Published in final edited form as:

Mol Ecol. 2017 April ; 26(7): 1832–1847. doi:10.1111/mec.14056.

Plasmid and clonal interference during post-horizontal gene transfer evolution

S. Bedhomme^{1,4}, D. Perez Pantoja^{2,3}, and I. G. Bravo⁴

¹Centre d'Ecologie Fonctionnelle et Evolutive, UMR CNRS 5175, 34293 Montpellier, France

²Department of Biochemistry and Molecular Biology, Faculty of Biological Sciences, Universidad de Concepción, Chile

³Programa Institucional de Fomento a la I+D+i (PIDi), Universidad Tecnológica Metropolitana, Santiago, Chile

⁴Laboratory MIVEGEC, UMR CNRS 5290, IRD 224, UM, Centre National de la Recherche Scientifique, Montpellier, France

Abstract

Plasmids are nucleic acid molecules that can drive their own replication in a living cell. They can be transmitted horizontally and can thrive in the host cell to high copy numbers. Plasmid replication and gene expression consume cellular resources and cells carrying plasmids incur fitness costs. But many plasmids carry genes that can be beneficial under certain conditions, allowing the cell to endure in the presence of antibiotics, toxins, competitors or parasites. Horizontal transfer of plasmid-encoded genes can thus instantaneously confer differential adaptation to local or transient selection conditions. This conflict between cellular fitness and plasmid spread sets the scene for multilevel selection processes.

We have engineered a system to study the short term evolutionary impact of different synonymous versions of a plasmid-encoded antibiotic resistance gene. Applying experimental evolution under different selection conditions and deep sequencing allowed us to show rapid local adaptation to the presence of antibiotic and to the specific version of the resistance gene transferred. We describe the presence of clonal interference at two different levels: at the within-cell level, because a single cell can carry several plasmids, and at the between-cell level, because a bacterial population may contain several clones carrying different plasmids and displaying different fitness in the presence|absence of antibiotic.

Understanding the within-cell and between-cell dynamics of plasmids after horizontal gene transfer is essential to unravel the dense network of mobile elements underlying the worldwide threat to public health of antibiotic resistance.

⁴Corresponding author: Stéphanie Bedhomme, Centre d'Ecologie Fonctionnelle et Evolutive, UMR CNRS 5175, 34293 Montpellier, France, Fax number: 00 33 (0)4 67 61 33 36, stephanie.bedhomme@cefe.cnrs.fr.

Data Accessibility:

The full dataset of illumina reads used in this study is deposited in Data Dryad (www.datadryad.org) at <http://dx.doi.org/10.5061/dryad.8m3t6>.

Author contributions: SB and IGB designed the study; SB produced and analysed the data; SB and DPP analysed NGS data; SB and IGB drafted the manuscript.

Keywords

codon usage preferences; plasmid; antibiotic resistance; clonal interference; multi-level selection

Introduction

Plasmids are replicative extrachromosomal DNA elements present in most bacteria. They can be transmitted either vertically or horizontally. Horizontal transfer occurs by conjugation for conjugative plasmids, by transformation of competent bacteria or transduction in certain cases for non-conjugative plasmids. From the plasmid's point of view, genetic information carried on a plasmid can be divided in essential gene(s), *i.e.* those necessary for plasmid replication, and non-essential genes. Non-essential genes in plasmids are very often also found on the chromosome in other species or lineages, meaning that they are not plasmid-specific. However, they are not a random sample of bacterial chromosomal genes. Instead it is long known that certain functions are over-represented in non-essential plasmid genes, such as antibiotic resistance, detoxification, virulence, catabolism of uncommon metabolites, or capacity to invade certain tissues. As noted by Eberhard, these functions most often correspond to "local adaptation to environmental conditions occurring sporadically in time and space" (Eberhard 1990). This observation led him to formulate the *local adaptation hypothesis* under which "this sporadic selection makes the maintenance of this local adaptation more likely on the plasmid than on the chromosome", basically thanks to increased horizontal mobility and higher copy number of plasmid-resident genes compared to chromosome-resident genes. In this study, we will focus on the local adaptation instantaneously conferred by a plasmid-resident antibiotic resistance gene upon transformation, and on the short-term evolutionary events that follow this transformation in conditions where the transferred gene is either beneficial or neutral/slightly deleterious.

Plasmids and more generally horizontal gene transfer (HGT) play a key role in the propagation of antibiotic resistances as shown by the occurrence of homologous antibiotic resistance genes on plasmids and/or on the chromosome of different lineages within a species, *e.g.* in *Staphylococcus aureus* (Lindsay 2014) or in different species, across the whole phylum of Firmicutes (Fernández Lanza *et al.* 2015). This last study also provides a full picture of the diversity of circulating plasmids and confirms on a large scale that carrying plasmid(s) is the rule rather than the exception in bacteria. This widespread presence of plasmids actually constitutes an apparent paradox, because it has been widely demonstrated that carrying a plasmid has a cost for the bacteria (Bouma & Lenski 1988). This carriage cost is directly linked to the number of copies of the plasmid (Harrison *et al.* 2012) and seems to be due more to the expression of plasmid genes than to the replication of the plasmid itself (Bragg & Wagner 2009, Lynch & Marinov 2015). Additionally to this metabolic cost, the presence of plasmids leads to multilevel selection as well as to a potential genomic conflict over the control of plasmid replication (Paulsson 2002): increasing replication favours the plasmid at the intracellular level but is likely to reduce the fitness of the cell, because of the cost of carriage. These plasmid costs led to the prediction (Bergstrom *et al.* 2000) that the outcome of any new bacteria-plasmid association should be either the elimination of the plasmid if the carried genes are not beneficial in the environmental

conditions, or the integration of the plasmid gene(s) in the bacterial chromosome if they are beneficial, thus retaining the benefit of these genes without having to pay the costs of plasmid carriage. Bergstrom and coworkers (2000) derived a mathematical model showing that the conditions for plasmid existence are very limited. However, this study considers the dynamics of plasmids and plasmid-carried genes in a model where variables such as the plasmid carriage cost or the probability of segregational loss have fixed values. It does not include the possibility of compensatory evolution after plasmid transfer. Diverse experimental studies (Bouma & Lenski 1988, Modi & Adams 1991, Dahlberg & Chao 2003, Dionisio *et al.* 2005), reviewed in Harrison and Brockhurst (2012), have shown that under selection pressures favouring the maintenance of plasmid genes, co-evolutionary processes occur that decrease plasmid carriage cost and increase plasmid stability. More recently, the nature of this compensatory evolution has been revealed by sequencing approaches: carriage cost reduction can be obtained through mutations in the plasmid replication machinery (Sota *et al.* 2010), by deletion of costly genes from the plasmid (Porse *et al.* 2016) or through mutations in global regulators that counteract the up-regulation of chromosomal genes by the plasmid and also reduce plasmid gene expression (Harrison *et al.* 2015, San Millan *et al.* 2015). Theoretical models integrating the effects of compensatory evolution have been built (San Millan *et al.* 2014, Peña-Miller *et al.* 2015) and predict that it contributes to plasmid stability improvement, but they are not sufficient to explain the ubiquitous presence of plasmids in most unicellular organisms.

Plasmid carriage costs are not the only costs to plasmid-mediated HGT. Another potentially important one is the mismatch in codon usage preferences between the transferred gene and the receiving genome (Baltrus 2013). Each species is indeed characterized by specific frequencies of use of the different codons within a synonymous codon family. An important evolutionary force shaping codon usage preferences is coevolution with the translation machinery: codon usage frequencies are strongly related with the copy number of corresponding tRNA genes, especially for highly expressed genes and for rapidly growing organisms (Rocha 2004). Genes transferred horizontally from a different organism can present a mismatch in codon usage preferences with the receiving genome, and such mismatch is known to affect translation accuracy and speed (Komar *et al.* 1999, Burgess-Brown *et al.* 2008). The combination of these two deleterious effects on protein synthesis results in the production of low quantities of functional proteins but also incurs important cellular costs to degrade non-functional proteins, either because of translation errors or incorrect folding (Drummond & Wilke 2008). Additionally, slow translation leads to ribosomal sequestration and can affect the translation level of other proteins (Plotkin & Kudla 2011). These mechanistic effects directly convert into evolutionary constraints: experimental studies show large variations in protein production and resulting fitness depending on the synonymous version of a key metabolic introduced gene (Agashe *et al.* 2013) and comparative approaches demonstrate that HGT is more likely to be successful between genomes with similar codon usage preferences (Tuller *et al.* 2011, Medrano-Soto 2004).

In an earlier study (Amoros-Moya *et al.* 2010), we mimicked the horizontal transfer of synonymous versions of an antibiotic resistance gene (*cat*, chloramphenicol acetyl transferase) *via* plasmid transfer in *Escherichia coli*. The cost of plasmid carriage was

quantified and we showed that mismatches in codon usage preferences generate differences in chloramphenicol resistance. These differences were totally erased after 400 generations of experimental evolution in presence of chloramphenicol, with all populations presenting equivalent growth rates and reaching equivalent high bacterial densities. Additionally, we established that both genomic compartments (plasmid and bacterial chromosome) contributed to the observed changes in fitness and resistance. Finally, a Sanger consensus sequence approach restricted to the *cat* gene and to its immediate 5'-untranslated regulatory sequence revealed the absence of mutations within the coding sequence and the presence of small indels upstream the start codon of the introduced gene.

In the present work, we have gone beyond the previous analyses of our experimental evolution system, addressing novel research questions focused on the fine-scale dynamic of plasmid sequence evolution. To do so, we have revisited the original material from the earlier experimental evolution approach (Amoros-Moya *et al.* 2010) with newer sequencing technology and have continued the evolution experiment beyond 1000 generations. The first goal was to identify mutations occurring in the entire plasmid sequence, including mutations present in low frequency and never reaching fixation. The specific questions asked here were: (i) Are there mutations outside the introduced gene, for example mutations affecting plasmid gene content or replication? (ii) Can we identify, even if at low frequency, synonymous mutations in the *cat* coding sequence that would bring the codon usage preferences closer to those of *E. coli*? The second goal was to decompose the diversity of the plasmid population: in the case where various mutants segregate in the population, is the diversity obtained by a mix of different monomorphic cell lineages or is there diversity within cell lineages? The third goal was to look for the signature of clonal interference (Gerrish & Lenski 1998) at two levels: between cells carrying different mutations on the plasmid and within cells between different mutants of the plasmid.

Materials and methods

Initial bacteria and plasmids

An experimental evolution approach had been started earlier (Amoros-Moya *et al.* 2010) and was continued in the present work to reach 1000 generations. The experiment had originally been started with three synonymous versions of the chloramphenicol acetyl transferase (*cat*) gene (EC 2.3.1.28) were designed and synthesized. For the first version, further named *cat*-Opt (EU751288), we chose for each amino acid the most-frequently used synonymous codon, in a set of highly expressed genes in *E. coli*, as defined by Puigbó and coworkers (2008). (<http://genomes.urv.cat/HEG-DB/>). For the second version, further named *cat*-GC (EU751289), we chose for each amino acid the GC-richest codon among the less-frequently used synonymous codons; and for the third version, further named *cat*-AT (EU751290), we chose for each amino acid the AT-richest codon among the less-frequently used synonymous codons (figure 1). These three synonymous versions were cloned in the same position into pUC57 by insertion in the multiple cloning site. The inserted gene is under the control of the *lac* promoter. The pUC57 plasmid also carries the *bla* gene, conferring ampicillin resistance. Each of the three constructs was transformed into *E. coli* Top10. In this *E. coli* strain, the *lac* repressor is deleted, so that *cat* gene expression is constitutive.

Experimental evolution

From each of the three initial samples, resulting from the transformations into *E. coli* of one of the three different synonymous version of the *cat* gene, six populations had been derived and experimentally evolved, three in presence of 25 $\mu\text{g.mL}^{-1}$ of chloramphenicol and three in presence of 20 $\mu\text{g.mL}^{-1}$ of ampicillin (figure 1). The chloramphenicol treatment corresponded to a strong selection pressure for a high chloramphenicol resistance. The ampicillin treatment had been conceived as a control treatment with a selection pressure to maintain the plasmid in the bacteria but without direct selection pressure on the *cat* gene. Each of the 18 populations was grown in 10 mL LB at 37 °C and 200 rpm orbital shaking, and serially transferred every day with a 1:100 dilution. The experimental evolution was maintained for 1000 generations (around 150 transfers).

Resistance evolution, plasmid and bacterial chromosome contribution

Bacterial growth of the three initial populations, a mock population (*E. coli* Top10 bacteria without plasmid) and the 18 evolved populations (at generation 1000) was measured across a range of chloramphenicol concentrations ([Cam] = 0, 5, 10, 15, 25, or 50 $\mu\text{g mL}^{-1}$) and across a range of ampicillin concentrations ([Amp] = 0, 5, 10, 20, 40, or 100 $\mu\text{g mL}^{-1}$). Growth measures were performed in a 96-well plate, with 200 μL Lysogeny Broth (LB) medium per well, incubated for 20 h at 37°C in a thermostated plate reader (TECAN Infinite 200; Tecan Group Ltd., Crailsheim, Germany), monitoring absorbance at 600 nm (OD600) every 30 min and automatically shaking the plate for 5 min before each measurement. The reference untransformed bacteria (mock) were grown in quadruplicate in each plate, without antibiotic, and all values in the plate were normalized with respect to the median value of these reference wells. For each growth assay, an overnight bacterial preinoculum was performed from a -20°C glycerol stock, and the bacteria were inoculated in the wells with an initial OD600 of 0.015. Each plate contained two times the same treatment, and the average of the two wells was taken and used as a data point in the analysis. Each population was measured independently from separate preinocula and in independent plates a minimum of four times. Measures in the absence of antibiotic were performed in eight independent preinocula and plates. From each growth curve, we extracted the maximum population density reached (maxOD) and the maximum growth rate (maxGR).

For each evolved population, at generation 1000, the evolved plasmids were extracted from a subculture of each population and transformed into the ancestral bacteria (*E. coli* Top10), thus generating populations combining the ancestral bacterial chromosome and the evolved plasmids. Growth curves of these transformed populations were obtained in a similar way as for evolved populations, both in absence of antibiotics and in presence of 25 $\mu\text{g.mL}^{-1}$ of chloramphenicol or 20 $\mu\text{g.mL}^{-1}$ of ampicillin (experimental evolution concentrations). For each of the seven populations sequenced (see below), we conducted an ANOVA with “antibiotic” (no antibiotic, ampicillin, chloramphenicol), “cell-plasmid combination” (c0-p0, c0-p1000, c1000-p1000) and their interaction as fixed factors and “max OD” as the variable. To further identify the genomic compartment responsible for the phenotypic changes, a Tukey-HSD test was performed.

CAT and BLA protein quantification through i-TRAQ analysis

Proteins were quantified using isobaric Tags for Relative and Absolute Quantitation (iTRAQ). This technique allows for the relative quantification of expressed proteins between one reference sample and problem samples. The three initial populations were cultured in 20 $\mu\text{g.mL}^{-1}$ ampicillin and the 18 evolved populations were cultured in their conditions of evolution. For evolved populations, the optical densities of the cultures of the three replicates of each (gene version * conditions of evolution) were equalized and cultures were pooled. Proteins were extracted, trypsin-digested and the N-termini were covalently labeled with mass-reporters. Samples were then pooled, fractionated by nano liquid chromatography and analyzed by tandem mass spectrometry (MS/MS). The fragmentation data were then submitted to a database search to identify the labeled peptides and hence the corresponding proteins. A detailed procedure is given in the supplemental material. The *cat*-ATg0 population was used as reference for the ratio calculation. The analysis presented here will focus on the two resistance proteins BLA and CAT.

Plasmid deep-sequencing

Based on growth characteristics and Sanger sequencing of the introduced gene between generations 0 and 358 (Amoros-Moya *et al.* 2010), a subset of populations was chosen to perform deep sequencing of the plasmid at generations 173, 358, 558 and 1000. For each of these time points, the plasmids of the following seven populations were deep-sequenced (figure 1): replicates 1 and 2 carrying *cat*-Opt and evolved in chloramphenicol (further named OptCam1 and OptCam2), replicate 1 carrying *cat*-GC and evolved in ampicillin (further named GCamp1), replicates 1 and 2 carrying *cat*-GC and evolved in chloramphenicol (further named GCCam1 and GCCam2) and replicates 1 and 2 carrying *cat*-AT and evolved in chloramphenicol (further named ATCam1 and ATCam2). Additionally, for OptCam1, GCamp1 and ATCam1, five clones were randomly picked and individually deep-sequenced. For population samples, an aliquot of the glycerol archive was transferred to 4 mL LB-antibiotic and bacteria were grown overnight at 37°C. For clone samples, an aliquot of the glycerol archive was plated on an LB-antibiotic agar plate. The next day, five colonies were picked, independently transferred to 4 mL LB-antibiotic, and bacteria were grown overnight at 37°C. Liquid cultures and plates contained 25 $\mu\text{g.mL}^{-1}$ chloramphenicol for populations evolved in chloramphenicol and 20 $\mu\text{g.mL}^{-1}$ ampicillin for populations evolved in ampicillin.

Total plasmid DNA was extracted from the liquid cultures using a Zymoclean plasmid extraction kit following manufacturer's instructions. Plasmid DNA was deep sequenced at Genoscreen (Lille, France). Libraries were produced following a Nextera XT protocol. DNA was broken into 50-500bp fragments thanks to transposons, and adapters were automatically ligated to generated ends. After size selection (100bp) and amplification, the libraries were controlled by loading on Labchip GX (Caliper). Single-read sequencing was then performed with HiSeq2000 technology. Sequencing of plasmids in individual clones allowed to address the question of segregation of different plasmids within a cell. Our protocol for plasmid sequencing from individual clones includes two growth steps, *i.e.* colony isolation and subsequent amplification. Strictly speaking, it does not allow to sequence the plasmidic content of a single cell, but rather of a nearly clonal bacterial population: growth of a large

bacterial colony on a plate from a single cell requires ca. 30 generations (Reams & Roth 2015), and subculturing of 1/10 of such a colony to saturate a 3 mL bacterial liquid culture requires five additional generations. We therefore cannot exclude that evolution through changes in plasmid copy number and/or generation of mutations may have occurred during the cell divisions leading from the cell starting the colony on the plate and the bacteria population from which plasmid DNA is extracted. However, this protocol is common practice to isolate and sequence clones from a population, the additional cellular divisions during which diversity could have been generated were conducted under the same antibiotic selection pressure as the populations from which the clones were derived, and the results have been analysed with this potential confounding factor in mind. Indeed, this additional amplification step is common to all samples used for the analysis of genetic diversity in nearly-clonal populations, so that the putative systematic bias thereby introduced should have not lead to differential accumulation of mutations for different cell lines.

Sequence read quality control, processing, mapping and variant calling

Quality control checks were performed on the raw FASTQ data using FastQC (version 0.10.1) (Andrews 2010). Sequencing reads were trimmed for sequencing adaptors using Trimmomatic (version 0.32) (Bolger *et al.* 2014) and the quality filtering and trimming was done by Prinseq-lite (version 0.20.4) (Schmieder & Edwards 2011). Briefly reads were trimmed for 'N' characters and low quality nucleotides (Phred score cutoff of 24) and later any read with an average Phred score below 32 and shorter than 90 nt was discarded. A random subset of 1 million filtered-in reads of each sample were mapped against the reference plasmid sequence by the Burrows-Wheeler algorithm using bwa-mem aligner (Li & Durbin 2009) with standard settings and stored as bam files. Overall the mean coverage depth was ~24,000X in each sample. For conversions from sam to bam files the samtools suite was used (Li *et al.* 2009). Nucleotide variants in each population were identified with the LoFreq (Wilm *et al.* 2012) variant calling algorithm using default options. The same program was used to determine the frequency of the different nucleotides for each position and to evaluate the reliability of the haplotypes. Only variants supported by at least eight reads were kept as a valid haplotype.

As most mutations detected with Lofreq mapped to a short sequence around the start codon of the *cat* gene and the coverage was high enough, it was possible to identify in this sequence stretch, whether two mutations identified by LoFreq in the same sample were actually carried by the same plasmid copy or not. For this purpose, we used the shell *grep* function to identify reads supporting haplotypes containing one or two different variants. The sequence stretch between positions 2398 and 2474 of reference plasmid was scanned for all possible variant combinations and the number of reads supporting each haplotype was counted using the *wc* function. Frequencies based in counting of reads supporting all possible haplotypes were determined. By using this approach the depth coverage of reads encompassing the scanned sequence stretch was ~3000 – 3500X in each sample.

Finally, the coverage data of all samples was analysed to detect potential duplications or deletions in the plasmid during experimental evolution.

Results

Evolution of resistance and relative contributions of plasmid and bacterial chromosome.

IC₅₀ (antibiotic concentration reducing by half antibiotic-free growth characteristics of a population) of chloramphenicol and ampicillin (and the confidence interval around them) were calculated for the ancestral and evolved populations from their growth characteristics across the antibiotic gradient. Chloramphenicol IC₅₀ data are presented in figure 2a. The higher IC₅₀ for Optg0 than for GCg0, which is itself higher than for ATg0 reflects the initial differential cost of codon usage preferences mismatch. These differences have been totally erased by generation 1000 for populations evolved in chloramphenicol with all of them showing a high IC₅₀ for both growth characteristics. This means that population evolved in chloramphenicol had recovered a high chloramphenicol resistance whatever the version of the chloramphenicol resistance gene was. For populations evolved in ampicillin, IC₅₀ values for chloramphenicol are lower than for populations at generation 0, indicating a loss of chloramphenicol resistance. All populations grew equally well in ampicillin, whatever the concentration was, such that it was impossible to fit a log-linear decrease. This suggested that all populations initially had and conserved a strong ampicillin resistance. A liquid medium determination confirmed that all populations were indeed able to grow at ampicillin concentration of 256 µg.mL⁻¹.

To get insight in the temporal dimension of chloramphenicol resistance, the maximum bacterial density reached in presence of 25 µg.mL⁻¹ of chloramphenicol (experimental evolution concentration) by the ancestral and 18 experimentally evolved populations is displayed on figure 2b. Values of maximum density for the ancestral populations (generation 0) confirmed the cost of mismatch in codon usage. Populations evolved in chloramphenicol had compensated the initial cost by generation 400 as there was no significant difference in bacterial growth between the populations carrying the three versions of the gene, and they actually grew better than the mock ancestral population in absence of chloramphenicol. These characteristics persisted at generation 1000. Populations evolved in ampicillin progressively lost the ability to grow in presence of chloramphenicol: at generation 400, all populations displayed a reduced maximum density compared to their respective ancestor, but significant differences remained between the populations carrying the three versions of the gene. At generation 1000, none of the populations evolved in ampicillin was able to grow in chloramphenicol.

Transformation of ancestral bacteria with plasmids retrieved from evolved populations at generation 1000 allowed determining the localisation of compensatory evolution in either the evolved plasmid or the evolved bacterial chromosome. For all populations, both “antibiotics” and “cell-plasmid combination” had a significant effect ($p < 0.001$ in all cases). The interaction was also significant except for population OptCam1 ($p = 0.084$). Tukey HSD tests allowed to determine the significant differences between conditions and are reported extensively in supplemental material. For populations carrying *cat-AT*- or *cat-GC*, initially paying a strong cost of codon usage mismatch, compensatory evolution had occurred in both the chromosome and the plasmid compartments: transformed bacteria had an intermediate profile between the ones of ancestral and evolved bacteria (figure 3), although the difference

between ancestral and transformed cells was only significant for populations carrying the AT version of the gene. For populations carrying the *cat*-Opt gene, the plasmid played little role if any in compensatory evolution and even have a negative effect for OptCam2, as the transformed bacteria grew worse in chloramphenicol than the ancestor (figure 3). For GCamp1, the transformed bacteria were unable to grow in chloramphenicol (figure 3).

CAT and BLA protein quantification

Quantities of CAT and BLA determined by the iTRAQ method for ancestral and evolved populations are given in table 2. Globally, populations evolved in chloramphenicol produced significantly higher quantities of CAT than ATg0 and similar quantities of BLA. Populations evolved in ampicillin tended to produce a lower quantity of CAT and higher quantities of BLA.

Deep-sequencing results

Deep-sequencing of the plasmids revealed that most mutations detected were located upstream of the start codon of the introduced gene (table 1). Very few mutations were observed in the coding sequence of the *cat* gene, and those occurring were not mutations shifting the codon usage of the introduced gene closer to that of *E. coli*, as they are mainly 1bp indels. The only mutation recorded outside the *cat* gene and its vicinity is a mutation in the *rep* gene, which was found only in one clone of OptCam1 and does not reach a high frequency.

Because most mutations detected were located in a restricted zone and that some samples harboured two or more of these mutations, an additional analysis was conducted to determine whether the mutations were located on the same read or on different reads. The goal was to identify the haplotype(s) (over a 77bp window) present in a population or in a cell (for virtually clonal samples) at a time point and thus determine whether clonal interference occurred. Figure 4 presents the proportions of the different haplotypes for the defined range and for the different populations and clones along time. First, this graphical representation clearly shows (see also table 1) that most mutations are population-specific, although one mutation (insertion of ATCCA in position 2422 of the ancestral plasmid) appeared independently in four different populations (carrying *cat*-GC or *cat*-Opt). Second, the ancestral haplotype was not detected in any of the evolved samples in populations carrying the *cat*-AT gene; it was present until generation 358 but not later in populations carrying the *cat*-GC gene; remained present at generation 1000 in population OptCam1; and was present until generation 558 in population OptCam2. Additionally, two independent mutations occurred very rarely on the same read, (figure 4, see inset colour code). The vast majority of haplotypes carried instead only one mutation, compared to the wild-type. This remained true when more than one mutation were found both in a population sample and in a clone sample.

Sharp differences in coverage along the plasmid allowed to determine that in GCamp1 samples (populations and clones) 1763 bp were deleted spanning the *cat* and *lacZ* genes, leaving the *bla* and *rep* genes intact. The proportion of reduced plasmid was calculated from the median coverage of the conserved and the deleted parts of the plasmid. Deletion of the

cat gene in the absence of selection pressure occurred very swiftly (figure 5): at generation 173, the reduced plasmid already represented 94% of the plasmids copies at the population level, with proportions in clones ranging from 67.29 to 99.97% strongly suggesting co-existence in a number of individual bacteria of the wild-type and the reduced plasmid. In subsequent generations, the reduced plasmid had been virtually fixed and the ancestral plasmid could not be detected.

Discussion

We describe here the results of rapid local adaptation to both the presence of antibiotic and to specific synonymous versions of an antibiotic resistance gene using artificial HGT followed by experimental evolution. The data here collected inform our understanding about the evolution of plasmids and provide exciting new results about the multilevel selection processes and adaptive dynamics occurring in evolving bacterial populations carrying plasmids.

Our experimental design aimed at simulating the outcome of horizontal transfer of antibiotic resistance genes with different levels of codon usage mismatch and presence or absence of the antibiotic. Our results demonstrate that phenotypic adaptation to the new conditions occurs rapidly. The cost of codon usage mismatch was important but chloramphenicol resistance recovery swiftly erased the initial differences in resistance conferred by the synonymous versions of the transferred resistance gene. Populations selected in ampicillin, on the contrary, lost the partial chloramphenicol resistance (figure 2). Transformation of ancestral bacteria with the evolved plasmids revealed that for populations evolved in chloramphenicol, there was contribution of both plasmid and bacteria chromosome to the chloramphenicol resistance improvement. This means that adaptation to the specific situation of an antibiotic resistance gene with a codon usage mismatch was achieved by changes occurring both *in cis* in the plasmid, mostly in the vicinity of the introduced gene, as well as *in trans*, on the chromosome. Other plasmid-cell coevolution experiments conducted in different contexts often lead to similar results with contribution of the two compartments to fitness improvement (Harrison & Brockhurst 2012). In this study we have focused further on the changes occurring *in cis* in the plasmid and their dynamics along experimental evolution.

The most conspicuous results of our experimental evolution are the streamlining changes modifying plasmid gene content in a population selected in the presence of ampicillin. The ablation of 1673 bp spanning the *cat* gene in populations initially carrying *cat*-GC (data presented here for replicate 1; replicates 2 and 3 are also affected) clearly fits with the prediction of changes in plasmid gene content to eliminate unnecessary genes. The reduced plasmid selected by experimental evolution in the absence of chloramphenicol only carries one accessory gene, *bla*, which is necessary to confer resistance to ampicillin. The loss of the *cat* gene explains the total absence of resistance to chloramphenicol for these populations and also for ancestral bacteria transformed with the evolved plasmid. The coverage analysis revealed that plasmid trimming occurred quite early during experimental evolution, as the reduced plasmid already represented a large majority of the plasmid population at generation 173 and was fixed at any further analysed generations (figure 5). Given the rapid spread of

the reduced plasmids it is striking that ablation of the *cat* gene has occurred exclusively in populations carrying the *cat*-GC version of the gene, because no differences in bacterial growth in ampicillin were found among the initial populations carrying the three versions of the *cat* gene (Amoros-Moya *et al.* 2010). Also, it is noticeable that the *bla* gene has not been lost in any of the lineages selected in the presence of chloramphenicol, *i.e.* under selection conditions that do not require maintenance of the ampicillin resistance gene. One peculiarity of this GC-rich version of the *cat* gene is its strong structure. Indeed, the large differences in nucleotide composition between the three synonymous versions of the *cat* gene cause ample variation in the stability of the associated secondary structures for DNA: the calculated free energy for the most stable DNA secondary structure of *cat*-GC (-131.8 kcal/mol) is 50% higher than for *cat*-OPT and 2.2-times higher than for *cat*-AT. Such strong structures, absent in the other two synonymous versions of *cat* may have facilitated the excision of the large fragment of the *cat*-GC plasmid, producing a version of the plasmid reduced from the original 3434bp to 1661bp.

In populations selected in chloramphenicol, we identified a series of short indels located upstream the *cat* start codon, which are likely to have a direct impact on the *cat* gene translation. The synonymous versions of the *cat* gene had been inserted in the multiple cloning site of the classical cloning vector pUC57. The *cat* gene was cloned within the coding sequence of the *lacZ* gene and its transcription was totally controlled by the upstream *lac* promoter. In the immediate 5' of the *cat* gene, we had engineered a spacer containing a stop codon in frame with the start codon of *lacZ*, such that all translation started at the start codon of *lacZ* was stopped at this point and all CAT protein translated would exclusively contain the amino acids encoded by the inserted *cat* gene (figure 6). After experimental evolution, all mutations identified in this region are located between the *lacZ* start codon and the spacer stop codon and trigger a +2 reading frame shift (del 1bp, ins 2bp, ins 5 bp) (figure 6). This frame shift removes the in-frame stop codon of the spacer and sets the *cat* gene in frame with the start codon of the *lacZ* gene. The net result in all cases is that all translation started at the *lacZ* start codon is continued through the *cat* start codon (figure 6). Our interpretation is that expression of the *cat* gene will thus benefit from the ribosome recruiting elements and translation enhancers naturally located in the 5'-untranslated region of the transcripts generated under the control of the *lac* promoter. Additionally, it has been shown that, within operons, genes located further from the end of the mRNA show a higher translation level because of the coupling between transcription and translation (Lim *et al.* 2011). A similar phenomenon might also result in a higher translation rate of the *cat* gene from the *LacZ* ATG than from its own ATG. Additionally, proteomic data confirmed that, at generation 1000, population evolved in chloramphenicol (and thus carrying the indels on the plasmids) had a significant increase in CAT protein production whereas those evolved in ampicillin did not present this pattern (table 2). These results are also consistent with the increase in chloramphenicol resistance in these populations (figure 2). The presence of the additional 26 amino acids in the N-terminus of the CAT protein seems to not have any effect on the antibiotic catalytic activity, as this protein is often used in biotechnology context as a reporter gene fused to the C-terminus of a target gene. The increase in molecular weight of the expressed CAT after the addition of the N-terminus stretch is obvious in the apparent protein size in electrophoresis (see figure 5 in Amoros-Moya *et al.* 2010). At first sight, the

increase in *cat* translation triggered by the different frame shift mutations should be the same, simply because the mechanism is the same. However, we observe that some mutations occurred only in one population and that there is some specificity for the association between one synonymous *cat* gene version and the frame shift mutation spreading in the population. For instance, the *2411 ins CC* mutation occurs almost exclusively in ATCam1, while the *2422 ins ATCCA* occurs in populations carrying the *cat*-Opt and the *cat*-GC genes (table 1 and figure 4). Because translation is especially sensitive to the strength of the RNA secondary structures around the ATG start codon, we propose that there may be non-favoured combinations between the repertoire of mutations harmonising the *lacZ* and the *cat* reading frames and the very divergent coding sequences with ample variation in GC content in the three versions of the *cat* genes.

Additionally to the genetic changes discussed above, both the increase in chloramphenicol resistance and the increase in resistance protein production could also be affected by an increase in plasmid copy number. This hypothesis could be discarded thanks to preliminary data of whole genome sequencing on these same populations. To perform this sequencing, total DNA extraction was performed at the population level and ratios of plasmid and chromosome coverage were used as an index of plasmid copy number (see supplemental material). In all cases, independently of the selection regime, the plasmid copy number seems to have been reduced over the course of evolution.

With our experimental protocol using synonymous versions of the *cat* gene, we initially expected to identify synonymous mutations that would shift codon usage preferences of this essential gene closer to those in the *E. coli* genome. Against our expectations, no mutation was identified that would improve the codon usage match: all except one mutations found in the coding sequence are 1bp indels. They are thus changing the reading frame and should be strongly deleterious. The spread and fixation of *2457 ins A* in OptCam1 is thus very difficult to interpret. One might argue that this may be a sequencing error because the insertion occurs within an A₉ string. However, the quality of the read is not inferior to that in other homopolymers, and there is no reason why this error would occur for all reads and only in this population, although the A₉ string was present in all populations carrying *cat*-AT or *cat*-Opt. We tend thus to interpret that this *2457 ins A* is a genuine mutation, although we cannot provide any convincing hypothesis to explain how it can reach fixation.

Sequencing of plasmids within clones in parallel to populations revealed that polymorphisms do occur at the cell level, with clones harbouring a mix of two or three variants of the plasmid. This is true both for the reduced plasmid cohabiting with the ancestral plasmid in GCamp1 and for plasmids carrying frame shift mutations upstream of the *cat* gene. In the case of the DNA samples extracted from nearly-clonal cultures, one can wonder whether this genetic diversity reflects the one present within the cell that started the colony or whether it was generated during the cell division leading from the colony-starter cell to the cell population from which DNA was extracted and sequenced. If the second was true, polymorphisms would be expected to appear randomly in all nearly-clonal samples. Here it is striking that the two cases where abundant polymorphism is detected (namely OptCam1 g358 and GCamp1 g178) at the clonal level correspond to cases where polymorphisms were also detected at the population level. On the contrary when population

level sequencing showed the presence of a single plasmid form, the cell sequencing level rendered the same result. Altogether, this strongly suggests that our clone sequencing approach reflects, at least qualitatively, the situation in individual cells of the population.

The presence of plasmids as independent replicative elements inside the cell paves the way to multilevel selection processes, because a mutation will not necessarily have fitness effects in the same direction at the plasmid level and at the cell level. For frame shift mutations, the advantage in terms of within-cell competition is not clear and the mutation is likely to be close to neutral for plasmid fitness. However, the advantage for the between-cells competition is clear, as it improves chloramphenicol resistance and results in large differences in cellular fitness. For the plasmid reduced to 50% of the original length, there is likely a fitness advantage at the within-cell level in terms of faster replication, as well as a fitness advantage at the between-cell level, because plasmid streamlining reduces the cost of plasmid replication and more importantly the cost of expression of a gene with codon usage mismatch.

Evolution in asexual microorganism has long been thought to proceed by successive selective sweeps and clonal replacement. In the last twenty years, experimental data have challenged this simple model and shown that natural populations often comprise several clones carrying individual beneficial mutations competing during adaptation. This situation, named “clonal interference” and modelled by Gerrish and Lenski (1998), was identified for example in RNA viruses (Miralles 1999) and *Saccharomyces cerevisiae* (Kao & Sherlock 2008). The description of the adaptive dynamics and the identification of clones in competition has become easier with modern sequencing technologies, applied for example to DNA bacteriophages adapting to a new host (Miller *et al.* 2011), or to *E. coli* adapting to the gut environment (Barroso-Batista *et al.* 2014) or to a glucose-limited environment (Maddamsetti *et al.* 2015). In our experimental setup, from the haplotype analysis conducted around the *cat* gene start codon, it can be concluded that it is very rare that two mutations are found on the same read and that haplotypes, in their vast majority, carry only one mutation (figure 4). This means that each sample, either population or individual clone, in which more than one mutation has been identified in this plasmid region is a showcase of “clonal interference”. At the level of between-cells competition, our results fit the picture of widespread clonal interference during adaptation of asexual microorganisms. It is important to note that clonal interference between plasmid versions cannot explain alone “zig-zag dynamics” as seen in GCCam1 or OptCam1. These dynamics are likely to be due to epistatic interactions with mutations occurring on the chromosome.

Our results are further pioneer in showing within-cell clonal interference between different versions of competing plasmids. To our knowledge, there is only one study on the role of population-level clonal interference during plasmid evolution (Hughes *et al.* 2012). It was conducted during the adaptation of a broad-host-range plasmid (pMS0506, 13.1 kb) to a new host (*Shewanella oneidensis*). After 1000 generations of selection for plasmid maintenance, plasmid stability was strongly increased and the associated mutations were short in-frame indels in the 5' region of the plasmidic *trpA1* gene, encoding the replication protein. Genotype dynamics suggested the widespread occurrence of clonal interference between-

cells, as in our data, but the sequencing strategy used in that work did not allow to check for clonal interference at the within-cell level.

The molecular basis of fitness increase provided by the mutations recovered upstream of the *cat* gene fit nicely with clonal interference rather than clone succession as the main mode of evolution. Mutations reconciling the reading frames of the *lacZ* and the *cat* genes must occur in a 63bp window and mandatorily induce a +2 frameshift without negatively affecting the secondary structure of the mRNA (Amoros-Moya *et al.* 2010). We have identified in total five different haplotypes carrying such mutations, some of them appearing in parallel in association with different *cat* versions. The combination in the same molecule of two of such +2 frame shift mutations would generate a +1 frame shift between the two genes. This would disengage translation of the CAT protein from the efficient translation start of the LacZ protein, restoring the original expression of the *cat* gene. These mutations are thus likely to present a strong reciprocal sign epistasis (Poelwijk *et al.* 2007) between them, which mechanistically explains why clonal interference is the evolution mode found in this experiment.

Our original, naïve expectation while designing this experimental setup was that synonymous changes would accumulate in the coding region of the *cat* genes, rendering their codon usage closer to that in the host. The stubborn absence of synonymous changes in the coding region, even at the 10,000x coverage achieved through deep sequencing, clearly shows that such codon usage amelioration process did not occur in this genetic configuration and at this time scale. Clonal interference could also provide an explanation for this. Indeed, the cost of codon usage mismatch is important, as evidenced by the extremely low initial chloramphenicol resistance of the bacteria transformed with the *cat-AT* gene. However, the increase in fitness associated with the optimisation of a single codon is most likely very small, especially when compared with mutations affecting translation initiation. For this reason, synonymous mutations improving codon usage match are unlikely to reach high frequency, as they will probably be outcompeted through clonal interference by mutations with larger fitness effect. A potential additional hindrance to the rise in frequency of synonymous mutations is multi-level selection occurring because the target gene is carried on a high copy number plasmid. Any synonymous mutation appearing in one copy of the plasmid would not confer any advantage for the replication of this specific plasmid molecule, and therefore would not increase the plasmid fitness with regards to within-cell competition. At the between-cells competition level, the fitness effect of one of the plasmid copies carrying a beneficial synonymous mutation is very likely to be negligible compared to other evolutionary forces, which in our experimental evolution setup are essentially drift and mutation loss due to transfer bottlenecks. Synonymous mutations on plasmid-encoded genes, especially those in high-copy number plasmids, should thus have an extremely low probability to spread in the population.

The results here presented highlight the need of further data collection and model development for multi-level selection processes to understand the within-cell and between-cell dynamics of horizontal gene transfer mediated by plasmids. Multi-level selection models have been conceived mainly for organelles (Rand 2001) and the only model for plasmid evolution (Paulsson 2002) focuses on plasmid mutations that affect its replication

rate, but does not address mutations with other effects. An important consequence of the multi-level selection argument is that the mode of gene evolution upon horizontal gene transfer is likely to strongly depend on the genomic compartment of the introduced gene (plasmid or chromosome) and in the case of plasmidic location, on the plasmid copy number. Given the impact on global health of antibiotic resistance spread mediated by horizontal gene transfer, such considerations may prove instrumental to understand gene flow through the heavily connected network of mobile elements and bacterial hosts.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors are indebted to Dolores Amorós-Moya for excellent work during the experimental evolution protocol. SB is funded by the European Research Council Grant HGTCODONUSE, contract number 682819. IGB is funded by the European Research Council Grant CODOVIREVOL, contract number 647916. D.P.-P. is funded by a VRID-UdeC grant.

Literature cited

- Agashe D, Martinez-Gomez NC, Drummond DA, Marx CJ. Good Codons, Bad Transcript: Large Reductions in Gene Expression and Fitness Arising from Synonymous Mutations in a Key Enzyme. *Molecular Biology and Evolution*. 2013; 30:549–560. [PubMed: 23223712]
- Amoros-Moya D, Bedhomme S, Hermann M, Bravo IG. Evolution in Regulatory Regions Rapidly Compensates the Cost of Nonoptimal Codon Usage. *Molecular Biology and Evolution*. 2010; 27:2141–2151. [PubMed: 20403964]
- Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010 Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Baltrus DA. Exploring the costs of horizontal gene transfer. *Trends in Ecology & Evolution*. 2013; 28:489–495. [PubMed: 23706556]
- Barroso-Batista J, Sousa A, Lourenço M, et al. Coop G. The First Steps of Adaptation of *Escherichia coli* to the Gut Are Dominated by Soft Sweeps. *PLoS Genetics*. 2014; 10:e1004182. [PubMed: 24603313]
- Bergstrom CT, Lipsitch M, Levin BR. Natural selection, infectious transfer and the existence conditions for bacterial plasmids. *Genetics*. 2000; 155:1505–1519. [PubMed: 10924453]
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014; 30:2114–2120. [PubMed: 24695404]
- Bouma JE, Lenski RE. Evolution of a bacteria/plasmid association. *Nature*. 1988; 335:351–352. [PubMed: 3047585]
- Bragg JG, Wagner A. Protein material costs: single atoms can make an evolutionary difference. *Trends in Genetics*. 2009; 25:5–8. [PubMed: 19010565]
- Burgess-Brown NA, Sharma S, Sobott F, et al. Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study. *Protein Expression and Purification*. 2008; 59:94–102. [PubMed: 18289875]
- Dahlberg C, Chao L. Amelioration of the cost of conjugative plasmid carriage in *Escherichia coli* K12. *Genetics*. 2003; 165:1641–1649. [PubMed: 14704155]
- Dionisio F, Conceição IC, Marques ACR, Fernandes L, Gordo I. The evolution of a conjugative plasmid and its ability to increase bacterial fitness. *Biology Letters*. 2005; 1:250–252. [PubMed: 17148179]
- Drummond DA, Wilke CO. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell*. 2008; 134:341–352. [PubMed: 18662548]

- Eberhard WG. Evolution in bacterial plasmids and levels of selection. *The Quarterly Review of Biology*. 1990; 65:3–22. [PubMed: 2186429]
- Fernández Lanza V, Tedim A, Martínez J, Baquero F, Coque T. The plasmidome of firmicutes: impact on the emergence and the spread of resistance to antimicrobials. *Microbiol Spectrum*. 2015; 3 PLAS–0039–2014.
- Gerrish PJ, Lenski RE. The fate of competing beneficial mutations in an asexual population. *Genetica*. 1998; 102–103:127–144.
- Harrison E, Brockhurst MA. Plasmid-mediated horizontal gene transfer is a coevolutionary process. *Trends in Microbiology*. 2012; 20:262–267. [PubMed: 22564249]
- Harrison E, Guymer D, Spiers AJ, Paterson S, Brockhurst MA. Parallel Compensatory Evolution Stabilizes Plasmids across the Parasitism-Mutualism Continuum. *Current Biology*. 2015; 25:2034–2039. [PubMed: 26190075]
- Harrison E, Koufopanou V, Burt A, MacLean RC. The cost of copy number in a selfish genetic element: the 2- μ m plasmid of *Saccharomyces cerevisiae*. *Journal of Evolutionary Biology*. 2012; 25:2348–2356. [PubMed: 22994599]
- Hughes JM, Lohman BK, Deckert GE, et al. The Role of Clonal Interference in the Evolutionary Dynamics of Plasmid-Host Adaptation. *mBio*. 2012; 3 e00077–12–e00077–12.
- Kao KC, Sherlock G. Molecular characterization of clonal interference during adaptive evolution in asexual populations of *Saccharomyces cerevisiae*. *Nature Genetics*. 2008; 40:1499–1504. [PubMed: 19029899]
- Komar AA, Lesnik T, Reiss C. Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. *FEBS letters*. 1999; 462:387–391. [PubMed: 10622731]
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25:1754–1760. [PubMed: 19451168]
- Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25:2078–2079.
- Lim HN, Lee Y, Hussein R. Fundamental relationship between operon organization and gene expression. *Proceedings of the National Academy of Sciences*. 2011; 108:10626–10631.
- Lindsay JA. *Staphylococcus aureus* genomics and the impact of horizontal gene transfer. *International Journal of Medical Microbiology*. 2014; 304:103–109. [PubMed: 24439196]
- Lynch M, Marinov GK. The bioenergetic costs of a gene. *Proceedings of the National Academy of Sciences*. 2015 201514974.
- Maddamsetti R, Lenski RE, Barrick JE. Adaptation, Clonal Interference, and Frequency-Dependent Interactions in a Long-Term Evolution Experiment with *Escherichia coli*. *Genetics*. 2015; 200:619–631. [PubMed: 25911659]
- Medrano-Soto A. Successful Lateral Transfer Requires Codon Usage Compatibility Between Foreign Genes and Recipient Genomes. *Molecular Biology and Evolution*. 2004; 21:1884–1894. [PubMed: 15240837]
- San Millan A, Peña-Miller R, Toll-Riera M, et al. Positive selection and compensatory adaptation interact to stabilize non-transmissible plasmids. *Nature Communications*. 2014; 5:5208.
- San Millan A, Toll-Riera M, Qi Q, MacLean RC. Interactions between horizontally acquired genes create a fitness cost in *Pseudomonas aeruginosa*. *Nature Communications*. 2015; 6:6845.
- Miller CR, Joyce P, Wichman HA. Mutational Effects and Population Dynamics During Viral Adaptation Challenge Current Models. *Genetics*. 2011; 187:185–202. [PubMed: 21041559]
- Miralles R. Clonal Interference and the Evolution of RNA Viruses. *Science*. 1999; 285:1745–1747. [PubMed: 10481012]
- Modi RI, Adams J. Coevolution in Bacterial-Plasmid Populations. *Evolution*. 1991; 45:656.
- Paulsson J. Multileveled selection on plasmid replication. *Genetics*. 2002; 161:1373–1384. [PubMed: 12238464]
- Peña-Miller R, Rodríguez-González R, MacLean RC, San Millan A. Evaluating the effect of horizontal transmission on the stability of plasmids under different selection regimes. *Mobile Genetic Elements*. 2015; 5:29–33.

- Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. *Nature Reviews Genetics*. 2011; 12:32–42.
- Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature*. 2007; 445:383–386. [PubMed: 17251971]
- Porse A, Schønning K, Munck C, Sommer MOA. Survival and Evolution of a Large Multidrug Resistance Plasmid in New Clinical Bacterial Hosts. *Molecular Biology and Evolution*. 2016; 33:2860–2873. [PubMed: 27501945]
- Puigbò P, Romeu A, Garcia-Vallvé S. HEG-DB: a database of predicted highly expressed genes in prokaryotic complete genomes under translational selection. *Nucleic Acids Research*. 2008; 36:D524–527. [PubMed: 17933767]
- Rand DM. The Units of Selection on Mitochondrial DNA. *Annual Review of Ecology and Systematics*. 2001; 32:415–448.
- Reams AB, Roth JR. Mechanisms of Gene Duplication and Amplification. *Cold Spring Harbor Perspectives in Biology*. 2015; 7:a016592. [PubMed: 25646380]
- Rocha EPC. Codon usage bias from tRNA's point of view: Redundancy, specialization, and efficient decoding for translation optimization. *Genome Research*. 2004; 14:2279–2286. [PubMed: 15479947]
- Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics (Oxford, England)*. 2011; 27:863–864.
- Sota M, Yano H, Hughes JM, et al. Shifts in the host range of a promiscuous plasmid through parallel evolution of its replication initiation protein. *The ISME Journal*. 2010; 4:1568–1580. [PubMed: 20520653]
- Tuller T, Girshovich Y, Sella Y, et al. Association between translation efficiency and horizontal gene transfer within microbial communities. *Nucleic Acids Research*. 2011; 39:4743–4755. [PubMed: 21343180]
- Wilm A, Aw PPK, Bertrand D, et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Research*. 2012; 40:11189–11201. [PubMed: 23066108]

Figure 1.

Experimental design. The table on the left describes for each synonymous version of the *cat* gene the GC content and the number of modified codons relative to the *cat*-Opt version.

Figure 2.

Phenotypic changes in experimentally evolved populations. a) Chloramphenicol IC_{50} in $\mu\text{g.mL}^{-1}$ for maximum OD (in black) and maximum growth rate (in grey) for ancestral (g0) and populations for 1000 generations (“Amp”, evolved in ampicillin; “Cam”, evolved in chloramphenicol) carrying the three synonymous versions of the *cat* gene (AT, GC, Opt). Error bars represent standard error. b) Normalized maximum optical density reached by the ancestral and the evolved populations at generation 400 and 1000 in presence of $25 \mu\text{g.mL}^{-1}$ of chloramphenicol.

Figure 3.

Maximum optical density reached by the ancestral populations (cell0 - plasmid0, in red), the evolved populations (cell1000-plasmid1000, in blue) and populations carrying the evolved plasmid in an ancestral chromosome background (cell0-plasmid1000, in green), when grown in the absence of antibiotics or in the presence of either chloramphenicol or ampicillin.

Figure 4. Haplotype proportions in each sample. Overlay-plots represent the changes in the population haplotype composition across time. For ATCam1 and OptCam1, the bar-plots below the indicated time points represent the haplotype composition for the five clones sequenced.

Figure 5.

Coverage along the plasmid sequence for GCamp1 population. Values for the population are shown in red and those for the clones in blue. The percentage of reduced plasmid given on the right have been calculated as $(1 - \text{median coverage on the deleted part} / \text{median coverage on the conserved part}) \times 100$. The linear plasmid map is depicted at the bottom, and the conserved part spanning the *rep* and the *bla* genes is indicated by discontinuous lines. Only generations 173 (upper panel) and 358 (lower panel) have been represented. Generations 558

and 1000 present patterns very similar to generation 358, with a minimum proportion of reduced plasmid of 99.91%.

Initial sequence

5' end of the lacZ gene spacer with an in frame stop codon 5' end of the cat gene

ATG ACC.ATG.ATT.ACG.CCA.AGC.TTG.CAT.GCA.GGC.CTC.TGC.AGT.CGA.CGG.GCC.CGG.GAT.CCG.ATC.CAA.TGA.ATT.CAT.GGA.CAC.CTA.CCG.TTA.CAT.CGA.AAA.AAA

Translation from the lacZ start codon
M T M I T P S L H A G L C S R R A R D P I Q *

Translation from the cat start codon
M D T Y R Y I E K K

2422 ins ATCCA
ATG ACC.ATG.ATT.ACG.CCA.AGC.TTG.CAT.GCA.GGC.CTC.TGC.AGT.CGA.CGG.GCC.CGG.GAT.CCG.ATC.CA.ATCC.AAT.GAA.TTC.ATG.GAC.ACC.TAC.CGT.TAC.ATC.GAA.AAA.AAA
M T M I T P S L H A G L C S R R A R D P I Q S N E F M D T Y R Y I E K K

2409 del G
ATG ACC.ATG.ATT.ACG.CCA.AGC.TTG.CAT.GCA.GGC.CTC.TGC.AGT.CGA.CGG.CCC.GGG.ATC.CGA.TCC.AAT.GAA.TTC.ATG.GAC.ACC.TAC.CGT.TAC.ATC.GAA.AAA.AAA
M T M I T P S L H A G L C S R R P G I R S N E F M D T Y R Y I E K K

2411 ins CC
ATG ACC.ATG.ATT.ACG.CCA.AGC.TTG.CAT.GCA.GGC.CTC.TGC.AGT.CGA.CGG.GCC.CCG.GGG.ATC.CGA.TCC.AAT.GAA.TTC.ATG.GAC.ACC.TAC.CGT.TAC.ATC.GAA.AAA.AAA
M T M I T P S L H A G L C S R R A P G I R S N E F M D T Y R Y I E K K

2415 del G
ATG ACC.ATG.ATT.ACG.CCA.AGC.TTG.CAT.GCA.GGC.CTC.TGC.AGT.CGA.CGG.GCC.CCG.ATC.CGA.TCC.AAT.GAA.TTC.ATG.GAC.ACC.TAC.CGT.TAC.ATC.GAA.AAA.AAA
M T M I T P S L H A G L C S R R A R I R S N E F M D T Y R Y I E K K

2399 del T
ATG ACC.ATG.ATT.ACG.CCA.AGC.TTG.CAT.GCA.GGC.CTC.GCA.GTC.GAC.GGG.CCC.GGG.ATC.CGA.TCC.AAT.GAA.TTC.ATG.GAC.ACC.TAC.CGT.TAC.ATC.GAA.AAA.AAA
M T M I T P S L H A G L A V D G P G I R S N E F M D T Y R Y I E K K

Figure 6.

Translation of the mRNA transcribed from the Lac promotor for the initial construct and for the frameshift mutations reaching high frequency in evolving populations. Insertions are indicated by yellow highlight and downwards triangles and deletions are indicated by yellow highlight and upwards triangles.

Table 1

Changes detected in the plasmid and populations in which they appear. Mutations in *italics* are included in the haplotype analysis and their respective frequencies in the different samples are given in figure 4.

Plasmid region	mutation	Populations presenting this mutation
<i>rep</i> gene	1469 C→T	OptCam1 (1 clone – 13.6%)
<i>cat</i> translation regulatory zone	<i>2399 del T</i>	GCCam1
	<i>2409 del G</i>	ATCam2
	<i>2411 ins CC</i>	ATCam1
	<i>2415 del G</i>	OptCam1
	<i>2422 ins ATCCA</i>	OptCam1, OptCam2, GCCam1, GCCam2
<i>cat</i> gene	<i>2454 ins G</i>	GCCam1
	<i>2457 ins A</i>	OptCam1, OptCam2
	<i>2458 del A</i>	OptCam1, ATCam1, ATCam2
	2483 ins T	GCCam1 (g173: 2.4%)
	2487 G→A	OptCam1 (1 clone – 1.9%)
First position after the <i>cat</i> gene stop codon	3148 ins A	GCCam2 (4 time points - 100%)
From the last bp of the <i>rep</i> gene to the intergenic region between <i>lac</i> and <i>bla</i> .	del 1972- 301	GCamp1 (all samples and time points)

Table 2

Relative quantification of CAT and BLA proteins. The three replicates of populations carrying the same synonymous version of the *cat* gene and evolved in the same antibiotic were pooled before protein extraction. Protein levels are expressed relative to the protein quantification in the population carrying the *cat*-AT gene right after transformation (generation 0).

Population	Chloramphenicol acetyltransferase		B-lactamase	
	Fold-change (relative to ATg0)	95 % Confidence interval (pval)	Fold-change (relative to ATg0)	95% Confidence interval (pval)
GC g0	1.6	0.92–2.8 (0.082)	1.3	0.87–1.9 (0.16)
OPT g0	1.3	0.9–1.84(0.18)	1.0	0.56–2.0 (0.84)
ATCam g1000	2.8	1.9–4.2 (1.7e-3)	1.04	Not determined (very low number of peptides identified)
GCCam g1000	12.2	2.7–53.8 (7.5e-3)	0.68	0.34–1.3 (0.22)
OptCam g1000	6.9	2.2–21.7 (7.7e-3)	1.1	0.67–1.7 (0.78)
ATamp g1000	1.4	0.99–2.02 (0.056)	1.6	1.2–2.2 (4.7e-3)
GCamp g1000	1.2	0.054–25.84 (0.89) (very low number of peptides identified)	1.2	0.58–2.6 (0.16)
OptAmp g1000	1.6	1.0–2.6 (0.051)	15.5	6.8–35.7 (1.9e-4)