘macro-recombination’ might reflect operation in nature of one or more unknown mechanisms of gene exchange transferring large blocks of DNA [18]. However, rare large transfer events have been reported during transformation experiments *in vitro* that used ‘gently prepared donor DNA’ [19]. Thus, this contrast might arise not from an undiscovered mechanism, but rather from an underappreciated aspect of *in vivo* transformation; i.e. naturally competent cells may operate *in vivo* in a way that is not modeled well in standard crosses carried out *in vitro* using donor DNA broken by hydrodynamic shear during purification. Such a natural mode of competence expression and operation is suggested by reports revealing that competent cells could lyse non-competent cells of the same or a different strain, when they are in close proximity [6, 20]. Since such interactions could provide competent cells with access to DNA from very recently

lysed ‘victims,’ the released DNA would potentially escape shear-induced breakage and might naturally extend the capability of the competence mechanism to acquisition of large contiguous tracts of divergent genetic information, as well as coincident transfers of non-contiguous tracts.

To begin to describe the genetic outcomes from experimental models of cell–cell lytic interactions that might better mimic exchange events occurring *in vivo*, a recent report compared, at the whole-genome scale, recombination events emerging from a classic ‘DNA + cells’ paradigm with those recovered from interactions between competent cells and nearby non-competent cells within the very dense cell assemblages of in biofilms or biofilm-like conditions [21]. A dramatically increased frequency of long recombination tracts in the biofilm mimics suggested that an explanation of macro-recombination occurring *in vivo* may indeed lie in DNA transfer via direct cell–cell interactions.

To define more clearly the range of outcomes of elementary cell–cell interactions, we recently developed a microfluidic device for confined isolation of pairs of cells under conditions where individual lytic events might be isolated for direct observation *in vitro* [22]. Because experimental conditions in microfluidic devices are so different from those in bulk cultures or *in vivo*, it is always a concern that a phenomenon of interest that has been recorded in bulk cultures may ‘evaporate’ once cells are confined within aqueous droplets in an oil–water–surfactant emulsion. Pneumococcal transformation in particular has long been known to be especially sensitive to environmental fluctuations, and pneumococcus itself is subject to, for example, detergent-induced autolysis. Thus, it is unknown whether the process of encapsulating cells in oil–water–surfactant emulsions could provide a suitable environment fully representative either of the standard culture media or of the natural environment *in vivo*. In the study described here, we sought to learn, by sequencing recombinants at the whole-genome scale, how well such isolated and directly observable lytic episodes may mimic the interactions previously documented for mixed bulk cultures of unconfined cells, where the specific cells participating in a single exchange were, by the experimental design, inherently unknowable. We found that the new droplets context yields outcomes similar to unconfined cells, validating the droplet method for the study of bacterial genetic transformation. The results show further that products of interactions between dispersed cells, both confined and unconfined, retain the hallmarks of macro-recombination. Thus, all classes of recombination events thought to occur in the natural host can now be accessible at the level of single observable competent cells, using this improved model of the natural process.

MATERIALS AND METHODS

Strains and media

Two previously described [21] pneumococcal strains, CP2204 (Rif^RNov^SSpc^SComA[−]ComE⁺) and CP2215 (Rif^RNov^RSpc^RComE[−]), were grown in chemically defined medium (CDM) [23, 24] with 1% casein hydrolysate tryptone medium (CAT) [25] to an optical density (OD₅₅₀) of 0.2, mixed with 2 ml of 100% glycerol per 12 ml culture and stored at −80°C. The strain CP2204, dependent on exogenous CSP to activate competence, is designated as DNA recipient or attacker. CP2215, lacking *comE*, was the non-competent DNA donor or target.

Competence of strain CP2204 was activated by an inducer cocktail consisting of CSP, bovine serum albumin (BSA), CaCl₂,

and CDM. CSP-1 [26] (NeoBiolab) was dissolved in distilled water to a concentration of 100 µg/ml, filtered through a 0.2 µm syringe filter (Sartorius), and stored at -20°C until use. BSA (Gold Biotechnology) was dissolved at 40 mg/mL in distilled water, sterilized, and stored at 4°C. The BSA stock was diluted to 4 mg/ml for short-term use in inducer cocktail. A sterile 0.1 M stock of CaCl₂ was diluted to 0.01 M for use in inducer cocktail. Inducer cocktail was prepared by mixing 28 µL of 100 µg/ml CSP, 14 µL of 4 mg/ml BSA, 70 µL of 0.01 M CaCl₂, and 588 µL of CDM.

Todd-Hewitt broth (Becton Dickinson and Company) supplemented with 2% yeast extract (THY) was used with 1.5% agar for selective plating [27] in 60 × 15 mm petri dishes (Falcon). For selection, the fourth layer of each plate consisted of antibiotics, either rifampicin (40 µg/ml) and novobiocin (10 µg/ml) (RN) or rifampicin (40 µg/ml) and spectinomycin (160 µg/ml) (RS). All plates were incubated for 2 days at 37°C. Transformation experiments were replicated four times. Recombinants' colonies from the fourth experiment (Supplementary Table S1, highlighted in blue) were used for sequencing.

Transformation experiment

Before transformation experiments, CP2204 and CP2215 were grown separately in CDM with 1% CAT and maintained at 37°C until reaching an OD of 0.25. Both strains were then chilled and centrifuged at 7000 × *g* for 7 min in a 4°C centrifuge (Eppendorf model 5804R). The pellets were then resuspended in cold CDM with 1% CAT, removing residues from dead cells and achieving a final OD of 4 for the experiments [22]. The two suspensions were mixed in a volume ratio of 1:1 and maintained on ice until use.

Transformation experiments were done in both confined (droplets) and unconfined environments. Previously described three-inlet flow-focusing microfluidic devices (Fig. 1) were used to generate 10 µm diameter aqueous droplets for cell encapsulation. The flow-focusing microfluidic devices had two aqueous inlets for inducer cocktail and bacterial cell mixture, respectively, plus one inlet for FC-40 with 5% surfactant, known as Pico-surf (Dolomite). The droplet-generating system was under control of three syringe pumps (Chemyx Inc); one operated at a speed of ~250 µL/h for delivering of Pico-surf and the others maintained at ~70 µL/h. The separate aqueous phase streams of inducer and cell mixture, which had a relatively low flowrate, were combined and forced through a 10 µm width narrow neck, into the outer Pico-surf stream, where pressure and viscous stresses break the mixed aqueous stream into uniform droplets on emerging from the narrowed orifice [28].

The droplet-generating process was conducted in a 4°C cold-room to maintain the concentrated cells dormant and postpone activation of the competence response in CP2204. Once a stable flow of droplets formed, a 20 cm long PTFE tube (~1 mm ID) was used to connect the outlet of the droplet generator to a vial on ice containing 300 µL of CAT, to prevent evaporation from the droplets. Droplets were collected for 30 min. Simultaneously, a 1:1 mixture of the cell mixture and inducer in a total volume similar to the total volume of collected aqueous droplets was maintained on ice in the cold room while droplets were collected.

Both the droplets emulsion and the unconfined cultures were then warmed to 37°C for activation of competence. After 25 min of incubation, the droplet emulsion was broken with Pico-break as described previously [22]. The droplet vial was first centrifuged at 100 × *g* for 30 s and the bottom layer of Pico-surf was pipetted out for disposal. About 200 µL of Pico-break (Dolomite) was added to the remaining droplets, mixed by

Figure 1. Droplet microfluidic device for bacterial cell encapsulation. (A) Microfluidic droplet generator. 10 µm diameter droplets were generated by flow-focusing of the three streams: Pico-surf (5% surfactant in FC-40), inducer, and bacterial cell mixture. Close-ups show flow-focusing nozzle (left) and image of cell encapsulating in droplets taken in a separate imaging chamber (right). (B) Detailed design of the droplet generator and the flow-focusing nozzle. Insets show 10 µm and 6 µm filters placed at inlets of oil and aqueous streams to prevent debris from blocking channels.

inversion, and centrifuged at 1000 × *g* for 1 min. The upper phase in the vial, containing released cells, was measured and pipetted out. Both the control and the droplet vial were then diluted 1000-fold into CAT before 1-h incubation at 37°C, followed by further dilution as needed, and selection of RN or RS recombinants, three plates for each dilution and for each selection. For cell viability, cultures were diluted 10⁵- and 10⁶-fold and plated with R, N, and S, in duplicate plates for each marker and each dilution factor.

DNA preparation and sequencing

Thirty recombinant colonies from each selection (RN and RS) of droplets and liquid attack experiments were isolated by stabbing with a sterile straight loop and transferred to 96-well plates with 150 µL of THY in each well. After 5 h at 37°C with 5% CO₂, cultures were streaked on blood agar plates containing Tryptone Soya Agar with 5% sheep blood (ThermoFisher Scientific). After 16 h at 37°C with 5% CO₂, isolated colonies were picked and transferred to 2 ml of CDM. To verify antibiotic resistance phenotypes of the cultures, the colonies were also transferred to agar plates with novobiocin (2 µg/ml) or with spectinomycin (40 µg/ml). Selective plates were incubated overnight at 37°C. The 2 ml CDM cultures were incubated at 30°C for 15 h and brought to 37°C

for the next 9 h. As the 2 ml cultures reached visible turbidity (OD of 0.1–0.4), they were adjusted to 12% glycerol and stored at -80°C until use for DNA purification. Their genomic DNA was purified according to the manufacturer's protocol (Quick-DNA Fungal/Bacterial Microprep kit, Zymo). Yields were $\sim 1\text{--}5\text{ }\mu\text{g}$ DNA per 15 ml THY culture at OD 0.25–0.3. About 88 DNA samples were processed at the UIC Sequencing Core Center for whole genome sequencing (WGS) by Illumina NextSeq with $2 \times 150\text{ bp}$ sequencing reads and with each recombinant sequenced to an average coverage of 300X. Raw sequencing data for cell + DNA and filter was adopted from the previous study [21] and mapped under the new mapping algorithm.

Mapping of recombination events

In this analysis, the reference genome was set as CP2215 (NCBI Reference Sequence: NZ_CP028436.1), unless otherwise stated. The two parental genomes, of strains CP2215 and CP2204, differ by 15 729 Single nucleotide polymorphisms (SNPs). The analysis of recombination events was accomplished through two steps: (1) generation of a consensus fasta file incorporating SNPs from recombinant strains, and (2) comparison of the consensus fasta files between recombinant and parent strains to determine the parental origin of individual SNPs and SNP blocks. These steps were incorporated into two scripts: (1) *Fasta_consensus* (Available at bit.ly/fasta_consensus) and (2) *Compare_fastas* (Available at bit.ly/compare_fastas). Additional software BWA MEM [29], Samtools [30], and Bcftools [31] were required for the *fasta_consensus* script. The algorithms behind these two steps are described more detail below.

Step 1: Generation of consensus fasta file (*Fasta_consensus* script). A new fasta file for either another parent, in this case CP2204, or the recombinant genome, was based on the finished assembly of the donor strain CP2215 ($\sim 2000\text{ kbp}$). Raw reads were aligned to the CP2215 reference genome using BWA MEM [32], and alignments were sorted, and apparent PCR duplicates were removed by Samtools [30]. Variant calling was performed using Samtools mpileup flags followed by Bcftools call [31] with the consensus caller algorithm (flag `-c`), and output as Variant Call Format (VCF) with genotype calls at every position in the genome, regardless of the presence of a variant. Base substitutions in the original reference genome were made based on the single nucleotide variants, and coverage and quality levels per nucleotide were assessed based on the following criteria: (1) sum of base quality (QUAL) > 40 , (2) depth (DP) > 5 , (3) mapping quality > 30 , and (4) allele frequency > 0.95 or < 0.05 . Positions that did not pass these quality checks were replaced with 'N.' The result was a new fasta file of equal length to the reference, but with high-confidence SNPs replaced with the allele present in the recombinant and low-quality bases masked with Ns. Details of quality filtering are available at bit.ly/make_pseudosequence.

Step 2: Comparison of consensus fasta files (*Compare_fastas* script). Fasta file of a recombinant strain was compared to two parental genomes, one parent as a donor and the other as a recipient. This analysis identified regions of the recipient that were derived from the donor parent. First, a list was compiled of single base positions where the donor matched the recombinant but not the recipient, and none of the donor or either parents' alleles were masked—this list was output as a position of donor-derived SNPs. To assemble putative genomic blocks that were derived from the donor, a second list was compiled of all unmasked positions where the recipient did not match the donor. The coordinates of each donor-derived SNP were then expanded based on this second list to the maximum region

where the recombinant consistently matched the donor. This result was output as a list of donor-derived genomic regions (start/end sites).

Integrated genome viewer (IGV) was used to visualize SNPs between two parental strains and the recombinants, and DNA tract and block regions. Circos [33] plot was used to visualize the genomic distribution of donor SNPs in recombinants' genomes.

RESULTS

Gene transfer from non-competent cells to competent cells in laboratory biofilms, or in very high density artificial transient but dense ($> 10^{12}$ cells/ml) assemblages [21], exhibits the characteristics of macro-recombination events that have been inferred to accumulate during pneumococcal evolution and adaptation in the human host [18]. To learn whether such complex exchange events might be inherent in individual cell–cell interactions, and not restricted to transfers within high-density populations, we characterized exchange events between much more highly dispersed cells (e.g. in mixed liquid suspension cultures, at $\sim 10^9$ cells/ml, as well as in complete isolation within femtoliter scale aqueous droplets). To explore whether physical confinement within microdroplets affects the nature of the recombination events arising at these low densities, a single mixture of a recipient and a donor strain was subjected to CSP-induced competence development and gene exchange, both as confined cells and as a parallel unconfined mixture.

Competence development by the recipient was dependent on supplemental CSP, as it carries a defective CSP-secretion allele (*comA::erm*), but the donor strain was competence-negative due to a deletion of the CSP receptor gene, *comE*, as detailed in 'Materials and Methods.' The two strains also differed in three selective markers (Nov^R and Spc^R of the donor strain and Rif^R of the competent recipient strain). Thus, gene transfer following competence induction could be estimated by appearance of Rif^RNov^R and Rif^RSpc^R recombinants, each of which would necessarily reflect lytic attack by the Rif^R competent cells upon the non-competent Nov^RSpc^R donors. To achieve a useful incidence of droplets containing donor-recipient pairs, exponential-phase cultures were concentrated by centrifugation in the cold, followed by resuspension in a chemically defined medium suitable both for droplet stabilization and for development of competence, at $\sim 10^9$ cells/ml, as previously described [22]. Competence develops rapidly after exposure to CSP [34], and gene-transfer events are already abundant by 25 min at 37°C .

Gene transfer during both classes of lytic interactions was initially estimated by counting recombinants expressing both the recipient Rif^R marker and either the Nov^R point marker or the Spc^R indel marker from the non-competent partner. Similar rates of transfer were observed in the two formats (Supplementary Table S1). A somewhat lower rate of transfer within droplets can be accounted for by the fact that encapsulation necessarily places over-half of the competent cells alone within droplets, precluding access to a target cell. Pico-surf (5% surfactant in FC-40) and Pico-break solutions, which were used in droplet generation and emulsion breaking, themselves also reduce transformant rates slightly as reported previously [22], but this effect was not statistically significant.

To characterize the recombination events occurring during these interactions more thoroughly and with a genomic perspective, we took advantage of the 15 729 SNPs distinguishing the two strains. Individual Nov^R and Spc^R transformants were isolated by selective plating and purified by single-colony isolation. DNA isolated from 88 independent recombinants was sequenced on

the Illumina platform at ~300X coverage. A schematic of the experimental workflow is shown in [Supplementary Fig. S1](#). By aligning sequence reads from each recombinant clone to the complete donor and recipient genomes, the SNPs distinguishing the interacting strains were used to map the positions and extents of recombination events, as described in 'Methods' and as illustrated in [Supplementary Fig. S2](#), showing alignments of sequence reads and use of the SNP markers to define limits of two exemplary recombination events.

The recombination events mapped for each recombinant demonstrated, as expected, transfer of the selected marker (1 divergent bp for Nov^R or a 753 bp replacement for Spc^R) ([Fig. 2](#) and [Supplementary Fig. S2](#)). These selected transfers were embedded within larger blocks of incorporated donor sequences, ranging in size from 2 bp to 37 167 bp, as expected. Some of the selected transfers were further embedded within a tract of two or three nearby recombinant blocks ([Supplementary Figs S3–S6](#)). Beyond these directly selected replacements, the recombinant genomes exhibited an average of ~3 additional distant recombination events, distributed broadly throughout the genome ([Fig. 3](#) and [Supplementary Table S2](#)). Across the entire genomes of recombinant strains, complex recombination tracts ranged in overall length from 858 to 89 120 bp ([Supplementary Figs S3–S6](#)).

The size distributions of recombination 'blocks' are compared ([Fig. 4](#)) in the four classes of recovered transformants—confined vs. unconfined cross formats and Nov vs. Spc selections. Comparisons are displayed as a function of event size for number of events ([Fig. 4A](#)), for total bp of information transferred by events in each event size class ([Fig. 4B](#)), as well as for the cumulative information transfer integrated from the smallest to the largest events ([Fig. 4C and D](#)). The mean recombination event size was slightly greater in droplets than that in liquid attack (2874 bp in DN, 4214 bp in DS, 1899 bp in LN, and 3564 in LS); however, the variance in this measure was so great that the differences in average size are not statistically significant. The total number of events per recombinant strain was also similar between the two formats of encounter ([Supplementary Table S2](#)). For both the Nov^R and the Spc^R selected recombinants, inspection of [Fig. 4C](#) shows that although the amount of DNA transferred per strain in small events was similar in all contexts, the transfer in larger (>10 kbp) blocks was greater for the droplet context than for the liquid mixtures. We conclude that the two cell–cell encounter formats have remarkably similar recombination outcomes, with an important proportion (~30%) of information transfer represented by blocks in sizes of 10–23 kbp.

Access to long donor DNA fragments can lead, in a single uptake event, to a cluster (tract) of recombination events (blocks), due either to interruption of strand displacement reactions during recombination or to distribution of parts of the donor strand to distinct sister chromosomes within a single competent cell, as illustrated previously by several examples of such events in transformation of the competent Gram-negative species, *Haemophilus influenzae* [35, 36]. Thus, genetic linkage conferred by transformation will be determined not just by transfer of large blocks, but also by the transfer of tracts of adjacent recombinant blocks. Because such extended events would be a significant source of local marker linkage, it is interesting to inspect the sizes of recombination tracts, repeating the comparison of cell–cell interaction between confined and unconfined cell mixtures. To facilitate this comparison, we grouped adjacent blocks into local tracts when blocks of donor sequence occurred within 100 kbp of each other and occupied at least 40% of a local region [35, 36]. For the comparison shown, singular

blocks, with no nearby blocks, are considered as tracts with 100% occupancy. [Supplementary Fig. S7](#) illustrates the combined contributions to information transfer of recombination tracts in recombinants recovered from confined vs. unconfined cultures. When accounting adjacent blocks as local tracts, recombination within both droplets and unconfined cultures also had similar outcomes. There were slightly fewer events per strain in the droplet confrontations, but the amount of transfer in each size class within droplets was equivalent to that in the liquid attack. In both contexts, tracts of lengths 30–90 kbp made significant contributions (~30%) to overall transfer of information.

Taken together, these mapping results indicate that both distinguishing features of macro-recombination—multiple coincident transfers and transfers of large blocks or tracts—are displayed in the outcomes of short-term cell–cell competence interactions occurring at low cell densities, in either dilute strain mixtures or between members of completely isolated pairs of cells. Thus, we conclude that confined cell pairs recapitulate, on a stage suitable for direct observation, the complex features of gene transfer events previously noted as occurring in dense mixed cultures and *in vivo*.

DISCUSSION

Confinement of bacteria within micro-scale microfluidic droplets is, in principle, an attractive alternative to supplement standard microbiological bulk-culture contexts, especially when cell–cell interactions are of interest, because they can allow direct observation and analysis at the level of individual cells. In the case of genetic transformation, only recently understood to occur in the context of cell–cell attack followed by incorporation of released DNA, we previously described design and operation of a device for confinement of such interacting pairs of cells within miniature aqueous droplets and showed that competence-dependent gene transfer occurred therein. The present results indicate that such transfers between confined cell pairs are suitable models of the natural gene-transfer interactions, because the properties of recombination events were indistinguishable between confined pairs or unconfined cell mixtures.

Macro-recombination was first described as a class of gene-transfer event inferred from close examination of whole-genome sequences of isolates from many widespread pneumococcal clones, representing evolution over periods of years, or decades [18]. The authors expressed surprise that some exchanges inferred to have happened *in vivo* appeared to be qualitatively distinct from those typical of genetic transformation *in vitro* and discussed the possibility that a different exchange mechanism might be operating during adaptation and evolution *in vivo*. More recently, Cowley *et al.* [21] explored the notion that macro-recombination might arise through genetic transformation when it operates in contexts of very high cell density such as found in natural biofilms. Comparing outcomes of recombination *in vitro* within artificial biofilms (over 2 days) or within artificial short-term high density mixed cell assemblages (10¹² cells/ml, over 40 min) with standard transformation of competent cells by purified genomic DNA, they reported both features of macro-recombination were characteristic of exchanges occurring at high cell densities *in vitro*, concluding that genetic transformation may indeed be a principal gene transfer mechanism operating *in vivo*, and that either cell–cell attack itself, or the occurrence of such interactions in high-density milieu could be at the root of the macro-recombination pattern. The present results, showing that both distinguishing

Figure 2. Recombination events mapped near selected markers in recombinants recovered from droplets (D) or liquid attack (L). (A) DN, (B) DS, (C) LN, and (D) LS. Each panel shows the 175 kbp region including the *Nov^R* marker (*gyrB*) at 770 kbp or *Spc^R* at 2120 kbp. Map tracks are ordered by the size of the single block passing through the indicated marker region (*Nov^R* or *Spc^R*). Arrows indicate genome positions of *gyrB* (*Nov^R*) or *Spc^R* markers. In each IGV panel, from top to bottom, tracks represent: (1) Genome coordinates in 175 kbp region of CP2215. Position of bp 1 indicated by red bar on genomes coordinates; (2) Gene: Annotated genes and selected gene names in CP2215; (3). Red and blue bars: Recombination blocks in genomes of recombinant strains listed on the left side of each track. *Nov* selection indicated by 'N'; *Spc* selection by 'S.'

features of macro-recombination also accompany CSP-induced interactions between dilute or isolated cells, indicate that the key context leading to macro-recombination is the interaction of a competent cell with DNA from a recently lysed target cell, rather than high cell density *per se*.

With availability of genome-scale maps of recombination events arising during interactions recorded over a 1000-fold range in cell density, it is interesting to compare the outcomes quantitatively. Outcomes of recombination events in the various *in vitro* models of two cell-cell encounter formats such as artificial cell assemblage at 10^{12} cells/ml, diluted cell cultures (liquid attack), isolated cell pairs in droplets, and interaction of competent cells and isolated DNA fragments are further compared in terms of blocks and tracts (including one-block tract). The comparison shows that even in the dispersed-cell format, cell-cell interactions yield an extended range of sizes of recombination events, as compared to standard cells + DNA interactions as shown in [Supplementary Table S2](#). While in the two formats used by Cowley *et al.*, competent cells + saturating DNA and dense cell mats, the number of recombination events per

transformed cell was essentially the same, the size of replacements was much larger in the latter format. At the much lower cell densities used in this study to examine cell-cell interactions, the numbers of recombination events per recipient cell were reduced by ~50%, but event sizes were larger than in the cells + DNA model, by 2-fold up to 5-fold.

Retrospective analyses of recombination events in the natural human host suffer necessarily from the filter of selection that they must pass through before becoming fixed and recovered from circulating strains for analysis [16, 37–39]. Because the time, place, and participants of inferred recombination events are unknown, it has been difficult to distinguish the direct recombination products of transformation from patterns sculpted by selection during their survival and spread in natural hosts. In WGS surveys, if lineages accumulating exchanges over periods up to decades [16, 38–40], the timing of exchange was virtually unknown. The time of exchanges was significantly more circumscribed in the interesting case of exchanges among 2–3 strains over a period of 7 months within a single pediatric patient [17], but the question whether macro-recombination

Figure 3. Genomic distribution of recombination events mapped in *Nov^R* and *Spc^R* transformants recovered from cell-cell interactions. Circos plots visualize mapped recombination events in droplets (D) and liquid attack (L). *Nov* selection is shown by 'N' and *Spc* selection is shown by 'S.' (A) DN, (B) DS, (C) LN, and (D) LS. Outer ring represents coordinates of CP2215 in bp starting from *dnaA*, with red marks showing selective markers (*Nov^R* at 770 kbp and *Spc^R* at 2120 kbp). Red lines in the first inner ring represent SNPs in CP2204 (recipient strain or attacker) that differ from CP2215 (donor strain or victim). Subsequent rings indicate recombination events in genomes of indicated transformants. Numbers in each inner ring indicate the transformant strain number. Black blocks in the inner rings represent recombination blocks.

reflected multiple consecutive exchanges or multiple exchanges within a single competence episode remains open. Recent observations of the macro-recombination pattern of exchanges occurring *in vitro* within a single 40–60 min competence episode in dense cell mixtures further sharpen the period of time for recombination events. They left open, however, the question of whether macro-recombination is inherent in the fundamental cell-cell competence interaction, or whether it arises only in multicellular environments. The present results indicate that

macro-recombination is not restricted to high-density populations but is the consequence of elementary cell attack/lysis/DNA uptake events characteristic of the natural transformation mechanism. Repeated retrieval of long blocks/tracts in recombination events from pneumococcal competence episodes suggests that this mode of exchange has a relaxed inherent limit on transfer. Once uptake has been initiated, there is no inherent mechanistically defined end of transferred material, except reaching an end of a DNA fragment. However,

Figure 4. Contribution of recombination blocks to total information exchange. Recombination events mapped throughout genomes of recombinants recovered from each condition were sorted by length (in 2 kbp increments) and aggregated to show average values per transformant strain. (A) Number of recombination events (blocks) in each size class. (B) Total amount of DNA transferred by events in each size class. (C) Cumulative amount of DNA incorporated by recipients, calculated by integrating the amount of DNA transferred by events up to the indicated event size. (D) Normalized cumulative transfer. DN and DS indicate recombinants recovered from confined droplets for Nov^R marker and Spc^R marker, respectively. LN and LS indicate recombinants recovered from liquid attack or unconfined culture for Nov^R marker and Spc^R marker, respectively.

regulation of competence does impose a self-limiting aspect, created mainly by activity of DprA in abrogating CSP sensitivity. This self-imposed limit to the period of competence may be the practical limiting factor in the length of transferred DNA.

CONCLUSIONS

We analyze and compare whole genomes of descendants of recombination events between two divergent parental strains in unconfined and confined contexts. This study demonstrates that droplet microfluidics can provide a new method to study bacterial genetic transformation at the level of individual cells. The droplet and bulk unconfined culture methods support similar gene-transfer rates and yields of recombination events, and the recombinants recovered from both methods shared similar distributions of recombination events at the genomic level, both in total amount of gene transfer and in the size range of individual events. Complex tracts of recombination, ranging from ~1000 to 80 000 bp, transfer multiple genes during single interactions in both contexts. In future studies, we expect improvements in our device design will allow real-time imaging of cell-cell attack and genetic transfer in individual droplets. Then, the descendant of interacting competent cells from a

specific droplet could be identified and recovered and resulting recombination events can be mapped by whole-genome sequencing. Investigating and understanding the gene transfer of pneumococcal strains at the whole-genome level offers an ability to improve study models for observing the full and precise mechanism of bacterial transformation as a cell-cell interaction, which can be critical to understanding the sources and evolution of antibiotics resistance in this common human pathogen.

SUPPLEMENTARY DATA

Supplementary data are available at *INTBIO Journal* online.

FUNDING

This work was supported by National Institute of Allergy and Infectious Diseases [R21AI133304]. Mark Maienschein-Cline was supported in part by NCATS through Grant UL1TR002003.

CONFLICTS OF INTEREST STATEMENT

There are no conflicts to declare.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Dr Stefan J. Green of Sequencing Core Facility of University of Illinois at Chicago (UIC) for support of WGS.

REFERENCES

- Seral C, Castillo FJ, Rubio-Calvo MC et al. Distribution of resistance genes tet(M), aph3'-III, catpC194 and the integrase gene of Tn1545 in clinical *Streptococcus pneumoniae* harbouring erm(B) and mef(A) genes in Spain. *J Antimicrob Chemother* 2001;47:863-6.
- Johnston C, Campo N, Berge MJ et al. *Streptococcus pneumoniae*, le transformiste. *Trends Microbiol* 2014;22:113-9.
- Li G, Liang Z, Wang X et al. Addition of hypertransformable pneumococcal isolates to natural transformation for in vivo fitness and virulence. *Infect Immun* 2016;84:1887-901.
- Håvarstein LS. Quorum sensing and peptide pheromones in streptococcal competence for genetic transformation. In: Dunne GM (ed.). *Cell-Cell Signaling in Bacteria*. Washington, DC, ASM Press, 1999, 9-26.
- Peterson SN, Sung CK, Cline R et al. Identification of competence pheromone responsive genes in *Streptococcus pneumoniae* by use of DNA microarrays. *Mol Microbiol* 2004;51:1051-70.
- Guiral S, Mitchell TJ, Martin B et al. Competence-programmed predation of noncompetent cells in the human pathogen *Streptococcus pneumoniae*: Genetic requirements. *Proc Natl Acad Sci U S A* 2005;102:8710.
- Håvarstein LS, Martin B, Johnsborg O, Granadel C, Claverys J-P. New insights into the pneumococcal fratricide: relationship to clumping and identification of a novel immunity factor. *Molecular microbiology* 2006;59:1297-1037.
- Claverys JP, Havarstein LS. Cannibalism and fratricide: mechanisms and raisons d'être. *Nat Rev Microbiol* 2007;5:219-29.
- Eldholm V, Johnsborg O, Haugen K et al. Fratricide in *Streptococcus pneumoniae*: contributions and role of the cell wall hydrolases CbpD, LytA and LytC. *Microbiology* 2009;155:2223-34.
- Croucher NJ, Finkelstein JA, Pelton SI et al. Population genomics of post-vaccine changes in pneumococcal epidemiology. *Nat Genet* 2013;45:656.
- Donati C, Hiller NL, Tettelin H et al. Structure and dynamics of the pan-genome of *Streptococcus pneumoniae* and closely related species. *Genome Biol* 2010;11:R107.
- Hiller NL, Sá-Leão R. Puzzling over the pneumococcal Pangenome. *Front Microbiol* 2018;9.
- van Tonder AJ, Bray JE, Jolley KA et al. Heterogeneity among estimates of the core genome and pan-genome in different pneumococcal populations. *bioRxiv* 2017. doi: 10.1101/133991:133991.
- Gurney T, Fox MS. Physical and genetic hybrids formed in bacterial transformation. *J Mol Biol* 1968;32:83-100.
- Croucher NJ, Harris SR, Barquist L et al. A high-resolution view of genome-wide pneumococcal transformation. *PLoS Pathog* 2012;8:e1002745.
- Chewapreecha C, Harris SR, Croucher NJ et al. Dense genomic sampling identifies highways of pneumococcal recombination. *Nat Genet* 2014;46:305-9.
- Hiller NL, Ahmed A, Powell E et al. Generation of genic diversity among *Streptococcus pneumoniae* strains via horizontal gene transfer during a chronic polyclonal pediatric infection. *PLoS Pathog* 2010;6:e1001108.
- Mostowy R, Croucher NJ, Hanage WP et al. Heterogeneity in the frequency and characteristics of homologous recombination in pneumococcal evolution. *PLoS Genet* 2014;10:e1004300.
- Trzcinski K, Thompson CM, Lipsitch M. Single-step capsular transformation and acquisition of penicillin resistance in *Streptococcus pneumoniae*. *J Bacteriol* 2004;186:3447-52.
- Steinmoen H, Teigen A, Havarstein LS. Competence-induced cells of *Streptococcus pneumoniae* lyse competence-deficient cells of the same strain during cocultivation. *J Bacteriol* 2003;185:7176-83.
- Cowley LA, Petersen FC, Junges R et al. Evolution via recombination: cell-to-cell contact facilitates larger recombination events in *Streptococcus pneumoniae*. *PLoS Genet* 2018;14:e1007410.
- Lam T, Brennan MD, Morrison DA et al. Femtoliter droplet confinement of *Streptococcus pneumoniae*: bacterial genetic transformation by cell-cell interaction in droplets. *Lab Chip* 2019;19:682-92.
- Chang JC, LaSarre B, Jimenez JC et al. Two group A streptococcal peptide pheromones act through opposing Rgg regulators to control biofilm development. *PLoS Pathog* 2011;7:e1002190.
- van de Rijn I, Kessler RE. Growth characteristics of group A streptococci in a new chemically defined medium. *Infect Immun* 1980;27:444-8.
- Tovpeko Y, Bai J, Morrison DA. Competence for genetic transformation in *Streptococcus pneumoniae*: mutations in sigmaA bypass the ComW requirement for late gene expression. *J Bacteriol* 2016;198:2370-8.
- Pozzi G, Masala L, Iannelli F et al. Competence for genetic transformation in encapsulated strains of *Streptococcus pneumoniae*: two allelic variants of the peptide pheromone. *J Bacteriol* 1996;178:6087-90.
- Morrison DA. Transformation and preservation of competent bacterial cells by freezing. *Methods Enzymol* 1979;68:326-31.
- Anna SL, Bontoux N, Stone HA. Formation of dispersions using "flow focusing" in microchannels. *Appl Phys Lett* 2003;82:364-6.
- Li H. *Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM*. arXiv e-prints [Internet]. March 01, 2013. Available from: <https://ui.adsabs.harvard.edu/abs/2013arXiv1303.3997L>.
- Li H, Handsaker B, Wysoker A et al. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25:2078-9.
- Danecek P, Auton A, Abecasis G et al. The variant call format and VCFtools. *Bioinformatics* 2011;27:2156-8.
- Li H. *Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM*. arXiv 2013; 1303.3997v1 [q-bio.GN].
- Krzywinski M, Schein J, Birol I et al. Circos: an information aesthetic for comparative genomics. *Genome Res* 2009;19:1639-1645.
- Havarstein LS, Coomaraswamy G, Morrison DA. An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus pneumoniae*. *Proc Natl Acad Sci U S A* 1995;92:11140-4.

35. Mell JC, Lee JY, Firme M et al. Extensive cotransformation of natural variation into chromosomes of naturally competent *Haemophilus influenzae*. *G3 (Bethesda)* 2014;**4**: 717–31.
36. Mell JC, Shumilina S, Hall IM et al. Transformation of natural genetic variation into *Haemophilus influenzae* genomes. *PLoS Pathog* 2011;**7**:e1002151-1.
37. Chaguza C, Cornick JE, Everett DB. Mechanisms and impact of genetic recombination in the evolution of *Streptococcus pneumoniae*. *Comput Struct Biotechnol J* 2015;**13**: 241–7.
38. Croucher NJ, Harris SR, Fraser C et al. Rapid pneumococcal evolution in response to clinical interventions. *Science (New York, NY)* 2011;**331**:430–4.
39. Croucher NJ, Mitchell AM, Gould KA et al. Dominant role of nucleotide substitution in the diversification of serotype 3 pneumococci over decades and during a single infection. *PLoS Genet* 2013;**9**:e1003868.
40. Golubchik T, Brueggemann AB, Street T et al. Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event. *Nat Genet* 2012;**44**:352–5.