

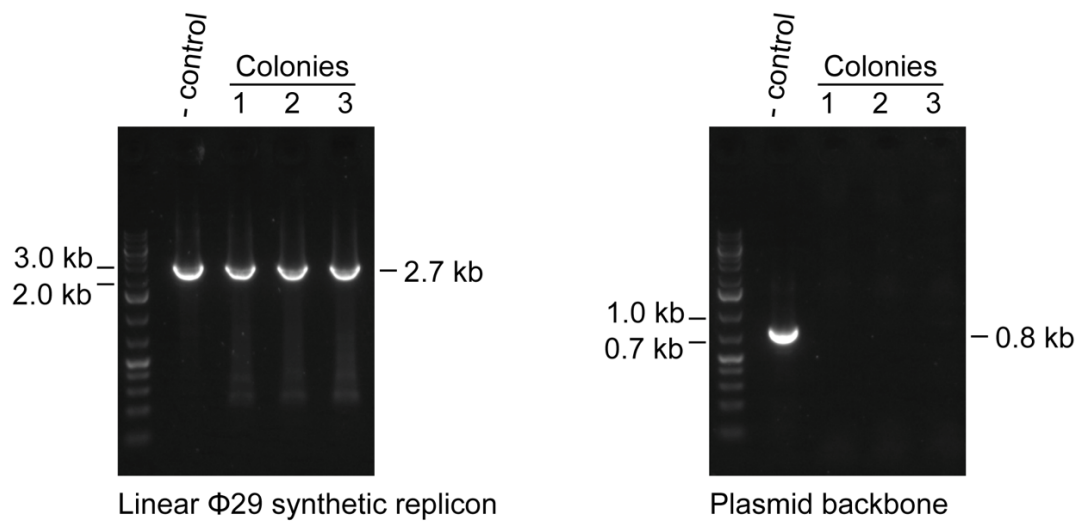
Highly mutagenic continuous evolution in *E. coli* using a Φ 29-based orthogonal replication system

In the format provided by the
authors and unedited

Table of Contents

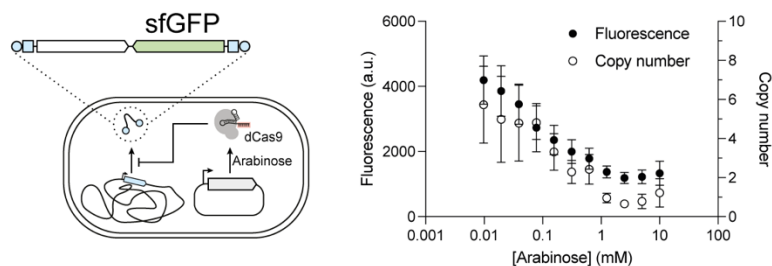
Supplementary Figures	2
Unprocessed gel images	26
Supplementary References	29

Supplementary Figures



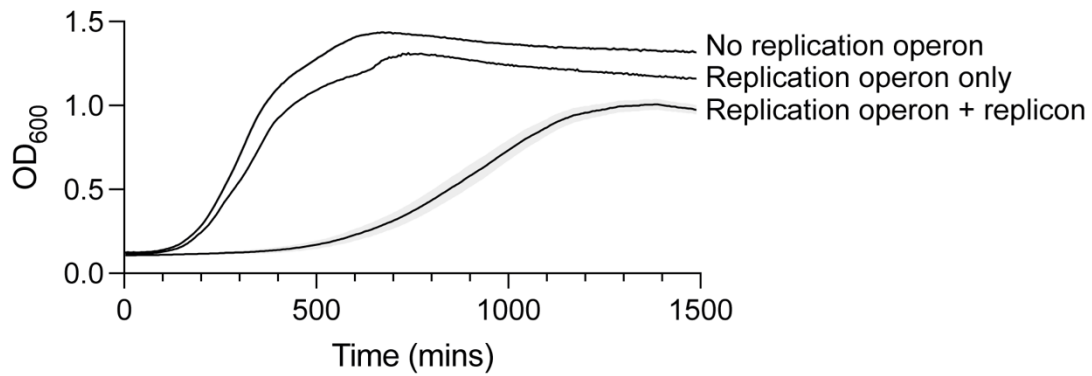
Supplementary Figure 1. Genotyping establishment of the linear Φ 29 synthetic replicon.

Colony PCRs to genotype for the TcR-GFP Φ 29 synthetic replicon (left) and the backbone of the circular plasmid (right) from which the PCR product used for establishing the replicon *in vivo* was amplified, as analyzed by agarose gel electrophoresis. The negative control is a circular plasmid which gives a band corresponding to the plasmid backbone which is absent from the linear Φ 29 synthetic replicons. The amplicon sequences were verified via Nanopore sequencing.



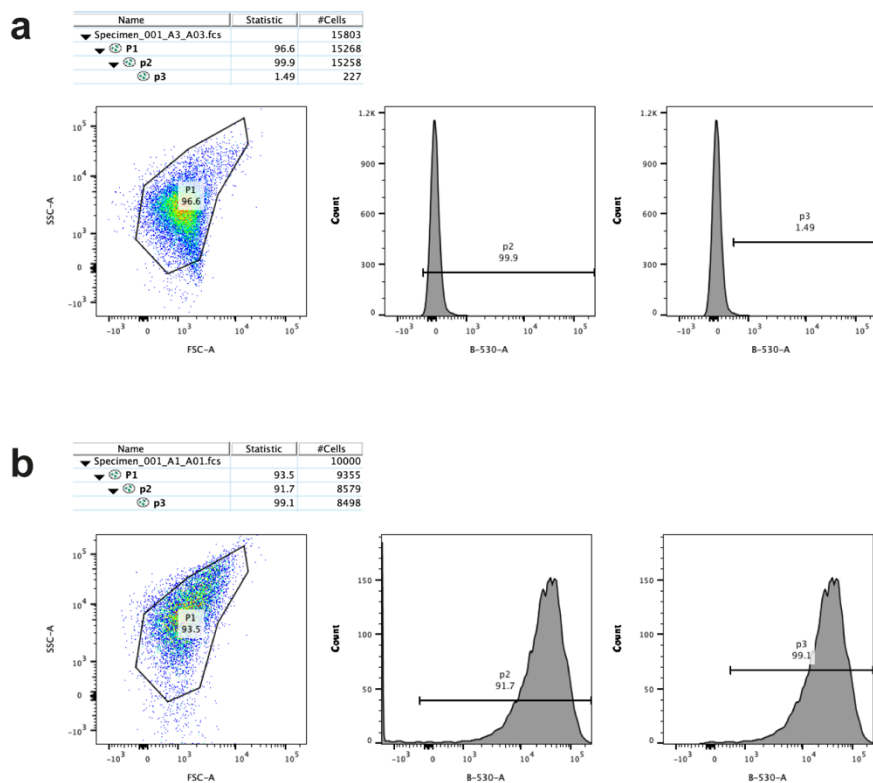
Supplementary Figure 2. Repressing the copy number of the Φ29 synthetic replicon.

The copy number of a TcR-GFP Φ29 synthetic replicon can be reduced by downregulating the expression from a genomically-integrated Φ29 synthetic replication operon via arabinose-inducible expression of dCas9 targeted to the promoter that is driving the replication operon. This change in copy number can be measured by GFP fluorescence and qPCR, and the two measurements are well correlated ($n = 3$ biological replicates, error bars indicate $\pm$ SD). Fluorescence values were normalised to culture density.



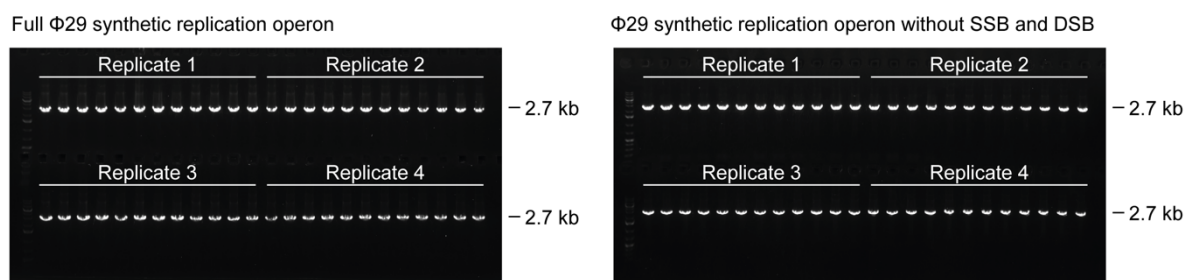
Supplementary Figure 3. Growth of *E. coli* harboring the Φ 29 synthetic replication operon and Φ 29 synthetic replicon.

Cell growth was measured via a plate reader ($n = 16$ biological replicates, the boundaries indicate $\pm$ SEM). Cells not transformed with the Φ 29 synthetic replicon were instead transformed with a colE1 plasmid harboring the same genes (encoding for GFP and tetracycline resistance). The Φ 29 synthetic replication operon was encoded by a single-copy plasmid. The data for the replication operon with the replicon are also shown in Figure 2.



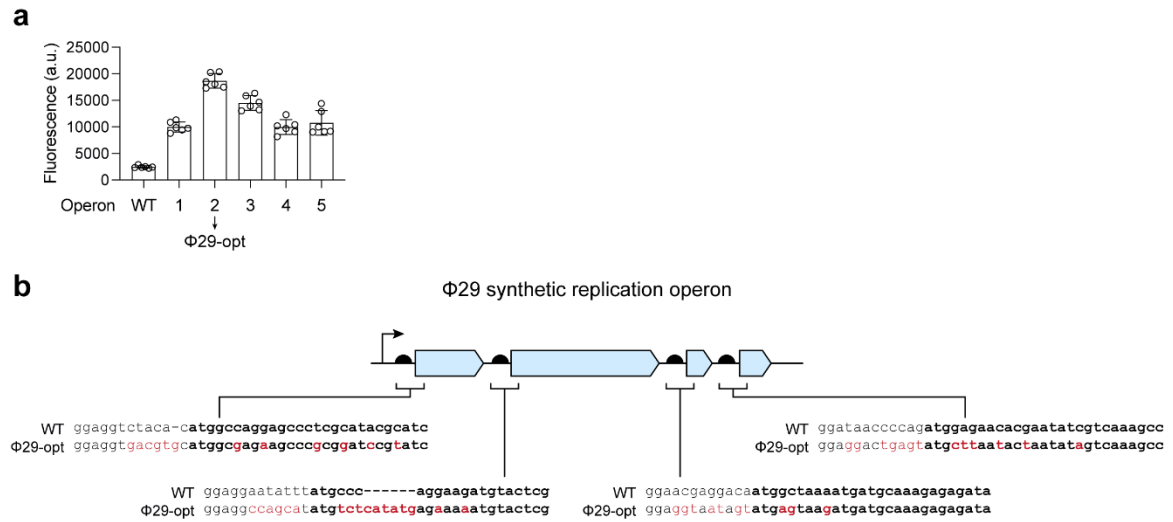
Supplementary Figure 4. Flow cytometry gating strategy.

a, Flow cytometry gating strategy for non-fluorescent negative control cells. **b**, Flow cytometry gating strategy for fluorescent cells harbouring a newly established TcR-GFP $\Phi 29$ synthetic replicon.



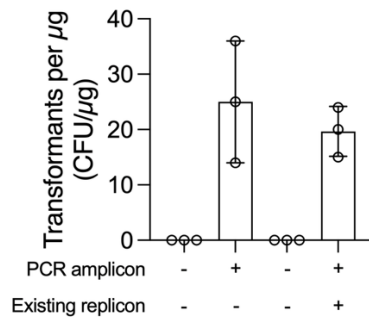
Supplementary Figure 5. Colony PCRs of a TcR-GFP Φ 29 synthetic replicon as maintained by a full-length Φ 29 synthetic replication operon or a variant without SSB and DSB.

We passaged cells harboring the Φ 29 synthetic replication operon, with or without the genes encoding for the SSB and DSB, on a single-copy plasmid and a TcR-GFP Φ 29 synthetic replicon for 50 generations and assessed whether the Φ 29 synthetic replicons remained intact via colony PCR using primers that anneal to the replicon terminal ends.



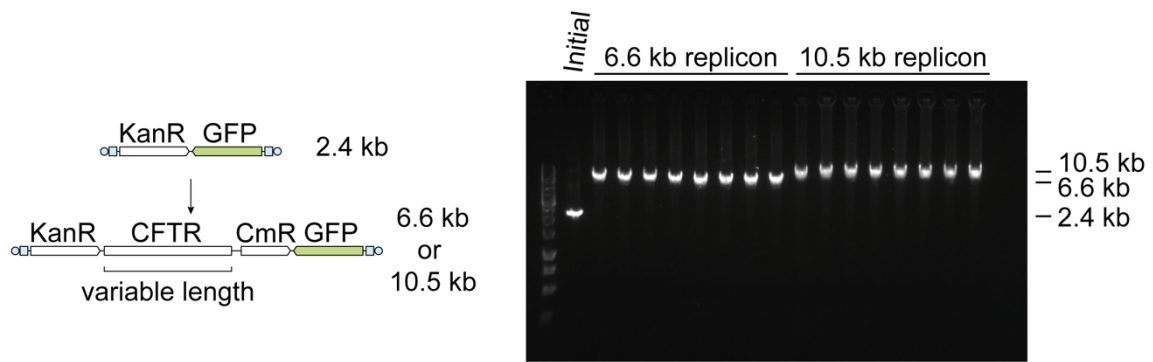
Supplementary Figure 6. Engineering an improved Φ29 synthetic replication operon.

a, Fluorescence of cells harboring a GFP-encoding Φ29 synthetic replicon as maintained by the initial (WT) Φ29 synthetic replication operon or engineered variants. The data for the WT and Φ29-opt operons is also shown in Figure 2 ($n = 6$ biological replicates, error bars indicate $\pm$ SD). Fluorescence values were normalised to culture density. **b**, Comparison of the WT Φ29 synthetic replication operon sequence with that of Φ29-opt. Sequences within genes are shown in bold. Nucleotide changes in Φ29-opt are highlighted in red.



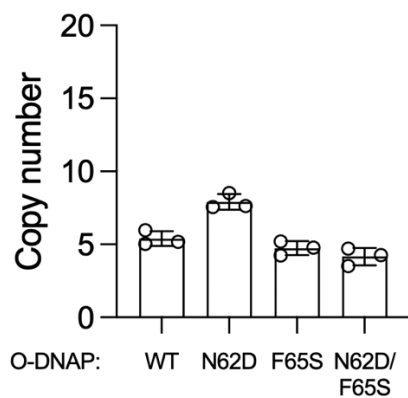
Supplementary Figure 7. Efficiency of establishing new $\Phi 29$ synthetic replicons from PCR amplicons using $\Phi 29$ -opt.

Transformation efficiency for establishing a TcR-GFP $\Phi 29$ synthetic replicon from a PCR amplicon in cells harbouring a genomically-integrated $\Phi 29$ -opt synthetic replication operon and a helper plasmid to additionally express *gam* and the genes encoding for the $\Phi 29$ SSB and DSB. Additionally, we examined these transformations in the context of cells which harboured an existing KanR-GFP $\Phi 29$ synthetic replicon but this did not further improve transformation efficiency ($n = 3$ technical replicates – three electroporations into the same batch of electrocompetent cells, error bars indicate $\pm$ SD).



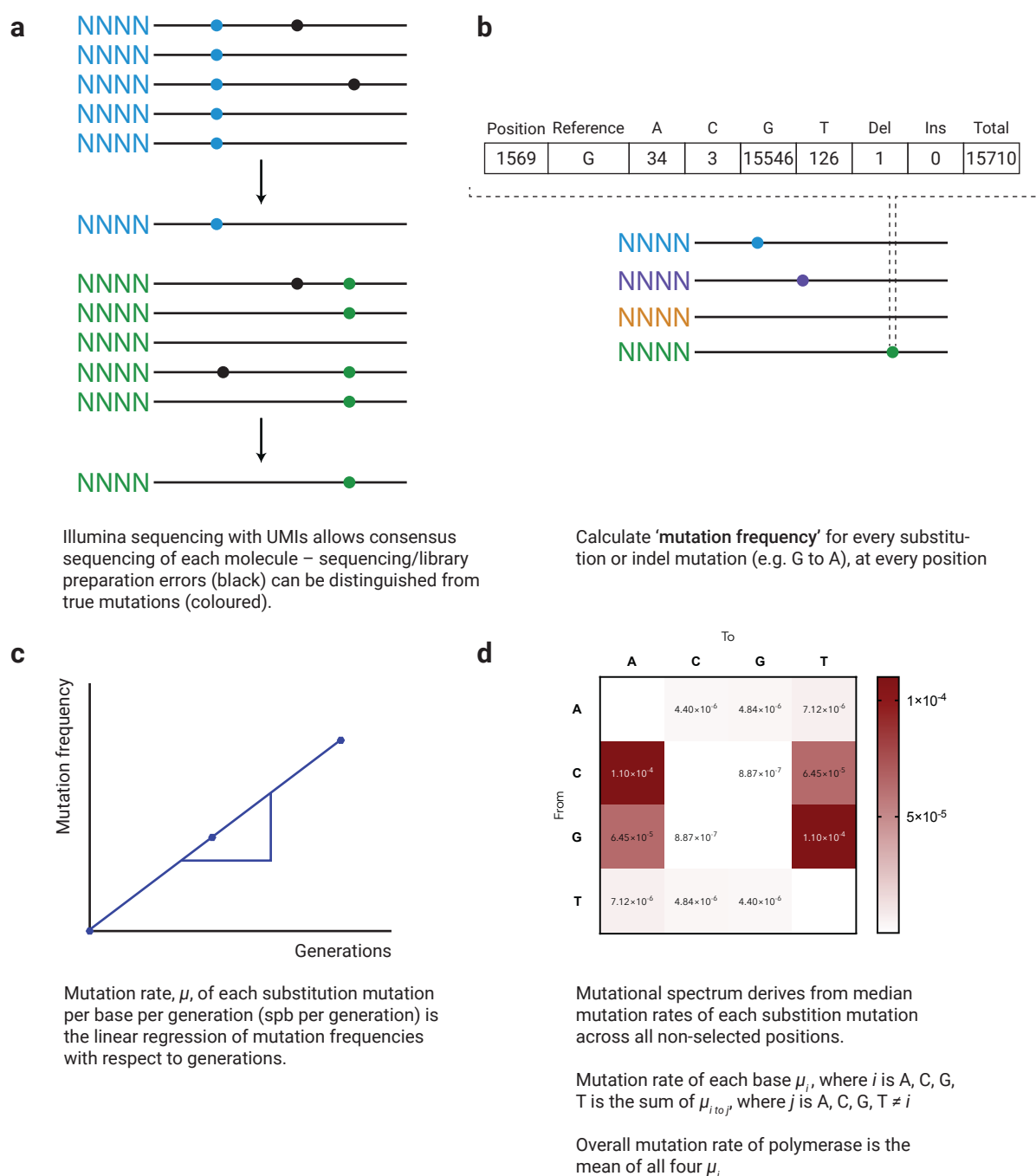
Supplementary Figure 8. Generating long Φ29 synthetic replicons.

We used lambda Red to insert sequences consisting of a variable portion of the human *CFTR* gene and a CmR gene onto a 2.4 kb KanR-GFP replicon, yielding replicons with lengths of 6.6 and 10.5 kb. Shown are PCR amplicons from overnight cultures set from 8 individual colonies per replicon. The amplicon sequences were also verified via Nanopore sequencing.



Supplementary Figure 9. Copy numbers of the KanR-CmR(Q38TAG) replicon as maintained by different O-DNAP variants for the fluctuation test.

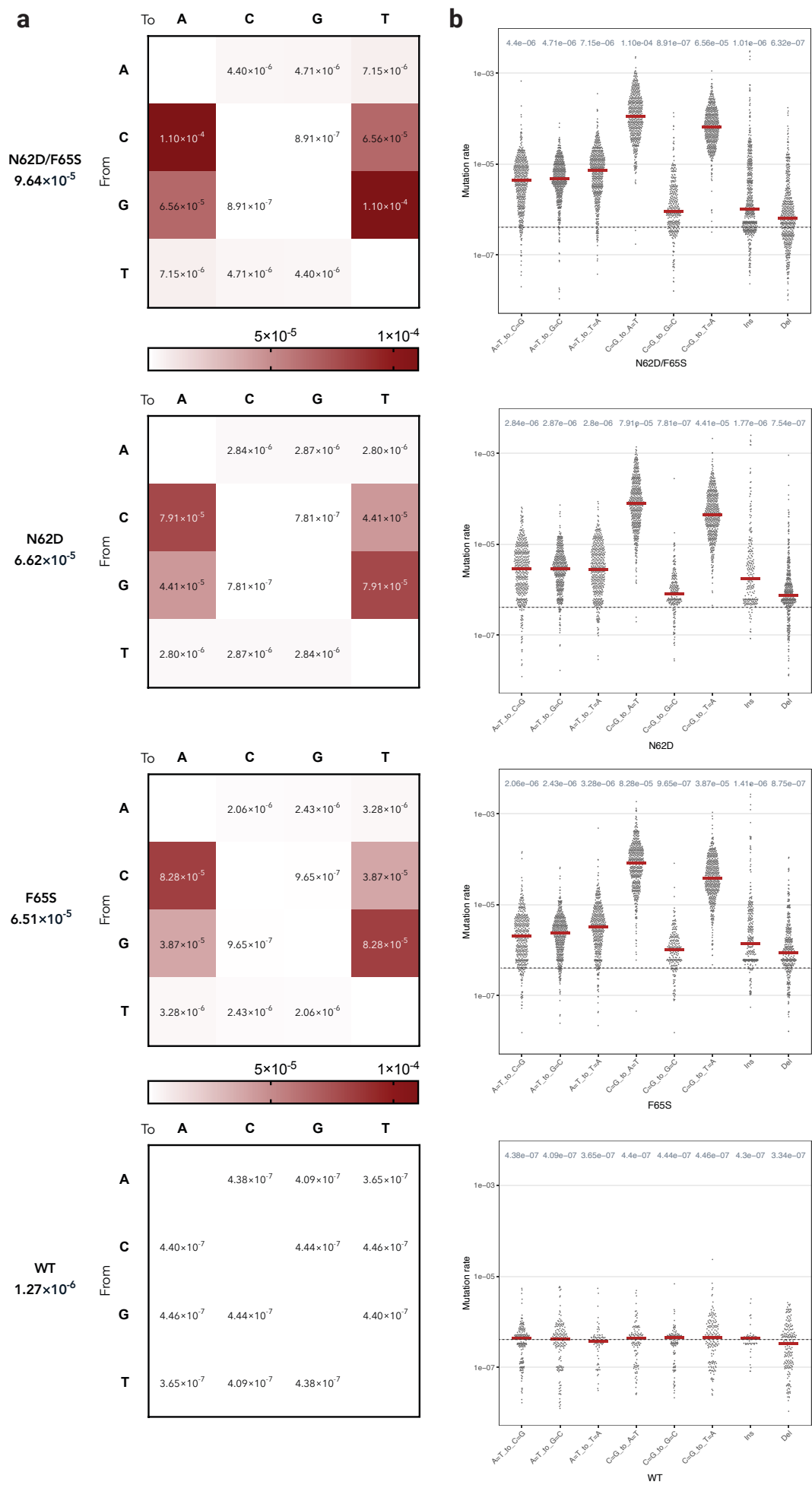
($n = 3$ technical replicates from 12 pooled biological replicates, error bars indicate $\pm$ SD)

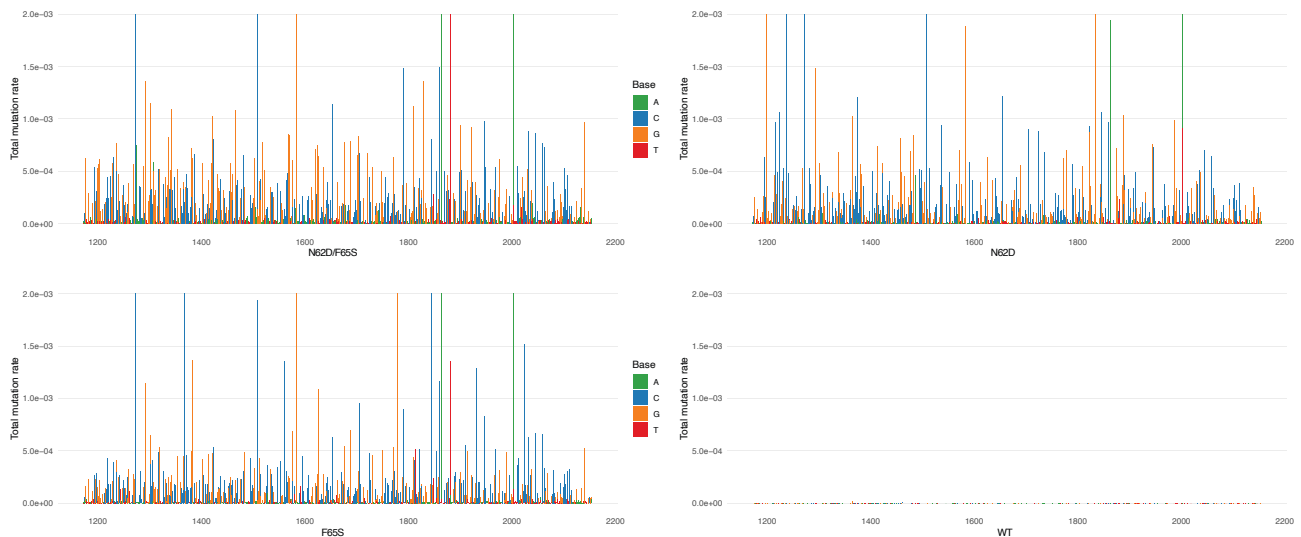


Supplementary Figure 10. Overview of analysis to estimate DNA polymerase mutation rates.

a, UMIs permit consensus sequencing to reduce 'false mutations' that arise from sequencing error¹. Here, coloured spots indicate 'true mutations' on the replicon; only if consensus is reached within the UMI family will it be called. Black spots indicate false mutations that arise from sequencing error. In this analysis, UMIs with >3 detected members were used to call consensus sequences. **b**, Mutation frequencies at each position are called by creating a pileup – base numbers at each position are counted, including insertions (counts both length and frequency) and deletions. Any 'N' basecalls (ambiguous nucleotide by Illumina sequencing) are ignored. Mutation frequencies per base and per indel are calculated as the

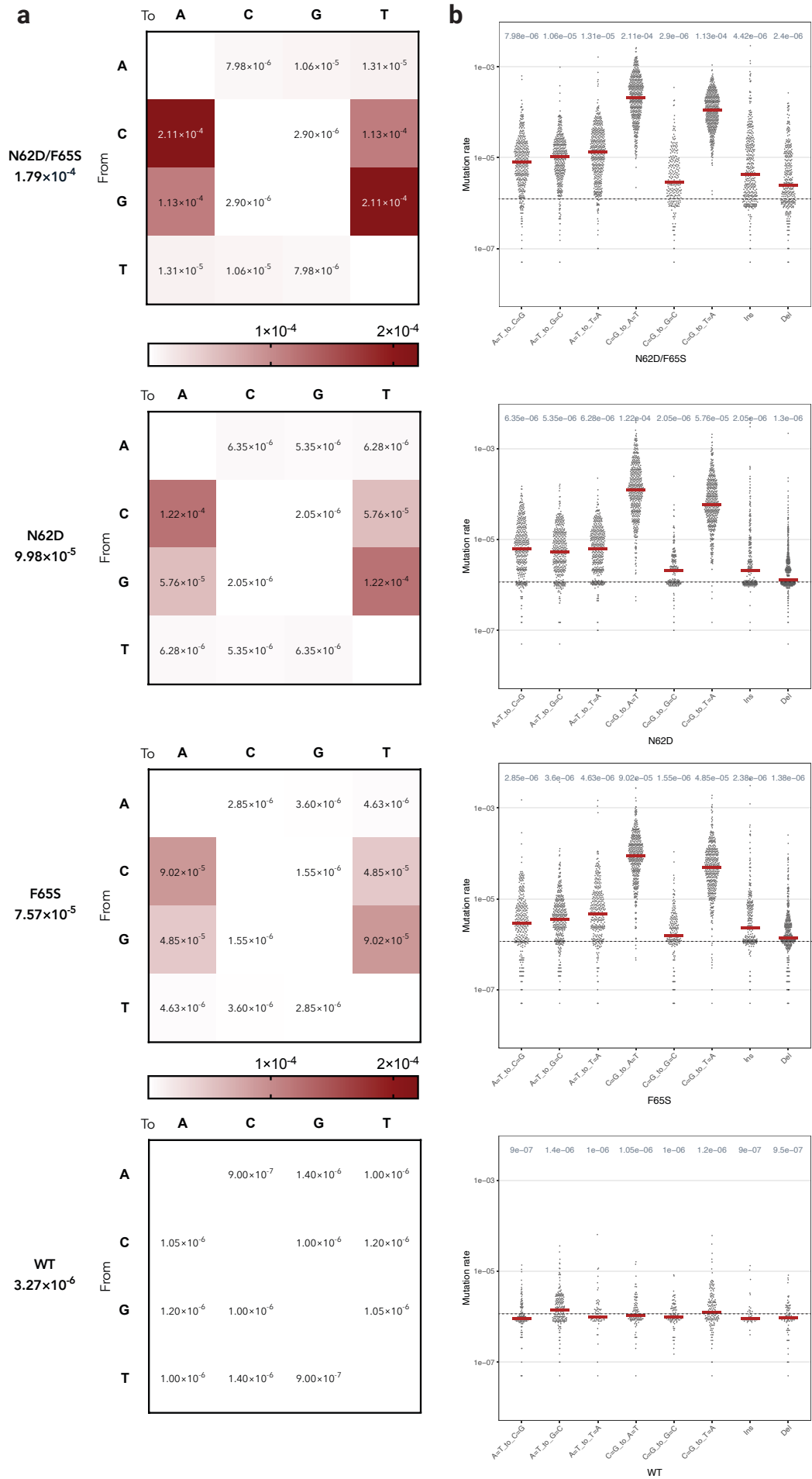
base count divided by total sequencing depth at that position. **c**, For each base substitution mutation e.g. C to T, we perform linear regression to find μ , the substitution mutation rate of each position per base, per generation (s.p.b. per generation). Each position produces 12 possible substitutions, and insertions and deletions; insertions are treated such that every inserted base is counted. Linear regression assumes that the mutation rate remains sufficiently low such that the count of original base in the sequence is assumed to remain constant with respect to acquired mutations. **d**, Mutation spectra are generated by taking the median mutation rate for each transformation across all non-selected positions (in this replicon – not the first and last 100 bp because of TP binding, and not the Kan resistance gene: 100:250 and 1172:2094). Because the polymerase operates in two directions and our sequencing is not stranded, we combined complimentary substitution mutations, e.g. C to T is indistinguishable from G to A, and so these two sets of mutation rates are pooled. The overall mutation rate is calculated as the mean of the 4 substitution mutation rate totals originating from every starting base. Indel rates were at least an order of magnitude lower and so were not included in this calculation (see Supplementary Figures 11 and 12).

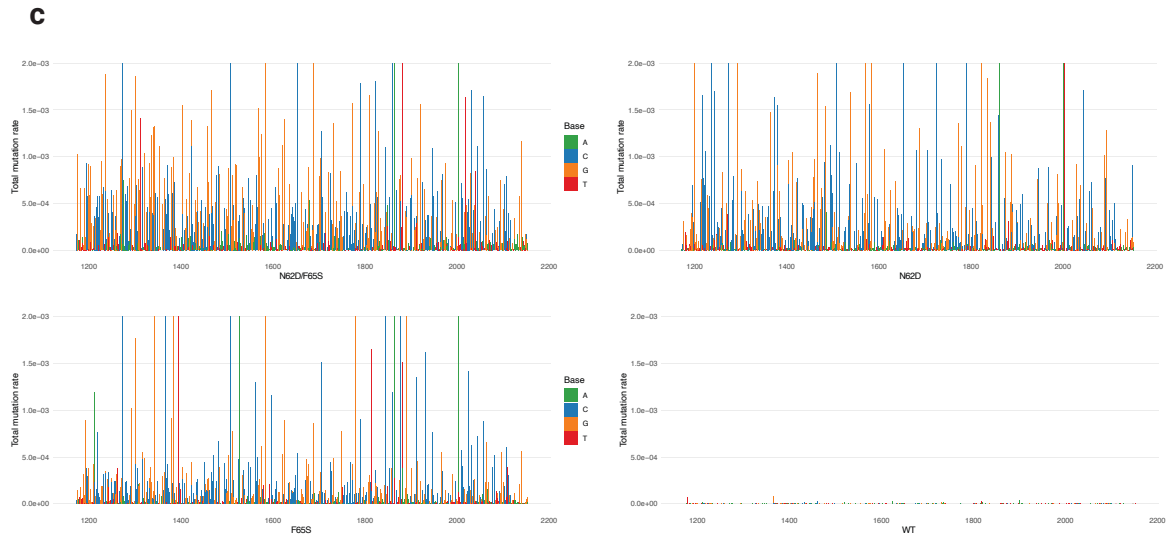


c

Supplementary Figure 11. Sequencing-based measurement of DNAP mutation rates using samples from 0, 20, and 50 generations.

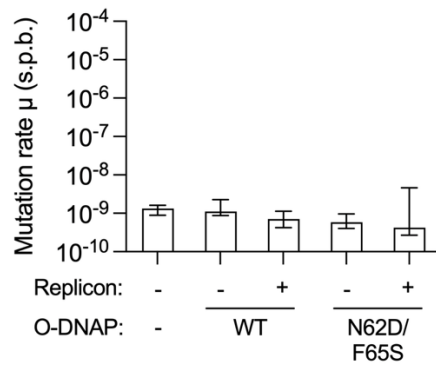
These data originate from a total of 1070 non-selected positions, number of A = 300, C = 244, G = 236, T = 290; negative mutation rates are filtered out. **a**, Mutation spectra for each of N62D/F65S, N62D, F65S and WT; all μ in s.p.b. per generation (same scale). The overall trend matches that of the fluctuation analysis at the CAT stop codon. In all polymerases, G=C base pairs mutate more frequently. The TAG stop codon (CTA in forward direction) has measured mutation rates at the G/C position (1902) are $5.1\text{e-}4$, $1.5\text{e-}4$, $1.9\text{e-}4$, and $5.0\text{e-}7$ for N62D/F65S, N62D, F65S and WT respectively. **b**, Mutation rates for each transformation, including indels (per base per generation), showing the variance for all non-selected positions in the linear replicon. Each point represents a single mutation at each position, and medians are quoted and indicated by the red line. Given the WT DNAP has a mutation rate of $\mu \approx 10^{-9}$ s.p.b. per generation, the measured mutation rates of the WT DNAP using this method provide an approximate limit of measurement (here $\sim 4\text{e-}7$ indicated by the dotted line). A theoretical limit of measurement can be estimated by $1/(\text{coverage} \times \text{generations})$; mean theoretical limit for 50 generations = $2\text{e-}6$. **c**, Summed mutation rates (all possible substitutions + indels) for each position for the unselected Cat gene and surrounding sequence (1172–2090), colours indicate original reference base. Mutation rates across this region do not appear to exhibit directional bias.





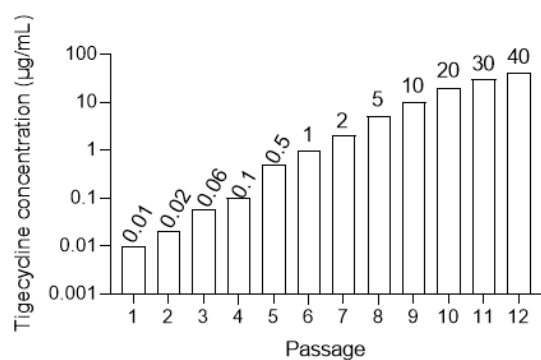
Supplementary Figure 12. Sequencing-based measurement of DNAP mutation rates using samples from 0 and 20 generations.

These data originate from a total of 1070 positions, number of A = 300, C = 244, G = 236, T = 290; negative mutation rates are filtered out. We hypothesise that performing the analysis over earlier time points (i.e. 20 generations only) may give a more accurate value for mutation rates, given that highly mutating sequences as defined by polymerase biases may be consumed over increased numbers of generations. **a**, Mutation spectra for each of N62D/F65S, N62D, F65S and WT, all μ in s.p.b. per generation (same scale). The overall trend of mutation rates in these polymerases does not change depending on number of generations included in the analysis. The TAG stop codon (CTA in forward direction) now has measured mutation rates of $5.8\text{e-}4$, $2.9\text{e-}4$, $1.1\text{e-}5$, and $-5.0\text{e-}6$ for N62D/F65S, N62D, F65S and WT respectively. **b**, Mutation rates for each transformation, including indels (per base per generation), showing the variance for all non-selected positions in the linear replicon. Each point represents a single mutation at each position, and medians are quoted and indicated by the red line. Given the WT DNAP has a mutation rate of $\mu \approx 10^{-9}$ s.p.b. per generation, the measured mutation rates of the WT DNAP using this method provide an approximate limit of measurement (here $\sim 2\text{e-}6$ indicated by the dotted line). A theoretical limit of measurement can be estimated by $1/(\text{coverage} \times \text{generations})$; mean theoretical limit for 50 generations = $5\text{e-}6$. **c**, Summed mutation rates (all possible substitutions + indels) for each position for the unselected Cat gene and surrounding sequence (1172–2154), colours indicate original reference base. Mutation rates across this region do not appear to exhibit directional bias.

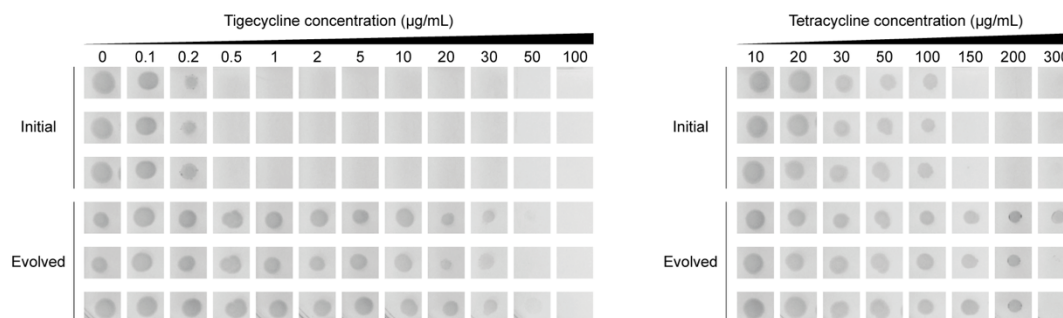


Supplementary Figure 13. Error-prone mutants of the Φ 29 DNAP do not increase the genomic mutation rate in the presence of a Φ 29 replicon.

We used fluctuation analysis to calculate the mutation rate from the frequency of a TAG stop codon reversion in a genomically-integrated chloramphenicol resistance gene (Q38TAG) ($n = 12$ biological replicates, error bars indicate the median $\pm$ upper or lower 95% bounds).

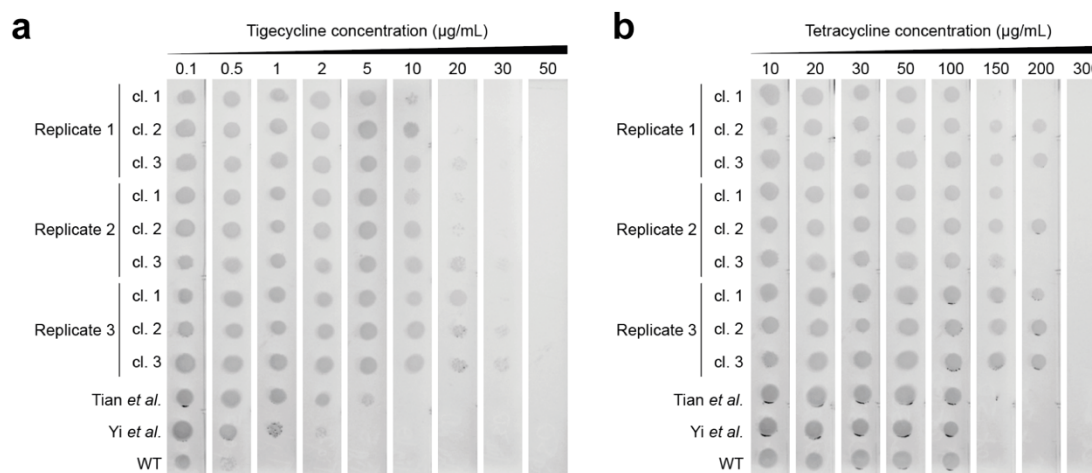


Supplementary Figure 14. Concentrations of tigecycline used at each passage of the continuous evolution.



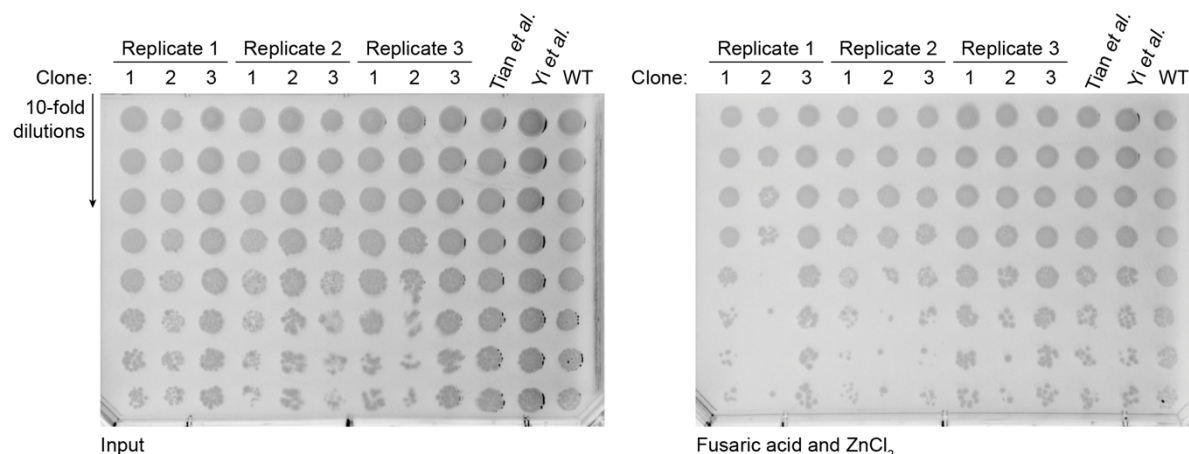
Supplementary Figure 15. Tigecycline and tetracycline resistance of pools of cells harboring evolved replicon-encoded *tetA* variants.

After passaging cells containing the N62D/F65S DNAP mutant and a *tetA*-encoding $\Phi 29$ synthetic replicon in increasingly higher concentrations of tigecycline, we obtained pools of mutants able to sustain growth at higher tigecycline and tetracycline concentrations. The first pool before and after evolution is also shown in Figure 4.



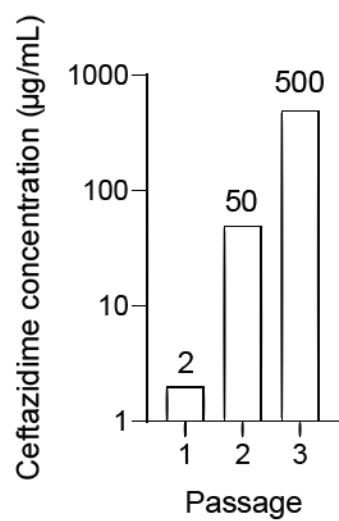
Supplementary Figure 16. Tigecycline and tetracycline resistance of cloned *tetA* variants.

Mutants obtained from passaging a *tetA-GFP* Φ 29 synthetic replicon maintained by the error-prone N62D/F65S DNAP mutant were cloned into *colE1* circular plasmids and transformed into DH10B cells to assess the level of tigecycline (**a**) and tetracycline (**b**) resistance conferred. Two previously reported mutants were tested as a comparison. The mutations found in each tested variant are listed in Supplementary Data 1 – the spots are arranged from top to bottom in increasing numerical order of the variants listed in the table.

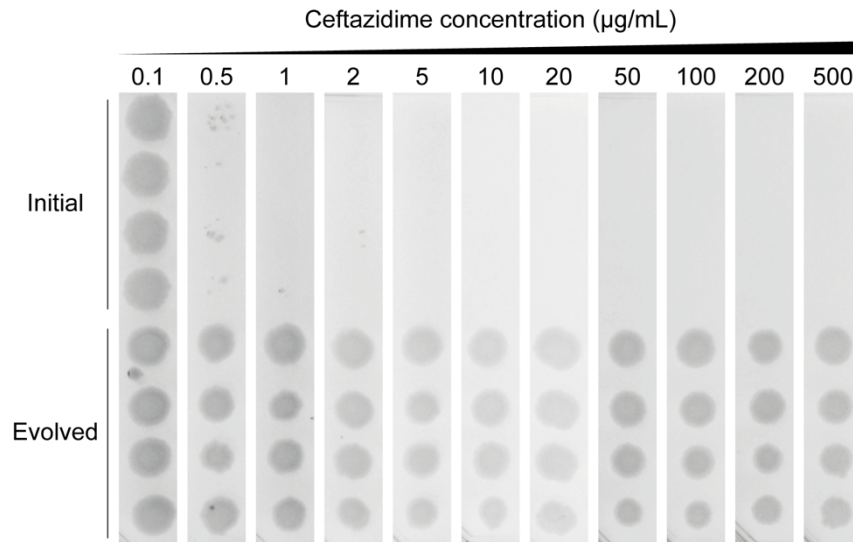


Supplementary Figure 17. Evolved *tetA* variants display a range of sensitivities to a combination of fusaric acid and ZnCl₂.

A combination of fusaric acid (48 mg/L) and ZnCl₂ (220 mg/L) can be used to counter-select against cells expressing TetA. WT TetA or mutants obtained from the continuous evolution for tigecycline resistance and cloned into colE1 circular plasmids were tested for their sensitivity to these compounds. Both the input and the counterselection plates were also supplemented with 10 mg/L tetracycline. Variants displaying enhanced sensitivity might be of use as improved dual positive-negative selection markers.

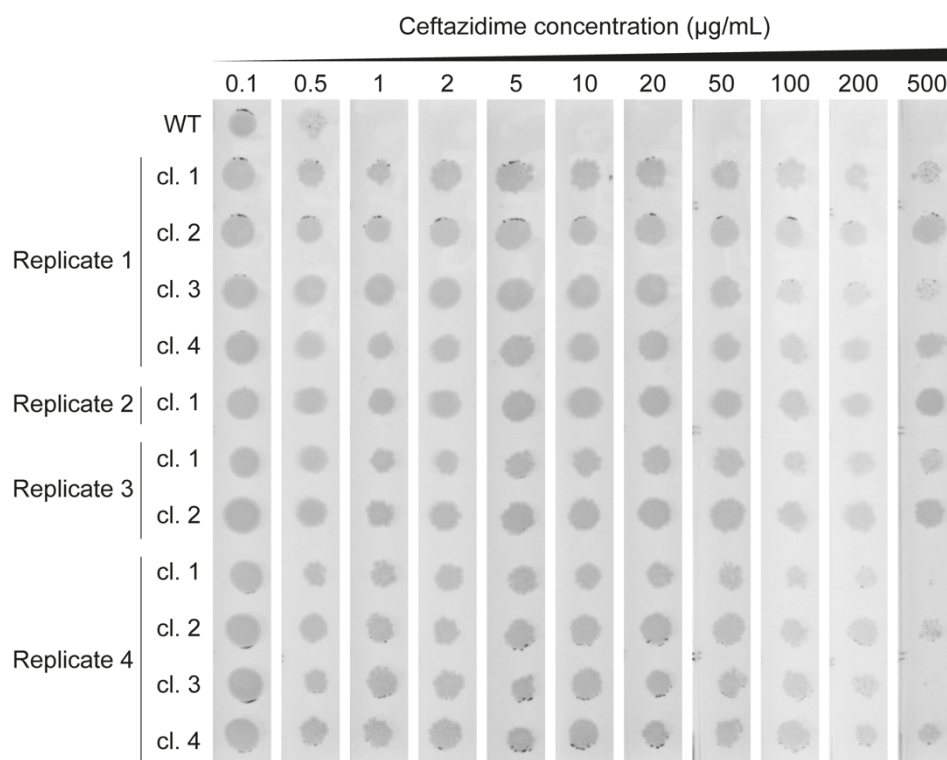


Supplementary Figure 18. Concentrations of ceftazidime used at each passage of the continuous evolution.



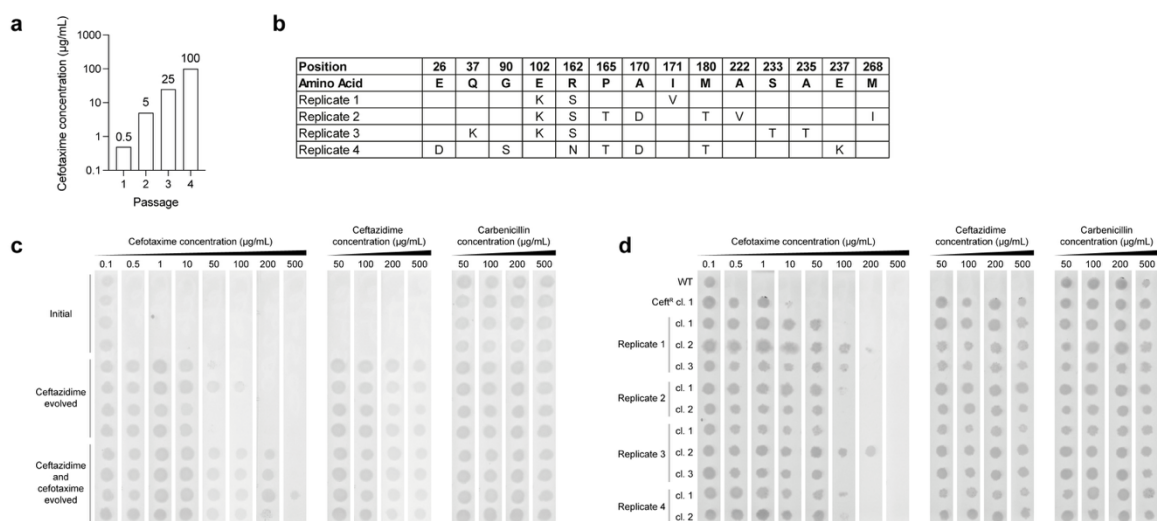
Supplementary Figure 19. Ceftazidime resistance of pools of cells harboring evolved replicon-encoded TEM-1 β -lactamase variants.

After passaging cells containing the N62D/F65S DNAP mutant and a β -lactamase-encoding Φ 29 synthetic replicon in increasingly higher concentrations of ceftazidime, we obtained pools of mutants able to sustain growth at higher ceftazidime concentrations. The first pool before and after evolution is also shown in Figure 4.



Supplementary Figure 20. Ceftazidime resistance of cloned TEM-1 β -lactamase variants.

Mutants obtained from passaging a β -lactamase-encoding Φ 29 synthetic replicon maintained by the error-prone N62D/F65S DNAP mutant were cloned into colE1 circular plasmids and transformed into DH10B cells to assess the level of ceftazidime resistance conferred. The mutations found in each tested variant are listed in Supplementary Data 2 – the spots are arranged from top to bottom in increasing numerical order of the variants listed in the table.



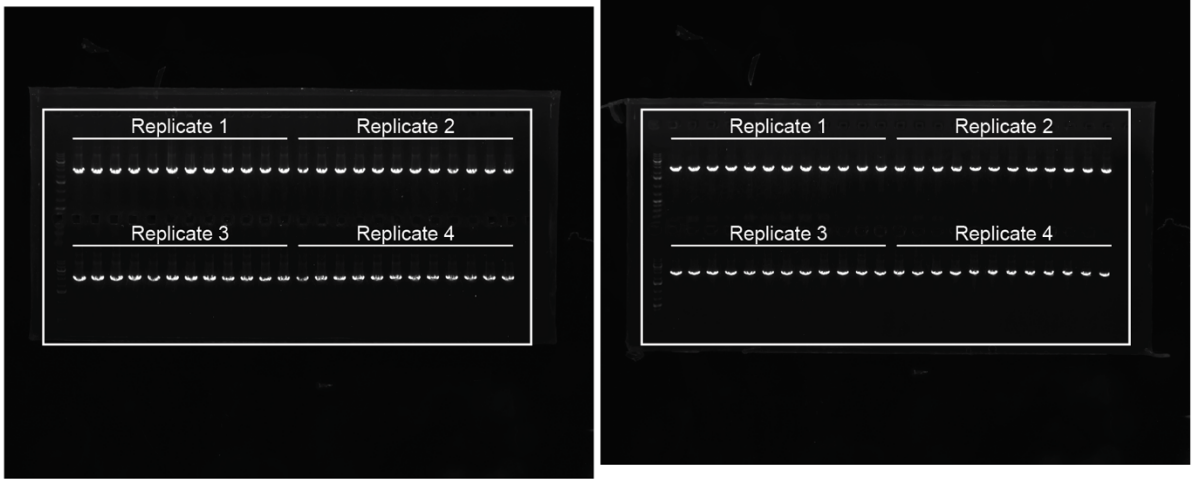
Supplementary Figure 21. Continuous evolution of multi-resistant TEM-1 β -lactamase using the Φ 29-based replication system.

After passing cells containing the N62D/F65S DNAP mutant and a β -lactamase-encoding Φ 29 synthetic replicon in increasingly higher concentrations of ceftazidime, we obtained pools of mutants able to sustain growth at higher ceftazidime concentrations. **a**, These pools of ceftazidime-resistant cells were then passaged in increasingly higher concentrations of cefotaxime while simultaneously selecting with ceftazidime and carbenicillin at 100 μ g/mL. **b**, Mutations, relative to the WT sequence shown above, found in the consensus of each replicate after passaging in cefotaxime. **c**, After passaging cells containing the N62D/F65S DNAP mutant and a β -lactamase-encoding Φ 29 synthetic replicon in increasingly higher concentrations of cefotaxime with steady concentrations of ceftazidime and carbenicillin, we obtained pools of mutants able to sustain growth at higher cefotaxime concentrations while maintaining ceftazidime and carbenicillin resistance. Shown are pools of cells before evolution (initial), after evolution for ceftazidime resistance (ceftazidime evolved) and after the final evolution for cefotaxime resistance (ceftazidime and cefotaxime evolved). **d**, Mutants obtained from the cefotaxime resistance evolution were cloned into colE1 circular plasmids and transformed into DH10B cells to assess the level of ceftazidime, cefotaxime and carbenicillin resistance conferred. Clone 1 from the ceftazidime resistant mutants is included as a control. The mutations found in each tested variant are listed in Supplementary Data 2.

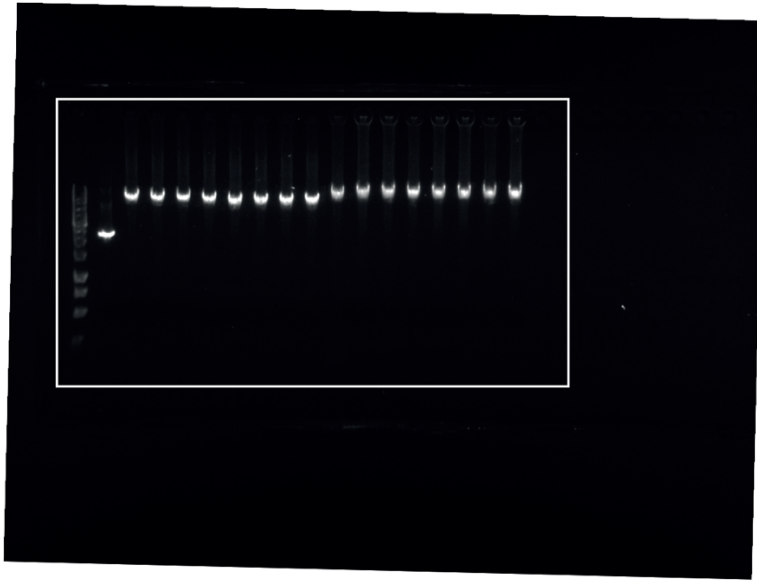
Unprocessed gel images



Unprocessed gel image from Supplementary Figure 1 with the cropped areas indicated by the white boxes.



Unprocessed gel images from Supplementary Figure 5 with the cropped areas indicated by the white boxes.



Unprocessed gel images from Supplementary Figure 8 with the cropped areas indicated by the white boxes.

Supplementary References

- 1 Salk, J. J., Schmitt, M. W. & Loeb, L. A. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. *Nat Rev Genet* **19**, 269-285, doi:10.1038/nrg.2017.117 (2018).