



The shufflon of IncI1 plasmids is rearranged constantly during different growth conditions

Michael S.M. Brouwer^{a,*}, Stephanie D. Jurburg^b, Frank Harders^b, Arie Kant^a, Dik J. Mevius^{a,c}, Adam P. Roberts^d, Alex Bossers^b

^a Department of Bacteriology and Epidemiology, Wageningen Bioveterinary Research, Lelystad, The Netherlands

^b Department of Infection Biology, Wageningen Bioveterinary Research, Lelystad, The Netherlands

^c Department of Infectious Diseases and Immunology, Utrecht University, Utrecht, The Netherlands

^d Department of Tropical Disease Biology, Liverpool School of Tropical Medicine, Liverpool, UK

ABSTRACT

One of the factors that can affect conjugation of IncI1 plasmids, amongst others, is the genetic region known as the shufflon. This multiple inversion system modifies the pilus tip proteins used during conjugation, thus affecting the affinity for different recipient cells. Although recombination is known to occur in *in vitro* conditions, little is known about the regulation and the extent of recombination that occurs. To measure the recombination of the shufflon, we have amplified the entire shufflon region and sequenced the amplicons using nanopore long-read sequencing. This method was effective to determine the order of the segments of the shufflon and allow for the analysis of the shufflon variants that are present in a heterogeneous pool of templates. Analysis was performed over different growth phases and after addition of cefotaxime. Furthermore, analysis was performed in different *E. coli* host cells to determine if recombination is likely to be influenced. Recombination of the shufflon was constantly ongoing in all conditions that were measured, although no differences in the amount of different shufflon variants or the rate at which novel variants were formed could be found. As previously reported, some variants were abundant in the population while others were scarce. This leads to the conclusion that the shufflon is continuously recombining at a constant rate, or that the method used here was not sensitive enough to detect differences in this rate. For one of the plasmids, the host cell appeared to have an effect on the specific shufflon variants that were formed which were not predominant in another host, indicating that host factors may be involved. As previously reported, the *pilV-A* and *pilV-A'* ORFs are formed at higher frequencies than other *pilV* ORFs. These results demonstrate that the recombination that occurs within the shufflon is not random. While any regulation of the shufflon affected by these *in vitro* conditions could not be revealed, the method of amplifying large regions for long-read sequencing for the analysis of multiple inversion systems proved effective.

1. Introduction

Multiple inversion systems or phase variation systems are an efficient strategy adopted by many, highly diverse bacteria to regulate expression of proteins or to introduce amino acid variability within a protein, allowing quick and reversible adaptation to environmental changes. The variations are mostly introduced by a site-specific recombinase which can recombine multiple contiguous regions of DNA, effectively shuffling the order and directionality of specific segments of the molecule. Some examples of phase variation include gene silencing of *cwpV* in *Clostridium difficile* (Emerson et al., 2009), control of the Mod methyl-transferases in *Haemophilus influenzae* (Atack et al., 2015) and control of the *Streptococcus pneumoniae* capsule protein gene *cap3A* (Waite et al., 2001) but also has analogous features with the more complex system of antibody maturation in higher organisms (de los Rios et al., 2015).

Plasmid-based phase variation systems have also been described including the p15B Min system in a phage-related plasmid in *Escherichia*

coli and the shufflon system present on plasmids of the I-complex in *Enterobacteriaceae*, which includes the IncI1 α , IncI1 γ , IncI2, IncK, IncB/O and IncZ plasmids (Komano, 1999; Sandmeier et al., 1991; Seiffert et al., 2017; Venturini et al., 2013). Due to the high percentage of homology in the DNA sequence between IncI1 α and IncI1 γ plasmids, the variations seen in these plasmids are similar, and these two plasmid types will be discussed collectively here as IncI1 (Brouwer et al., 2015).

The PilV protein encoded by IncI1 plasmids can have one of seven different C-terminal domains which are expressed and assembled at the tip of the thin pilus type IVb during conjugation in liquid media (Roux et al., 2012). This difference in the PilV C-terminal domain, which acts as an adhesin with conjugation recipients, result in differences in affinity for recipient cells of the plasmid during conjugation (Komano et al., 1995). The variations of PilV are generated by the shufflon phase variation system, which usually consists of four segments (A, B, C, D), which are located downstream of the 5' partial constant open reading frame (ORF) of the *pilV* gene (Fig. 1). Further downstream is the gene encoding the site specific recombinase Rci, which catalyses the

* Corresponding author.

E-mail address: mike.brouwer@wur.nl (M.S.M. Brouwer).

<https://doi.org/10.1016/j.plasmid.2019.03.003>

Received 8 January 2019; Received in revised form 12 March 2019; Accepted 14 March 2019

Available online 15 March 2019

0147-619X/ © 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

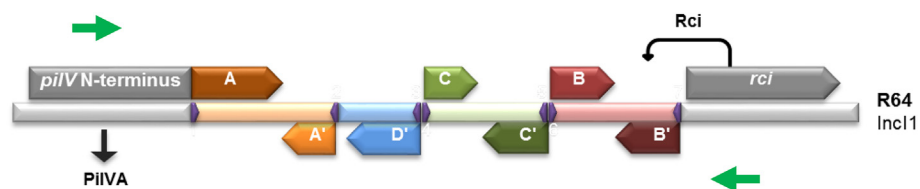


Fig. 1. Structure of the complete shufflon as first described in plasmid R64 (Komanon et al., 1987). Figure adopted from (Brouwer et al., 2015). PCR primers that were used for the amplification of the shufflon region are indicated by above and below the structure by green arrows. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

recombination between the *sfx* recombination sites present at the ends of the four segments, A–D, and results in rearrangement of the order and orientation of the segments (Gyohda et al., 2002). The shufflon is known to be continuously rearranged; resulting in a high diversity of sequences within a population of cells growing in planktonic culture, but the mechanism behind the regulation of the recombination reaction is unknown (Brouwer et al., 2015; Gyohda et al., 2006).

Despite recent advances in next-generation sequencing (NGS) technologies, studies of shufflon recombination have been hampered by the size of the shufflon region (approximately 1.9 kb), which is too large for short-read NGS technology to span in single reads or mate pairs. However, using long-read technologies such as PacBio SMRT sequencing or Oxford Nanopore Technologies MinION, the structural variation of the shufflon can be studied more easily, and this has already shown great heterogeneity within samples during the analysis of the IncI2 shufflon (Sekizuka et al., 2017).

The activity of mobile genetic elements can be influenced by the metabolic state of the host cell, and as such, the activity of the shufflon may be influenced by external factors. To determine this, we have measured the variation of the shufflon of several IncI1 plasmids over different growth phases, in the presence and absence of antibiotics and in different host cells, using ONT MinION long-read sequencing.

2. Materials and methods

2.1. Cell culture and sampling strategy

Shufflon analysis was performed in the original host *E. coli* in which the plasmids were first described, or in *E. coli* DH10B after electrotransformation (Invitrogen ElectroMAX DH10B, ThermoFisher Scientific). Features of the plasmids used in the study are described in Table 1.

Bacterial cells were grown on LB agar plates containing 1 µg/ml cefotaxime to select for presence of the plasmids. Liquid cultures were set up from a single colony and grown overnight in LB medium. For pESBL-4 and pESBL-283 plasmids, cells were cultured for 3 days with daily refreshing of the culture by diluting it 1:100 in fresh LB broth. Samples were taken at the start, middle and end of the exponential growth phase based on optical density (OD₆₀₀ approximately at respectively 0.15, 0.4, 0.8), see Fig. 2. At the start of day 3, cefotaxime was added to the culture medium to a final concentration of 1 µg/ml.

2.2. DNA isolation and sequencing

DNA was isolated using the Blood and tissue DNA isolation kit (Qiagen). Specific amplification of the shufflon region was performed

Table 1
Features of the plasmids and original isolates used for this study.

| Plasmid | Isolate | Origin | Shufflon segments | Reference |
|-----------|----------|---------|-------------------|-----------------------|
| pESBL-4 | ESBL-4 | Human | A, B, C, D | Smith et al. (2015) |
| pESBL-12 | ESBL-12 | Human | A, B, C, D | Brouwer et al. (2014) |
| pESBL-117 | ESBL-117 | Human | A, B, C | Brouwer et al. (2014) |
| pESBL-283 | ESBL-283 | Pig | A, B, C | Brouwer et al. (2014) |
| pESBL-305 | ESBL-305 | Poultry | A, B, C | Brouwer et al. (2014) |
| pESBL-355 | ESBL-355 | Poultry | A, B | Smith et al. (2015) |

using oligonucleotides Nanopore_shufflon_Fw (5'-NNNNNNATGACAG AAGGGCGAGTTC A -3') and Nanopore_shufflon_Rev (5'-NNNNNNGGT GCATTACGTTCCCTGGTC -3') with Biomix Red (Bioline), expected to produce an amplicon of 2860 bp for the entire shufflon region. Cycling program consisted of 4 min 94 °C, 25 cycles of 30 s 94 °C, 30 s 60 °C, 1 min 72 °C and terminating after 10 min final amplification at 72 °C.

PCR products were isolated using a PCR purification kit (Qiagen). Sequencing libraries were prepared according to the manufacturer's protocol using the Ligation sequencing kit (LSK108) and the Native Barcoding kit (EXP-NBD103) from Oxford Nanopore Technologies. Ligation reactions were performed using the NEB Blunt/TA ligase master mix and NEBNext Quick ligation module (New England Biolabs). Sequencing was performed on a MinION sequencer using flowcell type R9.4 (Oxford Nanopore Technologies) and sequencing was performed for 24–36 h.

2.3. Data analysis

Reads were basecalled using Albacore (v2.2.7, Oxford Nanopore Technologies). Demultiplexing and adapter trimming of reads were performed using Porechop (v0.2.1) at default settings (Wick, 2018). Amplicons contain 539 bp of *pilV*_{N-terminus} and 319 bp of *rci*. Any reads that did not contain either of these sequences at the ends were considered sequence errors and discarded.

Statistical analysis was performed using R 3.4.0 (R-Core-Team, 2014). The datasets were rarefied to 2100 randomly chosen reads per sample. The order of the shufflon segments per read were determined using BLAST against the expected sequences of each of the segments.

All raw data was uploaded to the European Nucleotide Archive and is available from accession numbers ERS3014708-ERS3014731.

3. Results and discussion

3.1. Shufflon analysis over time

The accepted biological function of the shufflon is to generate variation in the C-terminal domain of the *pilV* ORF. To determine the activity of the shufflon in various conditions, two IncI1α plasmids, pESBL-4 and pESBL-283, were transformed into *E. coli* DH10B, see Table 1. Cells were grown in liquid culture for 36 h during which samples were taken at several time points throughout different growth phases and after the addition of the antibiotic cefotaxime, to which both plasmids encode resistance (Fig. 2).

The complete shufflon region, including partial ORFs of the upstream *pilV* and the downstream *rci*, was amplified by PCR and sequenced by nanopore longread sequencing. Analysis of the sequenced reads showed that many variants of the shufflon were present and new variants were continuously measured at consecutive timepoints (Fig. 3a). For pESBL-4 the mean number of different variants per time point was 178 while the mean for pESBL-283 was 89 over the seven time points. Furthermore, the cumulative number of variants over all time points was 522 versus 157 respectively. This difference was caused by the different number of shufflon segments between these plasmids. The presence of segment D in pESBL-4 contributed to an exponential increase in the hypothetical biological variation (Brouwer et al., 2015; Sekizuka et al., 2017).

Analysis of the shufflon amplicons showed that many of the reads

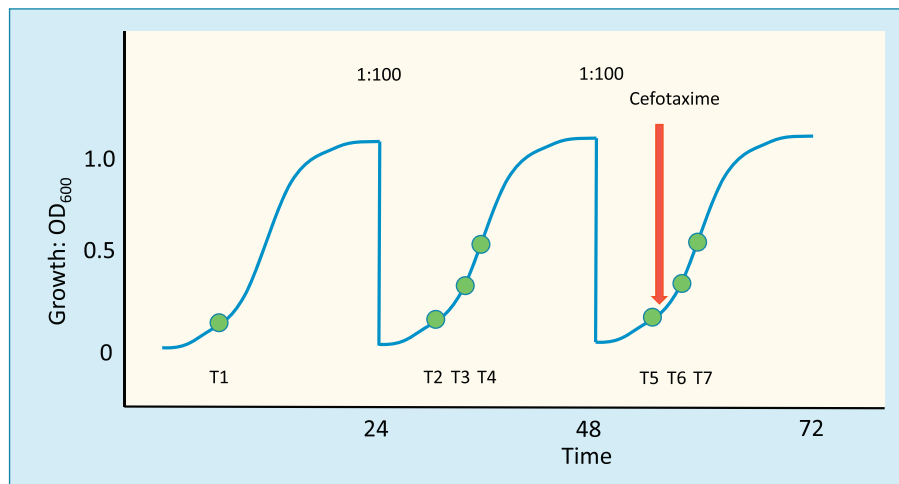


Fig. 2. Impression of the sampling scheme for experiment 1. Each green circle represents a time point and the appurtenant growth condition for which the sample was taken. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

did not contain all of the expected shufflon segments (Fig. 3b). pESBL-4 contained 4 segments in 59 to 64% of the reads while pESBL-283 contained 3 segments in 57 to 66% of the reads over all time points. Although the majority of plasmids still contain all segments over all time points, the loss of these segments *in vitro* is far greater than expected as most IncI1 plasmids described in the literature contain either 3 or 4 segments (Brouwer et al., 2015). It is therefore hypothesized that in most *in vivo* environments the plasmids will experience greater

selective pressure in which loss of shufflon segments will lead to severe loss of fitness compared to those plasmids that have retained all shufflon segments. Although loss of segments was measured from time point 1, the fraction of plasmids that retained all segments of the shufflon was stable over time, between 56 and 66% of the reads (Fig. 3b). Duplication of segments was observed in a small number of reads (0.7–0.8%) which had not previously been reported as it would not be possible to assemble using short-read NGS data.

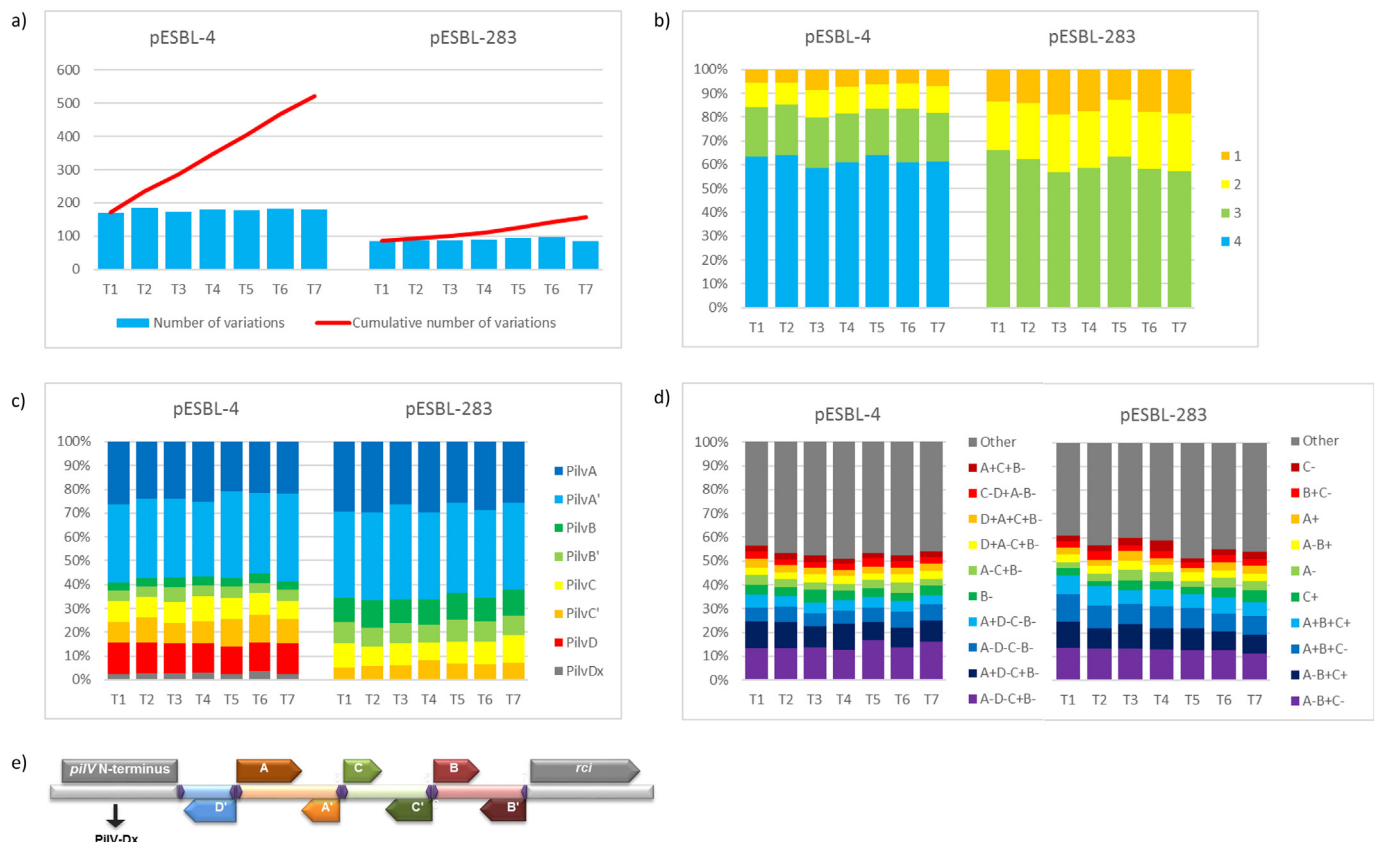


Fig. 3. Analysis of shufflon rearrangements through time. Plasmids pESBL-4 and pESBL-283 were both present in *E. coli* DH10B cells. (a) Blue bars show the number of shufflon variations measured at each timepoint. The red lines show the cumulative number of variations in the pool measured over the successive time points. (b) Counts of the number of shufflon segments that are present in the sequencing reads, measured over time. (c) Relative prevalence of each of the PilV ORFs that were formed, measured over time. (d) Relative distribution of the 10 most abundant shufflon variants, measured over time. (e) Novel variant of the PilV ORF PilvDx that was detected in low prevalence. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Analysis of the C-terminal end of the *pilV* ORF showed that for both pESBL-4 and pESBL-283, over all time points there was a preference for the *pilV-A* and *pilV-A'* ORFs, together making up 55–75% of all reads in each of the conditions (Fig. 3c). While *pilV-B* and *pilV-B'* were less frequent in pESBL-4, in pESBL-283 equal fractions of the ORFs of segments B and C were measured. In pESBL-4, *pilV-D* was detected in approximately equal fractions to *pilV-C* and *pilV-C'* but unexpectedly, the D segment was also detected in opposite orientation in a small fraction of reads, which would lead to a truncated PilV protein, referred to here as PilV-Dx, Fig. 3e.

Analysis of the complete shufflons indicated that certain variants appear to be present in the pool of variants more often than others (Fig. 3d). This finding is supported by previously published data on the IncI2 plasmids (Sekizuka et al., 2017). Based on the data presented here, we cannot conclude if these variants were formed early on in the experiment and continuously outcompeted all other variants due to their overrepresentation in the starting culture or if recombination favors these variants. However, as shufflon recombination is actively ongoing in the cultures, greater heterogeneity could be expected.

No particular difference between the various measured time points could be detected, either in the number of variants present in a population of plasmids or the rate at which they emerge as derived from the cumulative number of novel variants (Fig. 3a), suggesting that stress caused by high cell densities and nutrient deprivation or presence of an antibiotic had no measurable effects.

3.2. Shufflon analysis in different host cells

To determine if the type of host cell had any influence on the recombination of the shufflon, five IncI1 α plasmids were grown either in their original *E. coli* host or in *E. coli* DH10B. For these cells, only a single time point was measured, early exponential phase. In addition to the plasmids pESBL-4 and pESBL-283, pESBL-12, pESBL-117 and pESBL-355 were analyzed here. These plasmid encoded either 2, 3 or 4 shufflon segments (see Table 1 for details). The number of variants measured per plasmid was dependent on the number of shufflon

segments that the plasmid had at the start of the experiment (Fig. 4a). pESBL-4 and pESBL-12 encoded 4 segments and had 174–275 shufflon variants regardless of the host cell, pESBL-117 and pESBL-283 encoded 3 segments and had 113–141 variants, whereas pESBL-355 only encoded 2 segments and 62–69 variants were measured. The number of variants for pESBL-4 and pESBL-283 in *E. coli* DH10B were similar to those measured in the first experiment at the first time point.

The number of shufflon segments that was measured per read was similar for pESBL-4 in *E. coli* DH10B as measured in the first experiment, but for pESBL-283, this went from 57–66% of sequences containing 3 segments to 38% (Fig. 4b). However, this is comparable to the number of reads containing all 3 segments for pESBL-283 in the native cell (41%). For pESBL-12 in DH10B, 39% of reads contained 4 segments compared to 64% in the native cell, although this might be attributed to natural variation and the lack of further replicate experiments.

Analysis of the *pilV* ORFs that were formed in the shufflons showed little variation for the host cell of the plasmids (Fig. 4c). The ORFs of the A segment were present downstream of *pilV* in 48–79% of reads. In all samples of both experiments, there was some bias for *pilV-A'* (52–64%) over *pilV-A*. None of the other segments showed such a preference of either side of the segment that was conserved in all plasmids and host cells that were measured except for segment D which only encodes a single partial ORF, Fig. 3d. Nonetheless, *pilV-Dx* was detected in approximately 2.5% of reads for pESBL-4 in *E. coli* DH10B in both experiments. However, in experiment 2, for pESBL-4 in the native *E. coli* host and for pESBL-117 in the native host or *E. coli* DH10B, *pilV-Dx* was found only in 0.5% of reads.

The most highly abundant variants of the complete shufflons was analysed per plasmid, comparing the two different host cells and the results of the earlier experiment (Figs. 3d and 4d). Little commonality was detected for pESBL-4 in the native *E. coli* or *E. coli* DH10B; however, between the first and second experiment, an overrepresentation of variants A + B + C + D- and A-B + C + D- was seen. Although it can be argued that this might represent the shufflon orientation of the original transformant, both experiments were initially started from individual colonies of separate agar plates, making this scenario less

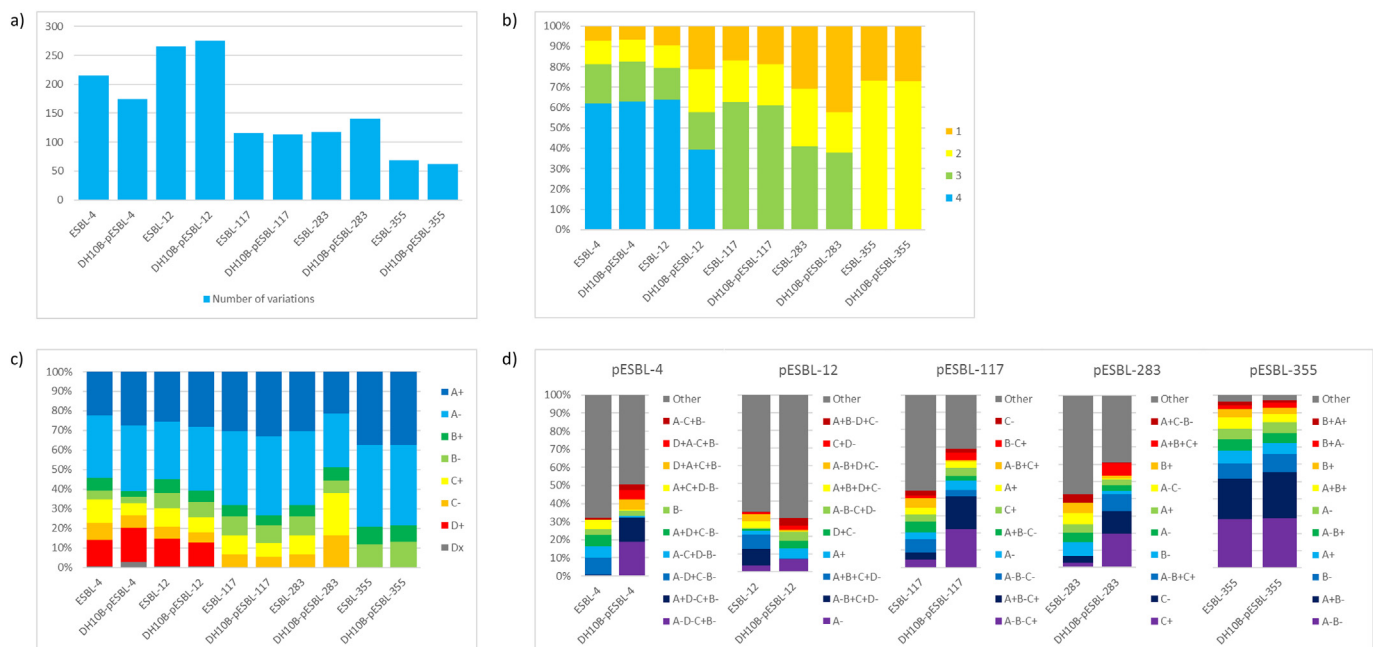


Fig. 4. Analysis of shufflon rearrangements in different hosts. Shufflon rearrangement was measured in plasmids pESBL-4, pESBL-12, pESBL-117, pESBL-283 and pESBL-355 measured either in the original *E. coli* isolate or after transformation into *E. coli* DH10B. (a) Blue bars show the number of shufflon variations measured. (b) Counts of the number of shufflon segments that are present in the sequencing reads. (c) Relative prevalence of each of the PilV ORFs that were formed. (d) Relative distribution of the 10 most abundant shufflon variants, calculated per plasmid for both host cell types. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

likely.

For pESBL-117, pESBL-283 and specifically pESBL-355, some more commonality between the data from the different hosts can be seen, which is at least partially caused by the decreased amount of variability when the experiment is started with less than 4 shufflon segments.

4. Conclusions

The results presented here demonstrate that the recombination of the shufflon is not a random process. This data on IncI1 plasmids show much similarity with the results previously presented for three IncI2 plasmids from a study which utilized a different long-read sequencing platform, PacBio, and directly sequenced complete plasmids (Sekizuka et al., 2017). As such, the coverage of the shufflon is much lower (96–134 reads per plasmid compared to 2100 amplicon reads analyzed here) but they also reported predominance of certain shufflon variants. Due to the difference in sequences of the segments between IncI1 and IncI2 plasmids, it is difficult to compare the predominant shufflon variants between these two studies. Nonetheless, the *pilV* ORFs *pilV-A* and *pilV-A'* were overrepresented in each of the plasmids of the two studies that contained segment A.

An unexpected finding in the analysis of our Nanopore data is the loss of shufflon segments that was detected. However, *in vitro* experiments using genetic fragments flanked by shufflon *sfx* sites previously reported low levels of loss of artificial shufflon segments when symmetric *sfx* sites were present (Gyohda et al., 2006). Here, we did not check for the symmetry of the *sfx* sites as the quality of the nanopore reads is not sufficient. *In vivo*, variants of the shufflon that have lost one or multiple segments have been described which may come at a fitness cost, depending on the environment in which the plasmid is present (Brouwer et al., 2015). The second unexpected finding was the lack of difference that was found at the different time points of the first experiment which were chosen to represent several growth stages and stress factors. Nonetheless, we could not observe differences in terms of recombination activity of the shufflon, something that could be investigated further using molecular reporter assays.

Conjugation is a major contributor to the spread of antimicrobial resistance genes, for which plasmids of the IncI complex (IncB/O/K/Z/IncI1/IncI2) contribute greatly in certain environments (Rozwandowicz et al., 2018). As deletion of *pilV* has been shown to significantly reduce the conjugal transfer of IncI1 in liquid environments, the shufflon system plays an important role for the selection of recipient cells (Kim and Komano, 1989). Limiting the transfer of AMR by reducing conjugation seems like an attractive strategy, but to determine if the shufflon is a suitable target, it is necessary to determine: 1) if and how the shufflon recombinase Rci is regulated, and 2) if the bias for certain shufflon variants and *PilV* ORFs is random or reproducible.

The method we have used here of amplifying large genomic regions and sequencing using long-read NGS will also be suitable for the

analysis of large numbers of DNA molecules of other multiple inversion systems. Analysis of large numbers of amplicons can indicate well what the distribution is within a heterogeneous population.

Acknowledgements

Financial support for this research came from the European Union's Horizon 2020 program under grant agreement No 773830, OH-EJP-H2020-JRP-AMR-2-ARDIG.

References

- Atack, J.M., et al., 2015. A biphasic epigenetic switch controls immunoevasion, virulence and niche adaptation in non-typeable *Haemophilus influenzae*. *Nat. Commun.* 6, 7828.
- Brouwer, M.S., et al., 2014. Complete genome sequences of IncI1 plasmids carrying extended-spectrum beta-lactamase genes. *Genome Announc.* 2.
- Brouwer, M.S., et al., 2015. IncI shufflons: assembly issues in the next-generation sequencing era. *Plasmid* 80, 111–117.
- de los Rios, M., et al., 2015. Structural and genetic diversity in antibody repertoires from diverse species. *Curr. Opin. Struct. Biol.* 33, 27–41.
- Emerson, J.E., et al., 2009. A novel genetic switch controls phase variable expression of CwpV, a *Clostridium difficile* cell wall protein. *Mol. Microbiol.* 74, 541–556.
- Gyohda, A., et al., 2002. Sequence-specific and non-specific binding of the Rci protein to the asymmetric recombination sites of the R64 shufflon. *J. Mol. Biol.* 318, 975–983.
- Gyohda, A., et al., 2006. Asymmetry of shufflon-specific recombination sites in plasmid R64 inhibits recombination between direct *sfx* sequences. *J. Biol. Chem.* 281, 20772–20779.
- Kim, S.R., Komano, T., 1989. Cloning and nucleotide sequence of the ColIb shufflon. *Plasmid* 22, 180–184.
- Komano, T., 1999. Shufflons: multiple inversion systems and integrons. *Annu. Rev. Genet.* 33, 171–191.
- Komano, T., et al., 1987. Distribution of shufflon among IncI plasmids. *J. Bacteriol.* 169, 5317–5319.
- Komano, T., et al., 1995. Mating variation by DNA inversions of shufflon in plasmid R64. *Adv. Biophys.* 31, 181–193.
- R-Core-Team, 2014. R: A Language and Environment for Statistical Computing. <https://www.R-project.org>.
- Roux, N., et al., 2012. Neglected but amazingly diverse type IVb pili. *Res. Microbiol.* 163, 659–673.
- Rozwandowicz, M., et al., 2018. Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. *J. Antimicrob. Chemother.* 73, 1121–1137.
- Sandmeier, H., et al., 1991. Gene organization in the multiple DNA inversion region min of plasmid p15B of *E. coli* 15T: assemblage of a variable gene. *Nucleic Acids Res.* 19, 5831–5838.
- Seiffert, S.N., et al., 2017. Plasmids carrying bla_{CMY}-2/4 in *Escherichia coli* from poultry, poultry meat, and humans belong to a novel IncK subgroup designated IncK2. *Front. Microbiol.* 8, 407.
- Sekizuka, T., et al., 2017. Elucidation of quantitative structural diversity of remarkable rearrangement regions, shufflons, in IncI2 plasmids. *Sci. Rep.* 7, 928.
- Smith, H., et al., 2015. Characterization of epidemic IncI1- γ plasmids harboring ambler class A and C genes in *Escherichia coli* and salmonella enterica from animals and humans. *Antimicrob. Agents Chemother.* 59, 5357–5365.
- Venturini, C., et al., 2013. Sequences of two related multiple antibiotic resistance virulence plasmids sharing a unique IS26-related molecular signature isolated from different *Escherichia coli* pathotypes from different hosts. *PLoS One* 8, e78862.
- Waite, R.D., et al., 2001. Spontaneous sequence duplication within an open reading frame of the pneumococcal type 3 capsule locus causes high-frequency phase variation. *Mol. Microbiol.* 42, 1223–1232.
- Wick, R. Porechop: Adapter Trimmer for Oxford Nanopore Reads. <https://github.com/rwrick/Porechop>.