

# A genomic view of experimental intraspecies and interspecies transformation of a rifampicin-resistance allele into *Neisseria meningitidis*

Kristian Alfsnes,<sup>1,\*</sup> Stephan A. Frye,<sup>2</sup> Jens Eriksson,<sup>2</sup> Vegard Eldholm,<sup>3</sup> Ola Brønstad Brynildsrud,<sup>4</sup> Jon Bohlin,<sup>4</sup> Odile B. Harrison,<sup>5</sup> Derek W. Hood,<sup>6</sup> Martin C. J. Maiden,<sup>5</sup> Tone Tønjum<sup>2,7</sup> and Ole Herman Ambur<sup>2,8</sup>

## Abstract

The spread of antibiotic resistance within and between different bacterial populations is a major health problem on a global scale. The identification of genetic transformation in genomic data from *Neisseria meningitidis*, the meningococcus (Mc), and other bacteria is problematic, since similar or even identical alleles may be involved. A particular challenge in naturally transformable bacteria generally is to distinguish between common ancestry and true recombined sites in sampled genome sequences. Furthermore, the identification of recombination following experimental transformation of homologous alleles requires identifiable differences between donor and recipient, which in itself influences the propensity for homologous recombination (HR). This study identifies the distribution of HR events following intraspecies and interspecies Mc transformations of *rpoB* alleles encoding rifampicin resistance by whole-genome DNA sequencing and single nucleotide variant analysis. The HR events analysed were confined to the genomic region surrounding the single nucleotide genetic marker used for selection. An exponential length distribution of these recombined events was found, ranging from a few nucleotides to about 72 kb stretches. The lengths of imported sequences were on average found to be longer following experimental transformation of the recipient with genomic DNA from an intraspecies versus an interspecies donor ( $P < 0.001$ ). The recombination events were generally observed to be mosaic, with donor sequences interspersed with recipient sequence. Here, we present four models to explain these observations, by fragmentation of the transformed DNA, by interruptions of the recombination mechanism, by secondary recombination of endogenous self-DNA, or by repair/replication mechanisms.

## DATA SUMMARY

Sequence data has been deposited in the European Nucleotide Archive (ENA) under accession number PRJEB27515.

## INTRODUCTION

In addition to replication and transcription, homologous recombination (HR) is one of the most fundamental genetic processes in nature conserved across all domains of life. The transfer of genes by horizontal gene transfer (HGT), through transduction, conjugation or transformation allows

bacteria of different species to share traits (e.g. antimicrobial resistance and virulence) on a very short timescale compared to random mutation and natural selection [1].

The opportunistic pathogenic bacterium, *Neisseria meningitidis* (the meningococcus; Mc) readily takes up and incorporates extracellular DNA into its chromosome through transformation [2, 3]. In contrast to many Gram-negative bacteria, Mc is competent for transformation throughout all growth phases [4] and does not depend on induction by environmental cues (reviewed by Seitz and Blokesch [5]).

Received 9 July 2018; Accepted 31 August 2018

**Author affiliations:** <sup>1</sup>Nasjonalt Folkehelseinstitutt, Oslo, Norway; <sup>2</sup>Department of Microbiology, Oslo University Hospital (Rikshospitalet), Oslo, Norway; <sup>3</sup>Department of Molecular Biology, Domain of Infection Control and Environmental Health, Norwegian Institute of Public Health, Oslo, Norway; <sup>4</sup>Department of Methodology Research and Analysis, Domain of Infection Control and Environmental Health, Norwegian Institute of Public Health, Oslo, Norway; <sup>5</sup>The Peter Medawar Building for Pathogen Research, University of Oxford, Oxford, UK; <sup>6</sup>Nuffield Department of Clinical Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford, UK; <sup>7</sup>Department of Microbiology, University of Oslo, Oslo, Norway; <sup>8</sup>OsloMet – Oslo Metropolitan University, Oslo, Norway.

\*Correspondence: Kristian Alfsnes, kalfsnes@gmail.com

**Keywords:** *Neisseria meningitidis*; *Neisseria lactamica*; experimental evolution; transformation; homologous recombination; genome sequencing.

**Abbreviations:** CI, confidence interval; DUS, DNA uptake sequence; Gc, gonococcus; HGT, horizontal gene transfer; HR, homologous recombination; indels, insertions and deletions; Mc, meningococcus; MMR, mismatch repair; NGS, next-generation sequencing; rif, rifampicin; RM, restriction modification; SNV, single nucleotide variant.

**Data statement:** All supporting data, code and protocols have been provided within the article or through supplementary data files. Two supplementary figures and three supplementary tables are available with the online version of this article.

Since the late 1950s, experimental transformation of the Mc with antibiotic-resistance markers has given accurate insights into the selectivity and mechanism of a highly evolved function [6, 7]. Since intraspecies transformation frequencies were early on found to be higher than interspecies transformation frequencies, preference for DNA was, prior to DNA sequencing, used for taxonomic classification of the neisseriae [4, 8]. Transformation, like sexual reproduction in general, allows for allelic re-assortment of homologous DNA [9]. The uptake and dynamic exchange of homologous DNA in competent bacteria may, through natural selection, maintain conserved regions of the genome and allow plasticity of others [10, 11]. Genomic comparisons of *Neisseria lactamica*, *Neisseria gonorrhoeae* (the gonococcus; Gc) and the Mc has shown that each species is a coherent genetic unit with more interspecies horizontal genetic exchange having taken place in the accessory than the core genome [12].

Selectivity for self-DNA in Mc and its close relative Gc is greatly influenced by preference for a specific 10–12-mer sequence, termed the DNA uptake sequence (DUS) [13, 14]. Possession of thousands of DUS copies is a genomic hallmark of the *Neisseriaceae* encompassing several dialects assorting to different species of the family [15]. Greater difference between DUS dialects severely impairs transformability. The two pathogenic *Neisseria* (Mc and Gc) and the commensal *N. lactamica* share the canonical 5'-ATGCCGCTGAA-3' DUS, facilitating their interspecies transfer. Clade-associated restriction-modification (RM) systems have also been shown to raise a barrier to transformation in Mc [16] and other bacteria [17]. Furthermore, the propensity to produce double-strand ends at unmethylated endonuclease recognition sites in endogenous DNA has been proposed to present more suitable material for HR and as such increase the rate of HR with self-DNA [18, 19]. Transformation of Mc using a heterologous marker, but not a homologous one, has been shown to depend on the number of susceptible RM sites in the donor DNA [20]. Several models for the interaction between RM systems and transformation have been proposed [21].

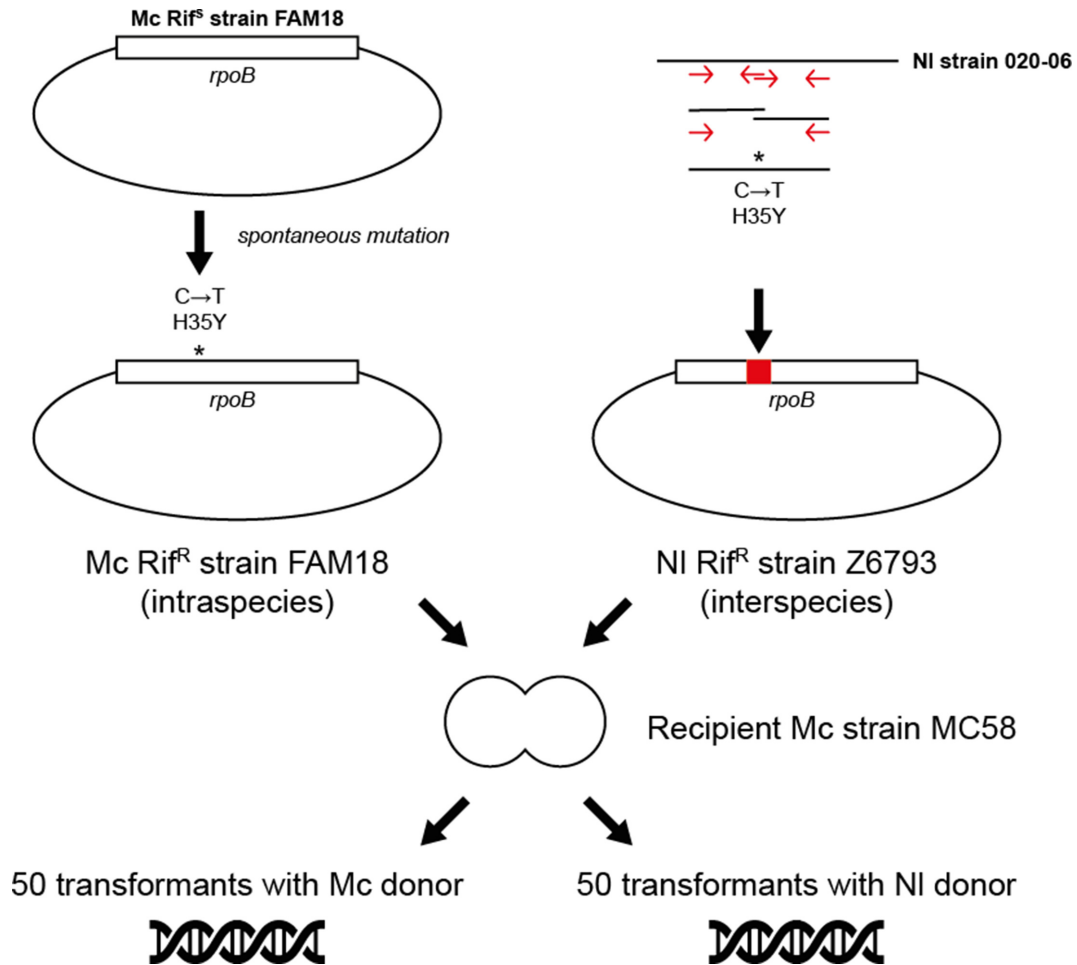
HR, the final step of transformation, in the Mc and Gc is dependent on the recombinase RecA, DprA and resolvases [22–25]. RecBCD has been shown to be required for efficient recombination during transformation in Gc [24]. The gonococcal RecA has been shown to possess considerably higher ATPase activity during strand displacement than that of *Escherichia coli* and found to catalyse more strand exchange through regions of microheterology [26]. At the genomic level, identification of recombination events and their endpoints in Mc and other bacteria has proven to be arduous. A particular challenge in endpoint identification is to distinguish donor-originating stretches from those that were originally present in the recipient when the sequences are identical [27, 28]. Comparing and cross-validating observed single nucleotide variants (SNVs) in a transformed recipient Mc to the donor DNA sequence enables a

## IMPACT STATEMENT

Exchange of genetic material (DNA) between bacteria is a major cause of the global spread and development of antibiotic resistance in pathogenic bacteria. Transformation, the uptake and recombination of exogenous DNA, may effectively transfer antibiotic-resistance genes or mutations between bacteria. *Neisseria meningitidis* can cause life-threatening disease in humans and is highly transformable. Here, we show for what is believed to be the first time the detailed genomic impact of experimental transformations of an antibiotic-resistance allele into the meningococcus with DNA from a same (intra) or a closely related (inter) species. Sequence alignment of the parallel transformed recipient genomes allowed for the identification of allelic replacement. The inclusion and comparison of both intraspecies and interspecies transformation, and the small antibiotic-resistance allele involving a single-nucleotide change, add to our understanding of an increasing problem. The depth of sequence coverage and cross-validated data-points ensure that the findings in the present study can reliably be added to the field of knowledge. Finally, in this study we suggest several hypotheses to explain the observed mosaic recombination patterns, increasing our understanding of horizontal gene transfer and bacterial evolution. The data is important for understanding transformation in a feared human pathogen at the genomic level for the first time.

conservative estimate of the extent of a recombination event. Sequence divergence itself strongly influences the efficiency of HR [29–32]. Thus, acquiring sufficient resolution of sequence polymorphisms to accurately identify the recombination endpoints is a trade-off with transformation efficiency in an experimental setting.

A study of HR events by whole genome sequencing following *in vitro* transformation in *Streptococcus pneumoniae* has shown that transformation can cause fragmented or mosaic HR patterns, with multiple non-contiguous segments originating from the same molecule of donor DNA [33]. Experimental transformation of *Streptococcus pneumoniae* revealed larger recombination event sizes in the transformed recipient when allowed cell-to-cell contact with their donor, compared to *in vitro* donor DNA transformation [34]. Heteroduplex correction by mismatch repair (MMR) was not found to associate with the mosaic patterns observed. In the same study, in addition, multiple sequence imports at selected and unselected loci across the genome were frequently found. On average, 1.4% of the recipient genome was found to have undergone recombination. Mosaic HR patterns have also been observed in a smaller fraction of *Haemophilus influenzae* transformants having used a similar protocol of genome sequencing following *in vitro*



**Fig. 1.** The construction and characteristics of the donor DNA, and the scheme for the intraspecies and interspecies transformation experiments. Asterisks indicate the SNV in the *rpoB* gene encoding rif resistance. The *rpoB* gene was placed within a circular chromosome (diagram not-to-scale). Red arrows indicate the primers used to generate the Rif<sup>R</sup> Z6793 donor. The recipient Mc strain MC58 is illustrated as a diplococcus. NI, *N. lactamica*.

transformation [35]. The mosaicism of unselected loci was generally found within a relatively short distance from the origin of replication and, therefore, was interpreted to associate with segregation of recombination products after replication [35]. On average, the length of recombined fragments was larger in *Haemophilus influenzae* than in *Streptococcus pneumoniae* (2.3 vs 6.9 kb), but in total amounted to replacement of approximately 1% of the genome in both species [33, 35].

Here, the genomic result of Mc transformation events was monitored by whole-genome sequencing and SNV analysis thereof. Genomic donor DNA from Mc strain FAM18 and a closely related species *N. lactamica* strain Z6793, both encoding rifampicin (rif) resistance from a single nucleotide mutation in the *rpoB* gene (Rif<sup>R</sup>), were used to transform Mc strain MC58 prior to whole-genome sequencing of the resulting Rif<sup>R</sup> transformants. The distribution of donor-specific SNVs relative to the recipient genome was

monitored using next-generation sequencing (NGS). This enabled unprecedented observations of the genetic consequences of transformation in this competent bacterium. The differences between the transformation of the two different donor DNAs and the resulting mosaic recombination events are described herein, and models to fit these observations reviewed.

## METHODS

### Experimental design

The well-characterized Mc strain MC58 (serogroup B) was chosen as the recipient for transformation (termed 'MA02') assays based on its available genome sequence and transformability (Fig. 1). The Mc strain FAM18 (serogroup C) was chosen as donor DNA (termed 'MA01') for its propensity to acquire rif resistance (Rif<sup>R</sup>) by spontaneous mutation relative to the spontaneous mutation frequency in the

recipient isolate, as well as its sequence divergence from strain MC58. Thereby, the genomic impact of intraspecies transformation could be observed. The annotated genomes of both Mc strains are available in GenBank (accession numbers NC\_003112.2 and NC\_008767.1). In order to test interspecies transformation of MC58, the *N. lactamica* genomic Rif<sup>R</sup> DNA had to be made first. This was achieved by transforming *N. lactamica* strain Z6793 with a Rif<sup>R</sup> PCR product amplified using *N. lactamica* strain 020-06 as the template (Fig. 1, right).

### Bacterial strains and transformation assay

The Rif<sup>R</sup> Mc strain FAM18 was obtained by a spontaneous single nucleotide mutation in *rpoB* with selection on gonococcal medium (GC) with rif (3 mg ml<sup>-1</sup>) agar plates (Fig. 1). The Rif<sup>R</sup> *N. lactamica* Z6793 was obtained by transforming Z6793 with two combined *N. lactamica* Rif<sup>R</sup> PCR constructs (with a total length of 412 nt). The PCR constructs were made using genomic DNA from *N. lactamica* isolate 020-06 as the PCR template (Fig. 1, Table S1, available with the online version of this article). Transformation of *N. lactamica* Z6793 with the *N. lactamica* Rif<sup>R</sup> PCR construct followed a previously described method for spot-transformation on GC-rif plates [36]. Mutation frequencies were determined from the number of colonies on the negative control plates. The isolates used in the transformation assays are listed in Table S2. Genomic DNA was prepared from re-plated colonies following a previously described method [37]. The presence of the single nucleotide mutation in FAM18 Rif<sup>R</sup> (corresponding to a H35Y change as described by Carter, Abadi, Yakubu and Pennington [38]) and the correct insertion of the PCR construct in Z6793 Rif<sup>R</sup> were confirmed using PCR and sequencing using the primers listed in Table S1. In the transformation assay, the rif sensitive (Rif<sup>S</sup>) recipient Mc strain MC58 was transformed with the intraspecific genomic DNA of Mc FAM18 Rif<sup>R</sup> or the interspecific DNA of *N. lactamica* Z6793 Rif<sup>R</sup>. Negative control experiments included genomic DNA of FAM18 Rif<sup>S</sup> and Z6793 Rif<sup>S</sup>, as well as water – no transformants were observed in these. Single colonies of the recipient MC58 were incubated in 0.5 ml GC medium for 45 min with 2 µg genomic donor DNA ml<sup>-1</sup>; 50 parallel transformation assays were set up for both the intraspecies and interspecies assays. Previous tests in the laboratory have confirmed a 2 µg ml<sup>-1</sup> final concentration of donor DNA to be sufficient for complete saturation of the bacterial suspensions [13]. A sample of 0.1 ml of each cell suspension was plated on GC-rif plates and incubated for 24 h. A single c.f.u. from each parallel executed assay was re-plated and incubated for another 24 h. Genomic DNA suitable for NGS was prepared from 50 Mc and 48 *N. lactamica* donor transformants according to a method described elsewhere [37].

### NGS DNA preparation and analysis

Two sets of sequencing libraries were prepared from the DNA extracted from each of the samples, one library using a Nextera XT DNA library preparation kit from Illumina,

and a second library using KAPA HyperPlus from Roche, both following the manufacturers' instructions. All DNA libraries were sequenced using the MiSeq sequencing platform from Illumina, the first library with 2×150 reads (v2 chemistry) and second library with 2×300 reads (v3 chemistry), following the manufacturer's instructions. FASTQ reads from the first and second round of sequencing were merged for analyses in order to increase the quality and coverage (output summary in Table S2).

### Consensus sequence

A recipient consensus sequence was prepared by using a three-step method. First with a *de novo* assembly using Spades v3.10.1 [39] of the FASTQ reads, careful-mode enabled to reduce the number of mismatches, and insertions and deletions (indels), and coverage cut-off set to auto to reduce low coverage regions. This was followed by contig ordering of the *de novo* assembled contigs to the MC58 reference sequence (NC\_003112.2) using Contiguator v2.7.5 (default settings) [40]. The resulting FASTA sequence was then used as the reference in the third step, a reference alignment using Bowtie2 v2.3.0 (default setting, with the maximum fragment length, 'maxins', set to 600 to include large reads from the 2×300 runs) [41]. SAMtools v1.4.1 (and subscripts BCFtools and vcfutils.pl) [42] was used to convert the reference alignment output from Sequence Alignment Map (SAM) to Binary Alignment Map (BAM) and consensus calling (default settings). The resulting MC58 recipient consensus sequence (MA02) was henceforth used as a reference for all the transformed samples.

### SNV calling

Sequences from the transformants in FASTQ format were first trimmed using Trimmomatic 0.36 (removing specified adapters with setting 2:30:10; leading/trailing, 3; sliding window, 3:15; and minimum length, 36) [43]. This step was omitted in the three-step method for recipient consensus sequence construction to reduce loss, whereas it was found to improve the quality of the reference alignment of the samples prior to SNV calling. Sequences were then aligned to the recipient consensus sequence (MA02) using Bowtie v2.3.0 (default settings, maxins 600). SNVs were called using SAMtools v1.4.1 and VarScan v2.3 (default settings, with strand-filter disabled) [44]. In order to observe SNVs only resulting from HR following transformation of the donor DNA, a strict sequence monitoring scheme was adopted (Fig. S1). Genome-wide SNVs were identified using a stringent cut-off set to 95 % frequency when calling the donor consensus sequences to avoid potential erroneous SNVs (Fig. S1a). Therefore, SNVs in the transformants not found in the respective donors (Fig. S1b) and SNVs in the transformants not matching the donor nucleotide variant (Fig. S1c) were removed. The donor-matching SNVs were further filtered using a cut-off set to 85 % frequency following assembly to remove ambiguous bases (Fig. S1d).

## Statistical analyses and visualization

All analyses were performed using R v3.4.3 with the packages ggplot2 v2.2.1 [45] and tidyr v0.8.0 (developed by Hadley Wickham) and MASS v7.3–50 [46] – the latter used to estimate shape parameters to check size distributions. Median values and empirical 95 % confidence intervals (CIs) extracted from bootstrapped intervals (2000 iterations) (with the ‘boot’ package) reported for nonparametric distributions. *Recombined regions* are stretches of sequences where all possible SNVs are in the donor variant, while interspersed regions lacking these donor SNVs were identified as *non-recombined regions*. The *recombination event lengths* are defined as the maximum distance between Rif<sup>R</sup>-associated SNVs in the recipients. Sequences were aligned between recipient MA02 and donors MA01 and ‘NlacRif’ of the region surrounding the resistance marker ( $\pm 50$  kb from Rif<sup>R</sup>) using the progressiveMauve algorithm (default settings – Muscle v3.6) [47], where indels relative to the recipient MA02 were excluded from the alignment. The sequence alignment was annotated using Prokka v1.13 [48], followed by manual inspection of annotated genes in Geneious v11.0.3 (whereupon genes that had been split into adjacent subunits were merged). Sequence alignments were scored using the megablast algorithm for highly similar sequences in BLASTN v2.8.0+ [49]. Compiling the annotated genes together with the BLAST alignment results was carried out using the GenoPlotR v0.8.7 package in R [50]. Sequence similarities were calculated using plotcon (default score matrix EDNAFULL) [51].

## RESULTS

The transformation frequency of MC58 with intraspecific and interspecific Rif<sup>R</sup> DNA ( $1.2 \times 10^{-4}$  and  $7.3 \times 10^{-6}$  c.f.u.<sup>-1</sup>, respectively) was found to be three orders of magnitude higher than the spontaneous mutation frequency of the recipient ( $6.9 \times 10^{-9}$  c.f.u.<sup>-1</sup>; Table S2). Initial trials revealed that recipient MC58 was transformable with DNA from *N. lactamica* isolate Z6793, but not with the sequenced and annotated *N. lactamica* 020–06. The *N. lactamica* donor strain Z6793 was itself competent for transformation and allowed recombination of the Rif<sup>R</sup> PCR product (mean  $1.2 \times 10^{-4}$  c.f.u.<sup>-1</sup>) in the preparation of interspecific *N. lactamica* donor DNA. A trial to spontaneously generate Rif<sup>R</sup> in the *N. lactamica* isolate Z6793 was not successful (Table S2). Both intraspecific and interspecific chromosomal donor DNA were extracted using the same method; gel electrophoresis revealed a large and indistinguishable fragment size distribution. Transformants of both donor DNA were phenotypically similar in size, coloration and growth rate. Forty-eight Rif<sup>R</sup> isolates generated in the interspecies transformations and fifty Rif<sup>R</sup> isolates generated in the intraspecies transformations were successfully prepared into genomic libraries.

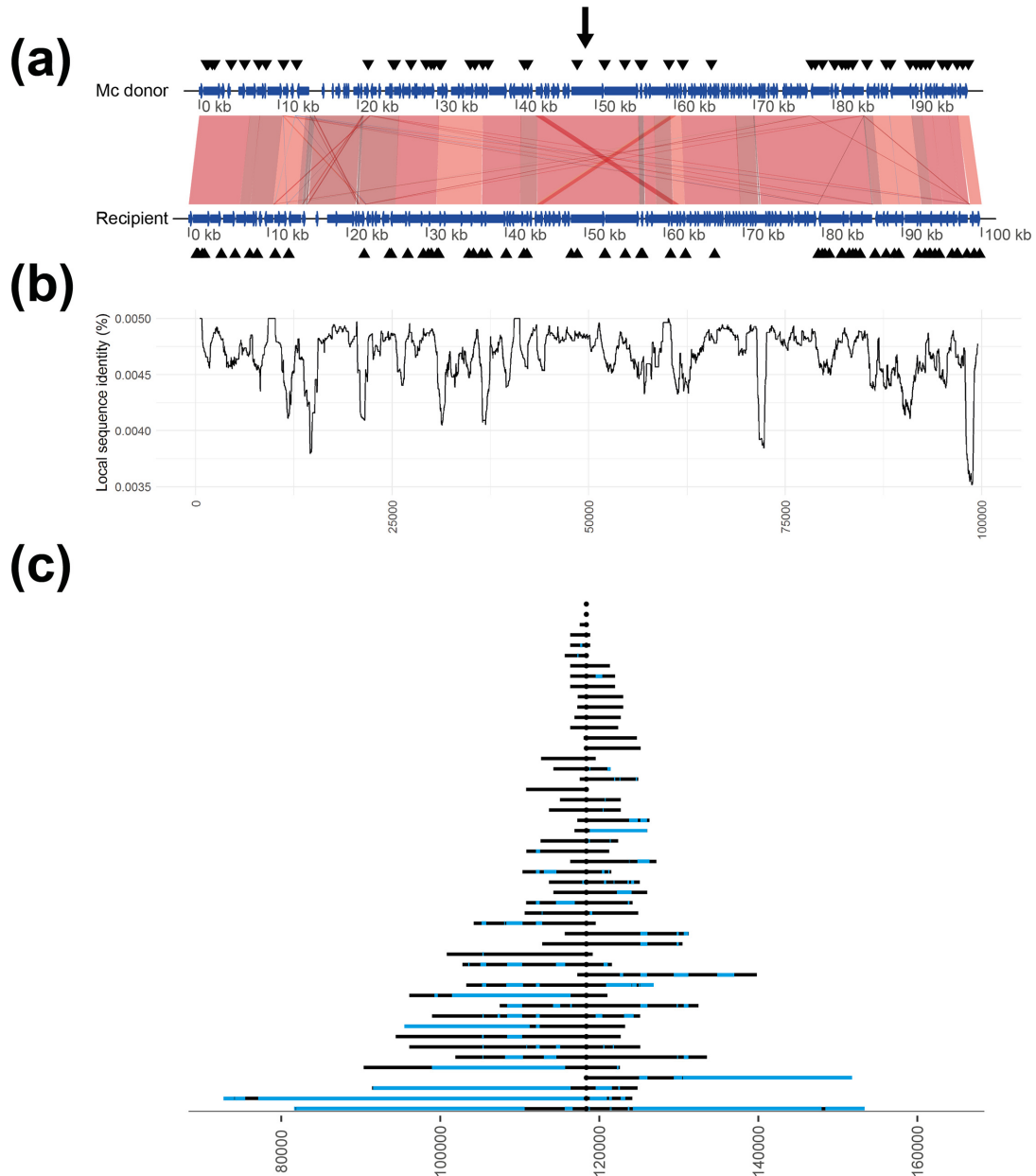
The analysis of the sequencing reads was designed to ensure that only SNVs representing the donor DNA were included in the study. Briefly, assembly of *de novo* contigs to the

reference sequence of MC58 was used to produce an intermediate reference sequence, the intermediate sequence was subsequently used for a final reference assembly resulting in the consensus recipient sequence henceforth used. The sequence reads of the donors and transformants were mapped against the consensus recipient sequence, and the resulting alignment was filtered in order to map only the transformants’ SNVs that matched with the donor’s SNVs. All samples were prepared and sequenced twice in order to increase the coverage and quality of the assemblies. Sufficient read coverage ( $>30\times$ ) following the reference-based assembly prior to SNV calling was observed for both donor and recipient sequences, as well as for the transformants (with a single exception – sample ‘M1’ with  $28\times$  mean coverage; Table S2).

Recombination endpoints could hypothetically be anywhere from  $\pm 1$  nt position between the last observed donor SNV and the non-observed donor SNV. A conservative approach was adopted in this study assigning only SNVs present in the donor consensus as confirmed recombined regions. Comparison of recipient and donor genomes revealed a few duplicated regions in the recipient, most importantly a 245 nt region 3.6 kb upstream and a 254 nt region 1.8 kb downstream of Rif<sup>R</sup> corresponding to the homologous genes *tufA* and *tufB* (Figs 2a and 3a). A few additional, smaller duplications were identified further away ( $>30$  kb) from Rif<sup>R</sup>, and a few small regions with reversed direction were found when comparing both donors to the recipient (amounting to a total of 582 and 609 nt in the 100 kb window, for the Mc and *N. lactamica* donors, respectively; Figs 2a and 3a). Duplicated regions may interfere with the reference sequence alignment if the size of the regions exceeds that of the sequence read length, the longer read lengths in this study ( $2 \times 300$ ) are likely long enough to bridge these duplicated regions overcoming this potential source of error.

SNVs outside the 100 kb region surrounding the Rif<sup>R</sup> (position 118 375 in the recipient consensus sequence) in the Mc donor transformants included: two closely positioned SNVs found in three transformants (M14, M23 and M32) at 9.8 kb (108 kb upstream of Rif<sup>R</sup>), and two closely positioned SNVs found in one transformant (M1) at 50 kb (68 kb upstream of Rif<sup>R</sup>), plus an additional six closely positioned SNVs found in the same transformant (M1) at 1.6 Mb (1.5 Mb downstream of Rif<sup>R</sup>) (Table S3). For the *N. lactamica* donor transformants, such removed SNVs included only a single SNV found in one transformant (L28) at 1.1 Mb (1 Mb downstream of Rif<sup>R</sup>) (Table S3).

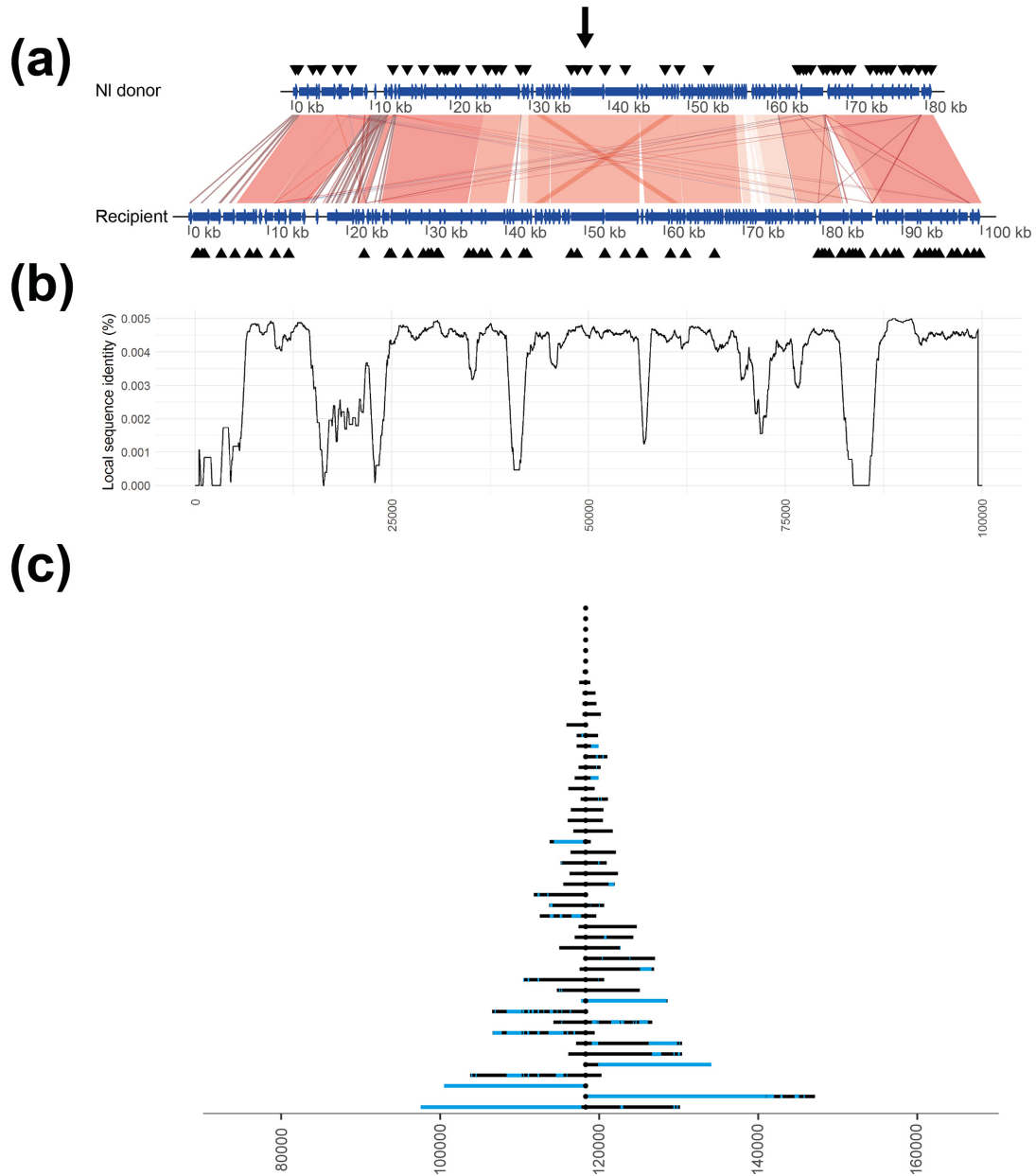
The recombination event lengths were exponentially distributed (one-sample Kolmogorov–Smirnov test with estimated exponential shape parameter,  $P=0.12$  and  $P=0.66$  for the Mc and *N. lactamica* donors, respectively; Figs 2c and 3c). The median recombination event lengths were found to be 10 620 nt (95 % CI 6254–22 578 nt) and 5750 nt (95 % CI 6845–20 682 nt) with intraspecific and interspecific DNA, respectively. Recombination event lengths for the



**Fig. 2.** Comparative overview of donor and recipient genomes, and the results of intraspecies transformations. (a) The aligned sequences of the recipient and the Mc donor (MA01), where a gradient from grey to dark red indicates increased alignment score; annotations are shown as blue arrows and the positions of the 10-mer DUSs are shown as black triangles. The black arrow corresponds to the  $Rif^R$  position in the recipient (at 118 375 in our consensus sequence). (b) The sequence similarity between the recipient and the Mc donor. Positions on the x-axis are as in (a). (c) The maximum extent of the recombination for each sample in the experiment (the recombination event endpoints), where blue regions represent missing donor SNVs (non-recombined regions). Positions on the x-axis corresponds to the consensus sequence,  $Rif^R$  is positioned in the centre (at 118 375 nt). Each recombination event endpoint is represented by an acquired donor SNV; however, due to the scale of the figure this (these) may not be visible.

transformants with the Mc donor were found to be significantly longer than with the *N. lactamica* donor (Wilcoxon rank sum test,  $P < 0.001$ ; Fig. 4, left panel). Two of the Mc transformants and five of the *N. lactamica* transformants contained an unaccompanied  $Rif^R$  SNV without surrounding SNVs of donor origin and were hence indistinguishable

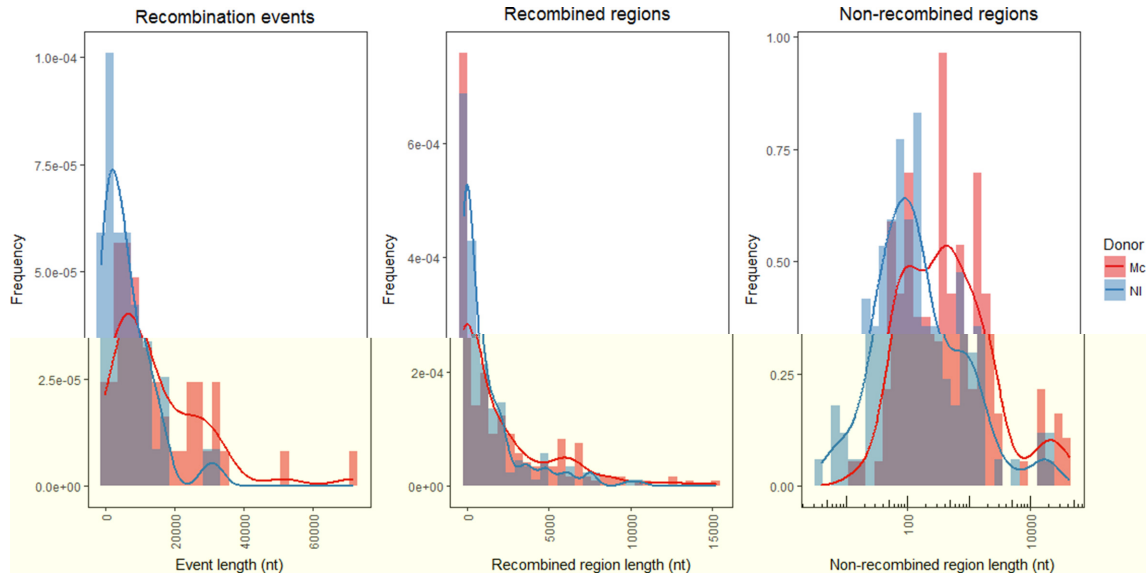
from spontaneous  $Rif^R$  mutants (Figs 2c and 3c). The median size of the recombined regions was found to be 867 nt (95 % CI 175–1695 nt) and 505 nt (95 % CI 308–998 nt) in the intraspecies and interspecies transformations, respectively, and not significantly different from each other (Wilcoxon rank sum test,  $P = 0.58$ ; Fig. 4, middle panel). The



**Fig. 3.** Comparative overview of donor and recipient genomes and the results of interspecies transformations. (a) The aligned sequences of the recipient and the *N. lactamica* donor (NlacRif), where a gradient from grey to dark red indicates increased alignment score; annotations are shown as blue arrows and the positions of the 10-mer DUSs are shown as black triangles. The black arrow corresponds to the Rif<sup>R</sup> position in the recipient (at 118 375 in our consensus sequence). (b) The sequence similarity between the recipient and the *N. lactamica* donor. Positions on the x-axis are as in (a). (c) The maximum extent of the recombination for each sample in the experiment (the recombination event endpoints), where blue regions represent missing donor SNVs (non-recombined regions). Positions on the x-axis correspond to the consensus sequence, Rif<sup>R</sup> is positioned in the centre (at 118 375 nt). Each recombination event endpoint is represented by an acquired donor SNV; however, due to the scale of the figure this (these) may not be visible. NI, *N. lactamica*.

median size of the non-recombined regions was found to be 436 nt (95 % CI 198–700 nt) and 137 nt (95 % CI 87–215 nt) in the intraspecies and interspecies transformations, respectively, and was found to be significantly longer in the intraspecies transformants (Wilcoxon rank sum test,  $P < 0.001$ ;

Fig. 4, right panel). No significant difference was observed when comparing the proportion of non-recombined regions to recombined regions between the two donors (Wilcoxon rank sum test,  $P = 0.53$ ; Fig. S2). The median numbers of non-recombined regions per sample were three (95 % CI 1–



**Fig. 4.** Distribution of recombination events (left), recombined regions (middle) and non-recombined regions (right) sizes for the transformants of the intraspecies (red) and interspecies (blue) donor DNA. Lines represent smoothed kernel density estimates of the histograms. NI, *N. lactamica*.

5) and two (95 % CI 1–5.5) in the intraspecies and interspecies transformations, respectively, and not found to be significantly different from each other (Wilcoxon rank sum test,  $P > 0.05$ ).

Regions of lower sequence similarity (Figs 2b and 3b) did not correspond with recombination event endpoints (Figs 2c and 3c). However, sequence similarity was found to be significantly higher in recombined regions compared to non-recombined regions for both donors (Wilcoxon rank sum test, both  $P < 0.001$ ; Fig. 5). In all transformants, the majority of all possible SNVs were found in intragenic regions (14 469 vs 1035 and 14 112 vs 951, intragenic vs intergenic positions for the transformants obtained using Mc and *N. lactamica* donor DNA, respectively). Accordingly, the majority of the SNVs were found in intragenic regions in the donor DNA (93.3 and 93.7 % in the Mc and *N. lactamica* donors, respectively). The majority of the intragenic SNVs were found in recombined regions irrespective of donor (65.6 and 74.3 % for the Mc and *N. lactamica* donors, respectively; Fig. 5). The number of intergenic SNVs was found to be nearly identical for both recombined and non-recombined SNVs having used the Mc donor (580 to 455, respectively), whereas a slightly higher number of recombined SNVs compared to non-recombined SNVs was found in these regions having used the *N. lactamica* donor (691 to 260, respectively; Fig. 5). Comparison of the variables revealed dependency of recombined versus non-recombined SNVs and position in the genome (intragenic vs intergenic) for the transformants obtained with Mc donor DNA (Pearson's Chi-squared test,  $X^2 = 39.05$ ,  $P < 0.001$ ), but they were independent in the transformants obtained with

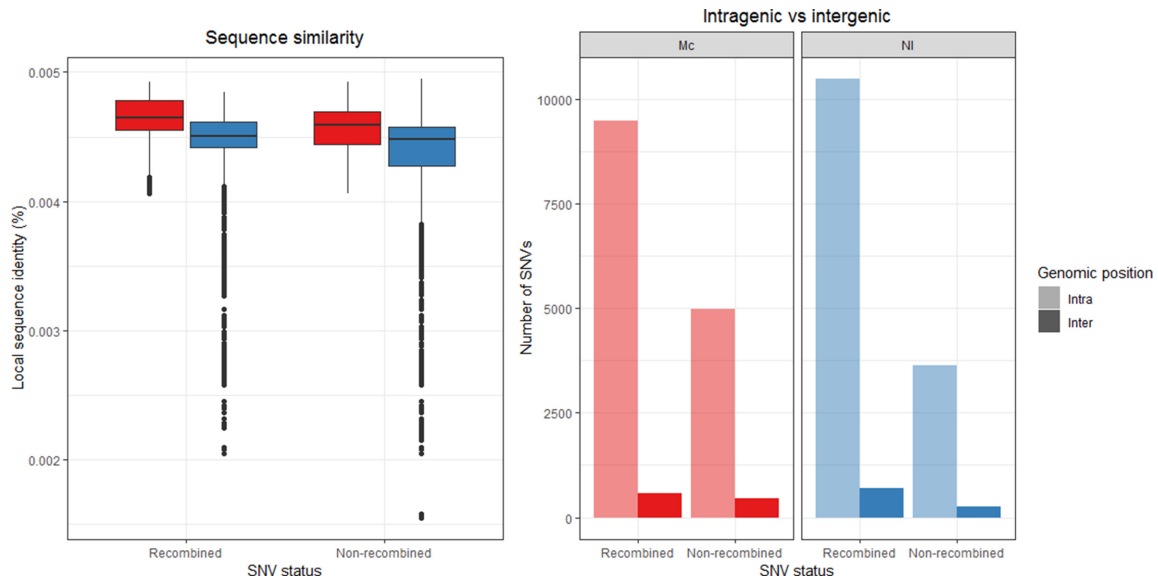
*N. lactamica* donor DNA (Pearson's Chi-squared test,  $X^2 = 1.27$ ,  $P = 0.26$ ).

## DISCUSSION

This experimental set-up was designed to facilitate transformation of an opportunistic pathogenic bacterium to antibiotic resistance through HR of intraspecific and interspecific donor DNA. HR is known to be of great importance for the evolutionary success of competent bacteria such as *N. meningitidis*; here, we enabled a view of the genomic impact of HR by NGS with whole-genome coverage and resolution.

For highly recombinant bacteria, it is advisable to use *de novo* assemblies to account for any intrachromosomal HR events. We employed a three-step method to prepare the recipient consensus sequence, using a combination of *de novo* assembly of the reads, followed by reference-based assembly of contigs and of reads. Mapping contigs against a reference sequence, and subsequent reference-based assembly of the reads against this scaffold, allowed gaps between the *de novo* contigs to be filled. Using this consensus sequence to map all SNVs, facilitated detection of the positions of recombination events in the donors and transformants, and allowed filtering out SNVs that were not found in the donor DNA. Strict criteria, i.e. the exclusion of non-donor SNVs and ambiguous bases following assembly, were employed in order to remove SNVs that could not be assigned to HR of the donor DNA. The excluded non-donor SNVs and SNVs with ambiguous frequencies may represent HR and shuffling of self-DNA and/or a combination of self-DNA and donor DNA. The diploid genome in the Mc may





**Fig. 5.** Sequence similarity of the SNVs in recombined regions and non-recombined regions (left), and the genomic position (intra-genic/intergenic) of the SNVs in recombined regions and non-recombined regions (right) for the transformants of the Mc (red) and *N. lactamica* (blue) donor DNA. NI, *N. lactamica*.

allow for frequent shuffling, gene conversion and HR of heterologous self-DNA [52, 53]. Other causes for such false-positive SNVs may be spontaneous mutations or other error-prone mechanisms, such as replication slippage, antigenic variation or even sequencing errors. Indels have been shown to be incorporated during transformation and HR at low rates [54], and only maintained in a heteroduplex with the presence of matching sequences upstream and downstream [33]. Additionally, due to the complexity involved in mapping and anchoring indels relative to the recipient consensus, indels were not analysed in the present study. A similar scheme of (cross-) validation of SNVs following HR was employed by Mell, Shumilina, Hall and Redfield [55]. Croucher *et al.* [33] calculated recombined region length, median length ('L50'), using the middle position in the region between the last observed donor SNV and the non-observed (expected) donor SNV. These regions (referred to as *boundary regions*) were found to be very variable in length in the present study. Therefore, calling the median was not considered informative here and a conservative estimate using the last observed SNVs was used instead.

The experimental transformations of live bacteria in this study generated recombination of donor DNA corresponding to a single region surrounding the  $Rif^R$ , the selected rif-resistance allele, with very few exceptions. It is possible that despite the strict cut-off criteria and SNV validation, SNVs found outside of the  $Rif^R$  region in these few samples were considered false positives. These SNVs could also represent genuine recombination events; however, given the small number of clustered SNVs in these cases this seemed less likely. The few SNVs (14 and 1 with the Mc and

*N. lactamica* donors, respectively; Table S3) found outside the 100 kb region were therefore excluded from further analyses. A similar pattern of SNVs found in high density around a selective marker has been shown for pneumococcal transformation, suggested to be the result of HR of a single donor molecule [33].

A three orders of magnitude difference in transformation frequency (efficiency) and mutation frequency in the MMR-proficient donor isolate MC58 was expected to be sufficient to avoid an issue with spontaneous  $Rif^R$  mutations in the transformation assays [56]. The interspecies *N. lactamica* donor has been shown to exhibit protein orthologue similarity at 94 % relative to the recipient Mc, compared to the intraspecies Mc donor shown with protein orthologue similarity of more than 98.7 % [57]. Thus, the transformed and recombined donor genes are not expected to express significant change in the recipient (i.e. growth or morphology), that is of course not including the *rif* resistance allele in *rpoB* – consequently selective pressures are not expected outside of the antibiotic-resistance marker.

Successful transformation of Mc with closely related species, such as *N. lactamica*, has been extensively studied (reviewed by Ambur [6]). The inability to transform the recipient Mc strain MC58 with chromosomal DNA from the *N. lactamica* strain 020–06 remains unresolved and warrants further investigation. The two strains share the same DUS (atDUS) [15], but display different RM profiles [58]. The transformation efficiency of *N. meningitidis* and the subsequent extent of HR is lower for interspecies events, due to divergent (non-homologous) DNA and absent or different DUS motifs [15, 59, 60], and/or different RM systems [16] or

CRISPR organization [61]. Mc intraspecies transformation ratios of streptomycin resistance have been shown to be five times higher than *N. lactamica*-Mc interspecies transformations [60]. A small number of transformed isolates (two and five of the Mc and *N. lactamica* transformants, respectively) was observed with unaccompanied Rif<sup>r</sup> SNVs, possibly resulting from spontaneous single-nucleotide mutations or caused by partial recombination or recombination of very small DNA fragments. The observed exponential increase in recombination event lengths in both intraspecies and interspecies transformations (Figs 2c and 3c) may support HR of partial or very small DNA fragments to mutation, as shown previously for small heterologous DNA fragments of both intraspecific and interspecific origin in *Acinetobacter baylyi* [62].

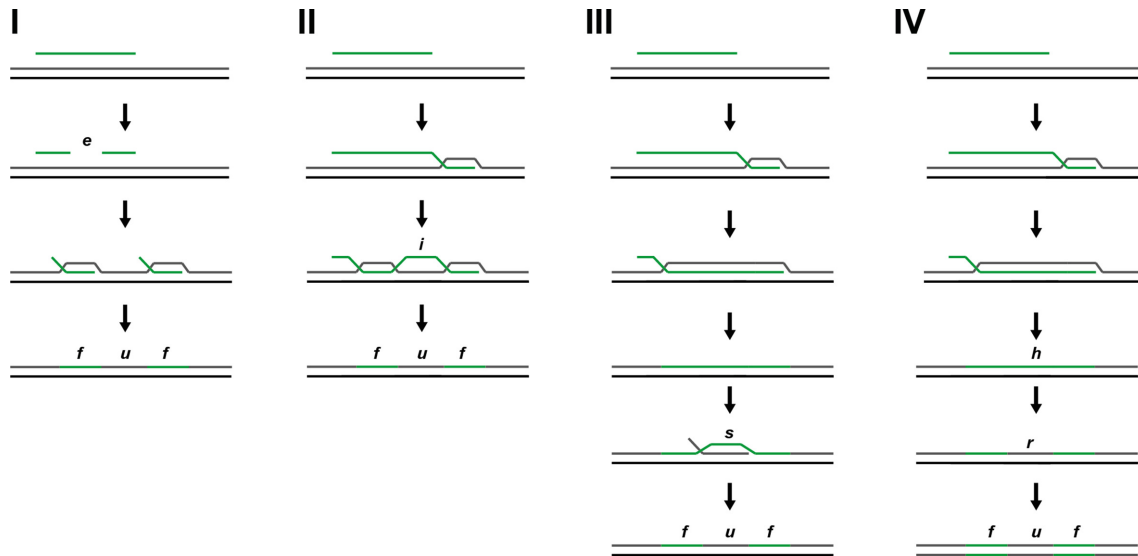
A wide range of recombination event lengths was observed in this study (Figs 2c and 3c), with significantly larger lengths for the transformants of the intraspecific (Mc) donor DNA compared to the interspecific (*N. lactamica*) donor DNA, with median event lengths of 10.6 and 5.8 kb, respectively (Fig. 4, left panel). It should be noted that a generalized conclusion of intraspecific versus interspecific transformation patterns in the genus *Neisseria* could not be made from this study, as this would require multiple comparative intraspecific and interspecific donors and recipients. Nevertheless, the study does show the variability of transformation patterns using two distinct donor sequences, and are in line with the differences in Mc-Mc and *N. lactamica*-Mc transformation rates found by Hoke and Vedros having used a streptomycin-resistance marker [60]. The observed range of recombination event lengths was in agreement with previous estimates in *N. meningitidis* (1.5–9.9 kb in work by Linz, Schenker, Zhu and Achtman [63]) and *Haemophilus influenzae* (1.2–16.6 kb in work by Mell *et al.* [55]), but considerably larger than estimates from *Helicobacter pylori* (1.3 kb in work by Lin *et al.* [64]). HGT by means of generalized transduction or conjugation and recombination of fragments from 40 to 107 kb, or when including mobile elements, up to 240 kb, has been suggested between closely related *E. coli* clones [65]. If such lengths can be observed in the case of natural transformation, and whether exogenous DNA of such recombination substrate lengths would remain intact, remains to be determined. The wide distribution of recombination event lengths observed in this study is similar to that shown in transformation and HR in *Streptococcus pneumoniae* [33, 34], but different to the complete heterogenic distribution of recombined fragments shown in experimental HGT in *E. coli* [66]. Even though the recombination endpoints did not significantly correlate with reduced sequence similarity in this study, it is expected that overall reduced homology will limit the efficiency of transformation and HR, and likely be the cause of the significantly larger recombination events observed in the intraspecies relative to the interspecies transformations.

Fragmented integration of DNA during HR as shown in the present study (Figs 2c and 3c) has been shown previously

[33–35, 55, 64]. It is possible that the length of these recombined regions is underestimated in this study because of the strict criteria calling recombined regions only from the last observed donor SNVs. The length of the non-recombined regions, however, may be overestimated as these were called in-between the last observed donor SNVs (Fig. 4, middle and right panels). The distribution of recombination events was found to be similar in transformants with both donor DNAs, whereas non-recombined regions were found to be significantly larger for the intraspecies transformants compared to the interspecies transformants. The large recombination event lengths observed in the present study may not cause epistatic interactions to be at odds with HR [67], indeed through evolutionary pressures these two mechanisms may work in synergy. The comparable proportion of non-recombined regions to recombination event lengths in both transformations indicate that the larger non-recombined regions found in intraspecific transformants were not a direct consequence of the longer observed recombination fragment lengths. Biological effects and subsequent evolutionary pressures have been suggested to explain the observation of non-recombined genes in the transformation of divergent strains of *A. baylyi* [68], the mechanisms involved in how specific genes would avoid HR, and the absence of HR outside a delimited region shown in this study, remain to be addressed. In the present study, non-recombined regions were not found to be more likely in intergenic regions; however, care must be taken since the intergenic region makes up far fewer SNVs than the intragenic regions (Fig. 5, right panel).

Here, we suggest four models for the observed mosaic patterns of recombined regions interspersed with non-recombined regions, although these may not be mutually exclusive. (I) One model for the origin of the non-recombined regions during HR is by cleavage of the donor DNA by endonucleases (i.e. the models proposed by Johnston *et al.* [21]), counteracted by the potential protection of endogenous DNA from digestion by methylation [35, 64] (Fig. 6, I). DNA is transferred across the inner membrane as ssDNA through ComA, and aided by the PilG protein and other inner-membrane components during Mc transformation [69–71], and increased Mc transformation efficiency in the absence of recipient-specific NlaIV restriction sites in plasmid donor DNA but not genomic homologous DNA has been demonstrated [20]. No correlation was shown between the position of methylation sites and recombination endpoints in *Helicobacter pylori* transformation [64]; however, certain type I endonucleases may cleave the DNA at variable positions away from the recognition sequence [72]; thus, making it difficult to assess correlations [68].

(II) A second model involves the propensity for interruptions of the recombination mechanism and fidelity itself, causing collapse of the D-loop (Fig. 6, II). Secondary structures on the DNA strand (such as transcriptional terminators, stem-loop structures) and/or stretches of non-homologous DNA or indels have been shown to reduce



**Fig. 6.** Models of HR following transformation in *N. meningitidis*. Model I shows fragmentation (e.g. by endonuclease digestion of DNA - *e*) of the transformed DNA prior to recombination, in HR of two fragments (*f*) flanking a non-recombined region (*u*). Model II shows interruptions (*i*) in the recombination mechanism resulting in interspersed integration of DNA. Model III shows how secondary recombination of self-DNA (*s*) may introduce a region (*u*) without the SNVs in the original recombination pattern. Model IV shows how strand-dependent MMR (*r*) following recombination can introduce non-recombined regions (*u*) in the original recombination pattern – asynchronous heteroduplex segregation of the donor region on one strand (*h*) during DNA replication may also introduce non-recombined regions.

transformation efficiency [54]. In the present study, a significantly higher sequence similarity was found in the SNVs making up the recombined regions compared to the sequence similarity in SNVs in the non-recombined regions (Fig. 5, right panel), supporting this model; however, no correlation was found between recombined regions and sequence similarity in a genome-wide transformation assay with *Streptococcus pneumoniae* [33].

(III) A third model suggests continuous recombination of the donor strand, whereas secondary recombination events with endogenous self-DNA introduce what is interpreted to be ‘non-recombined regions’ (Fig. 6, III). Mc and its close relative Gc have been shown to be diploid (whereas other relatives, such as *N. lactamica*, have been found to be haploid) [52, 53]. Efficient gene conversion and intracellular transfer of DNA confirmed by antigenic variation has been shown in Mc in the absence of transformation [73, 74]. As such, available self-DNA should be present and readily exchanged within the cell, supporting this model. It is difficult to find support for this model in the present study. A certain distribution and limitations to the size of the intracellular self-DNA may be expected and consequently the size of the non-recombined regions. The observed wide range of the non-recombined regions seems plausible to support secondary recombination of self-DNA, as a greater sequence similarity and consequently more efficient secondary recombination may explain the increased lengths of the apparent non-recombined regions observed between the

recipient and the Mc donor compared to the *N. lactamica* donor (Fig. 4, right panel).

(IV) A fourth model also suggests continuous recombination of the donor strand, where other mechanisms may influence the resulting recipient’s genome. Strand-specific post-replication repair pathways (e.g. the MMR pathway) may correct a portion of the recombined donor DNA resulting in a region with self-DNA (Fig. 6, IV); however, a study of pneumococcal transformation failed to identify a role of MMR in the generation of such mosaicism [33]. Asynchronous heteroduplex segregation of donor DNA on a single strand during DNA replication may result in breaking up and shuffling of the recombined region, introducing self-DNA into the recombined regions in the daughter cells [35] (Fig. 6, IV). The reported uniform (and hence random) propensity for gene conversion of *tuf* genes in *Salmonella* [75] may similarly in the present study explain the size distribution of non-recombined regions (Fig. 4, right). The diploid nature of Mc (as mentioned in III) may provide the homologous DNA for gene conversion in the case of double strand breaks (as shown in *Salmonella enterica* [76]).

All four models rely on error-prone strand-specific DNA repair pathways (e.g. MMR), allowing for heteroduplex recombination products to arise, otherwise these would be corrected (to self-DNA) prior to DNA replication. Saturation of the MMR following transformation in both *Streptococcus pneumoniae* and *N. meningitidis* has been observed [77, 78]. The mechanisms (repair and asynchronous

segregation) described in the fourth model can be engendered subsequently to any of the other hypotheses, making efforts to identify a single mechanistic explanation for this complex scenario difficult. Finally, the biological consequences (i.e. genes affected) and the varying evolutionary pressure of these may influence the actual isolates we observe following transformation and HR. Detrimental changes, complete or partial genes, or allelic changes of the transformed recipient may be severe such that it is not recovered, or it may favour those that through secondary recombination events restore the specific parts of the genome (e.g. Fig. 6, III and IV).

The present study extends previous efforts to observe and understand the mechanisms involved in HR in naturally transformable bacteria, allowing pathogens to transfer adaptive genetic elements introducing beneficial traits and other changes, such as antibiotic resistance. The combination of experimental transformation and sequencing revealed a complex pattern of both DNA transfer and a wide range of recombination event lengths – with recombined regions interspersed with non-recombined regions. The four models presented may in combination, or individually, explain the pattern of interspersed recombined and non-recombined SNVs observed. Further understanding of these mechanisms may identify targets in the bacterial life cycle that effectively neutralize the ability to successfully transfer and take up genetic elements. This would be extremely beneficial in the combat against infectious bacteria and the ongoing struggle to deal with HGT and antimicrobial resistance.

#### Funding information

This work was supported by GLOBVAC (project 220901) and Centre of Excellence (CoE) (Centre for Molecular Biology and Neuroscience, project 142035) grants from the Research Council of Norway to T.T., as well as support from the Norwegian Oxford scholarship UK to K.A.

#### Acknowledgements

The authors would like to thank Mary Deadman and Katherine Makepeace at the University of Oxford for their contributions to the initial experiments, and Sumana Kalyanasundaram and Merete Molton Worren for their contribution to the analysis of the preliminary data. The authors would also like to thank Kaare Magne Nielsen for valuable feedback on the manuscript.

#### Conflicts of interest

The authors declare that there are no conflicts of interest.

#### Data bibliography

1. Alfsnes et al. ENA. PRJEB27515 (2018).

#### References

- Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. *Nature* 2000;405:299–304.
- Alexander HE, Redman W. Transformation of type specificity of meningococci; change in heritable type induced by type-specific extracts containing desoxyribonucleic acid. *J Exp Med* 1953;97:797–806.
- Davidson T, Tønnum T. Meningococcal genome dynamics. *Nat Rev Microbiol* 2006;4:11–22.
- Sparling PF. Genetic transformation of *Neisseria gonorrhoeae* to streptomycin resistance. *J Bacteriol* 1966;92:1364–1371.
- Seitz P, Blokesch M. Cues and regulatory pathways involved in natural competence and transformation in pathogenic and environmental Gram-negative bacteria. *FEMS Microbiol Rev* 2013;37:336–363.
- Ambur OH. Inter- and intraspecies transformation in the *Neisseria*: mechanism, evolution and DNA uptake sequence specificity. In: Davies JK and Kahler CM (editors). *Pathogenic Neisseria: Genomics, Molecular Biology and Disease Intervention*. Poole: Caister Academic Press; 2014. pp. 59–76.
- Catlin BW. Transformation of *Neisseria meningitidis* by deoxyribonucleates from cells and from culture slime. *J Bacteriol* 1960;79:579–590.
- Jyssum K, Lie S. Genetic factors determining competence in transformation of *Neisseria meningitidis*. 1. A permanent loss of competence. *Acta Pathol Microbiol Scand* 1965;63:306–316.
- Ambur OH, Engelstädter J, Johnsen PJ, Miller EL, Rozen DE. Steady at the wheel: conservative sex and the benefits of bacterial transformation. *Philos Trans R Soc Lond B Biol Sci* 2016;371:20150528.
- Johnsborg O, Eldholm V, Håvarstein LS. Natural genetic transformation: prevalence, mechanisms and function. *Res Microbiol* 2007;158:767–778.
- Lorenz MG, Wackernagel W. Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol Rev* 1994;58:563–602.
- Bennett JS, Bentley SD, Vernikos GS, Quail MA, Cherevach I et al. Independent evolution of the core and accessory gene sets in the genus *Neisseria*: insights gained from the genome of *Neisseria lactamica* isolate 020-06. *BMC Genomics* 2010;11:652.
- Ambur OH, Frye SA, Tønnum T. New functional identity for the DNA uptake sequence in transformation and its presence in transcriptional terminators. *J Bacteriol* 2007;189:2077–2085.
- Goodman SD, Socca JJ. Identification and arrangement of the DNA sequence recognized in specific transformation of *Neisseria gonorrhoeae*. *Proc Natl Acad Sci USA* 1988;85:6982–6986.
- Frye SA, Nilsen M, Tønnum T, Ambur OH. Dialects of the DNA uptake sequence in *Neisseriaceae*. *PLoS Genet* 2013;9:e1003458.
- Budroni S, Siena E, Dunning Hotopp JC, Seib KL, Serruto D et al. *Neisseria meningitidis* is structured in clades associated with restriction modification systems that modulate homologous recombination. *Proc Natl Acad Sci USA* 2011;108:4494–4499.
- Thomas CM, Nielsen KM. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol* 2005;3:711–721.
- Arber W. Genetic variation: molecular mechanisms and impact on microbial evolution. *FEMS Microbiol Rev* 2000;24:1–7.
- Vasu K, Nagaraja V. Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol Mol Biol Rev* 2013;77:53–72.
- Ambur OH, Frye SA, Nilsen M, Hovland E, Tønnum T. Restriction and sequence alterations affect DNA uptake sequence-dependent transformation in *Neisseria meningitidis*. *PLoS One* 2012;7:e39742.
- Johnston C, Martin B, Polard P, Claverys JP. Postreplication targeting of transformants by bacterial immune systems? *Trends Microbiol* 2013;21:516–521.
- Hovland E, Beyene GT, Frye SA, Homberset H, Balasingham SV et al. DprA from *Neisseria meningitidis*: properties and role in natural competence for transformation. *Microbiology* 2017;163:1016–1029.
- Koomey JM, Falkow S. Cloning of the *recA* gene of *Neisseria gonorrhoeae* and construction of gonococcal *recA* mutants. *J Bacteriol* 1987;169:790–795.
- Mehr IJ, Seifert HS. Differential roles of homologous recombination pathways in *Neisseria gonorrhoeae* pilin antigenic variation,

- DNA transformation and DNA repair. *Mol Microbiol* 1998;30:697–710.
25. Rocha EP, Cornet E, Michel B. Comparative and evolutionary analysis of the bacterial homologous recombination systems. *PLoS Genet* 2005;1:e15.
  26. Stohl EA, Gruenig MC, Cox MM, Seifert HS. Purification and characterization of the RecA protein from *Neisseria gonorrhoeae*. *PLoS One* 2011;6:e17101.
  27. Didelot X, Lawson D, Darling A, Falush D. Inference of homologous recombination in bacteria using whole-genome sequences. *Genetics* 2010;186:1435–1449.
  28. Smith JM, Smith NH, O'Rourke M, Spratt BG. How clonal are bacteria? *Proc Natl Acad Sci USA* 1993;90:4384–4388.
  29. Falush D, Torpdahl M, Didelot X, Conrad DF, Wilson DJ et al. Mismatch induced speciation in *Salmonella*: model and data. *Philos Trans R Soc Lond B Biol Sci* 2006;361:2045–2053.
  30. Fraser C, Hanage WP, Spratt BG. Recombination and the nature of bacterial speciation. *Science* 2007;315:476–480.
  31. Hanage WP, Spratt BG, Turner KM, Fraser C. Modelling bacterial speciation. *Philos Trans R Soc Lond B Biol Sci* 2006;361:2039–2044.
  32. Roberts MS, Cohan FM. The effect of DNA sequence divergence on sexual isolation in *Bacillus*. *Genetics* 1993;134:401–408.
  33. Croucher NJ, Harris SR, Barquist L, Parkhill J, Bentley SD. A high-resolution view of genome-wide pneumococcal transformation. *PLoS Pathog* 2012;8:e1002745.
  34. Cowley LA, Petersen FC, Junges R, Jimson D Jimenez M, Morrison DA et al. Evolution via recombination: cell-to-cell contact facilitates larger recombination events in *Streptococcus pneumoniae*. *PLoS Genet* 2018;14:e1007410.
  35. Mell JC, Lee JY, Firme M, Sinha S, Redfield RJ. Extensive cotransformation of natural variation into chromosomes of naturally competent *Haemophilus influenzae*. *G3* 2014;4:717–731.
  36. Dillard JP. Genetic Manipulation of *Neisseria gonorrhoeae*. *Curr Protoc Microbiol* 2011;Chapter 4:Unit4A.2.
  37. Wilson K. Preparation of genomic DNA from bacteria. *Curr Protoc Mol Biol* 2001;56:2.4.1–2.4.2.
  38. Carter PE, Abadi FJ, Yakubu DE, Pennington TH. Molecular characterization of rifampin-resistant *Neisseria meningitidis*. *Antimicrob Agents Chemother* 1994;38:1256–1261.
  39. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.
  40. Galardini M, Biondi EG, Bazzicalupo M, Mengoni A. CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. *Source Code Biol Med* 2011;6:11.
  41. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–359.
  42. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–2079.
  43. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–2120.
  44. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 2012;22:568–576.
  45. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Basel: Springer; 2016.
  46. Venables WN, Ripley BD. *Modern Applied Statistics with S-PLUS*. New York: Springer Science & Business Media; 2013.
  47. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 2004;14:1394–1403.
  48. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.
  49. Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. *J Comput Biol* 2000;7:203–214.
  50. Guy L, Kultima JR, Andersson SG. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* 2010;26:2334–2335.
  51. Rice P, Longden I, Bleasby A. EMBOS: the European molecular biology open software suite. *Trends Genet* 2000;16:276–277.
  52. Tobiasson DM, Seifert HS. The obligate human pathogen, *Neisseria gonorrhoeae*, is polyploid. *PLoS Biol* 2006;4:e185.
  53. Tobiasson DM, Seifert HS. Genomic content of *Neisseria* species. *J Bacteriol* 2010;192:2160–2168.
  54. Stuy JH, Walter RB. Addition, deletion, and substitution of long nonhomologous deoxyribonucleic acid segments by genetic transformation of *Haemophilus influenzae*. *J Bacteriol* 1981;148:565–571.
  55. Mell JC, Shumilina S, Hall IM, Redfield RJ. Transformation of natural genetic variation into *Haemophilus influenzae* genomes. *PLoS Pathog* 2011;7:e1002151.
  56. Richardson AR, Yu Z, Popovic T, Stojiljkovic I. Mutator clones of *Neisseria meningitidis* in epidemic serogroup A disease. *Proc Natl Acad Sci USA* 2002;99:6103–6107.
  57. Treangen TJ, Ambur OH, Tonjum T, Rocha EP. The impact of the neisserial DNA uptake sequences on genome evolution and stability. *Genome Biol* 2008;9:R60.
  58. Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE – a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res* 2015;43:D298–D299.
  59. Graves JF, Biswas GD, Sparling PF. Sequence-specific DNA uptake in transformation of *Neisseria gonorrhoeae*. *J Bacteriol* 1982;152:1071–1077.
  60. Hoke C, Vedros NA. Taxonomy of the neisseriae: deoxyribonucleic acid base composition, interspecific transformation, and deoxyribonucleic acid hybridization. *Int J Syst Bacteriol* 1982;32:57–66.
  61. Zhang Y, Heidrich N, Ampattu BJ, Gunderson CW, Seifert HS et al. Processing-independent CRISPR RNAs limit natural transformation in *Neisseria meningitidis*. *Mol Cell* 2013;50:488–503.
  62. Harms K, Lunnan A, Hülter N, Mourier T, Vinner L et al. Substitutions of short heterologous DNA segments of intragenomic or extragenomic origins produce clustered genomic polymorphisms. *Proc Natl Acad Sci USA* 2016;113:15066–15071.
  63. Linz B, Schenker M, Zhu P, Achtman M. Frequent interspecific genetic exchange between commensal neisseriae and *Neisseria meningitidis*. *Mol Microbiol* 2000;36:1049–1058.
  64. Lin EA, Zhang XS, Levine SM, Gill SR, Falush D et al. Natural transformation of *Helicobacter pylori* involves the integration of short DNA fragments interrupted by gaps of variable size. *PLoS Pathog* 2009;5:e1000337.
  65. Dixit PD, Pang TY, Studier FW, Maslov S. Recombinant transfer in the basic genome of *Escherichia coli*. *Proc Natl Acad Sci USA* 2015;112:9070–9075.
  66. Maddamsetti R, Lenski RE. Analysis of bacterial genomes from an evolution experiment with horizontal gene transfer shows that recombination can sometimes overwhelm selection. *PLoS Genet* 2018;14:e1007199.
  67. Arnold BJ, Gutmann MU, Grad YH, Sheppard SK, Corander J et al. Weak epistasis may drive adaptation in recombining bacteria. *Genetics* 2018;208:1247–1260.
  68. Ray JL, Harms K, Wikmark OG, Starikova I, Johnsen PJ et al. Sexual isolation in *Acinetobacter baylyi* is locus-specific and varies 10,000-fold over the genome. *Genetics* 2009;182:1165–1181.
  69. Chen I, Dubnau D. DNA transport during transformation. *Front Biosci* 2003;8:s544–s556.
  70. Facius D, Meyer TF. A novel determinant (*comA*) essential for natural transformation competence in *Neisseria gonorrhoeae* and the effect of a *comA* defect on pilin variation. *Mol Microbiol* 1993;10:699–712.

71. Frye SA, Lång E, Beyene GT, Balasingham SV, Homberset H *et al.* The inner membrane protein PilG interacts with DNA and the secretin PilQ in transformation. *PLoS One* 2015;10:e0134954.
72. Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA *et al.* A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res* 2003; 31:1805–1812.
73. Swanson J, Morrison S, Barrera O, Hill S. Piliation changes in transformation-defective gonococci. *J Exp Med* 1990;171:2131–2139.
74. Zhang QY, Deryckere D, Lauer P, Koomey M. Gene conversion in *Neisseria gonorrhoeae*: evidence for its role in pilus antigenic variation. *Proc Natl Acad Sci USA* 1992;89:5366–5370.
75. Paulsson J, El Karoui M, Lindell M, Hughes D. The processive kinetics of gene conversion in bacteria. *Mol Microbiol* 2017;104:752–760.
76. Arwidsson O, Hughes D. Evidence against reciprocal recombination as the basis for *tuf* gene conversion in *Salmonella enterica* serovar Typhimurium. *J Mol Biol* 2004;338:463–467.
77. Alexander HL, Richardson AR, Stojiljkovic I. Natural transformation and phase variation modulation in *Neisseria meningitidis*. *Mol Microbiol* 2004;52:771–783.
78. Humbert O, Prudhomme M, Hakenbeck R, Dowson CG, Claverys JP. Homeologous recombination and mismatch repair during transformation in *Streptococcus pneumoniae*: saturation of the Hex mismatch repair system. *Proc Natl Acad Sci USA* 1995;92:9052–9056.

#### Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at [microbiologyresearch.org](http://microbiologyresearch.org).