genes

Origins left, right and centre: increasing the number of initiation sites in the *Escherichia coli* chromosome

Juachi U. Dimude[1], Monja Stein[1], Ewa E. Andrzejewska[1], Mohammad S. Khalifa[1], Alexandra Gajdosova[1], Renata Retkute[2], Ole Skovgaard[3] and Christian J. Rudolph[1,*]

*Corresponding author: christian.rudolph@brunel.ac.uk

[1]Division of Biosciences, College of Health and Life Sciences, Brunel University London, Uxbridge, UB8 3PH, UK

[2]School of Life Science, The University of Warwick, Gibbet Hill Campus, Coventry, CV4 7AL, UK

[3]Department of Science and Environment, Roskilde University, DK-4000 Roskilde, Denmark

ABSTRACT

The bacterium *Escherichia coli* contains a single circular chromosome with a defined architecture. DNA replication initiates at a single origin called *oriC*. Two replication forks are assembled and proceed in opposite directions until they fuse in a specialised zone opposite the origin. This termination area is flanked by polar replication fork pause sites that allow forks to enter but not to leave. Thus, the chromosome is divided into two replichores, each replicated by a single replication fork. Recently we analysed replication parameters in *E. coli* cells in which an ectopic origin termed *oriZ* was integrated in the right-hand replichore. Both head-on replication-transcription conflicts at highly transcribed *rrn* operons and the replication fork trap were identified as major obstacles to replication. Here we describe replication parameters in cells with ectopic origins termed *oriX* and *oriY* integrated into the left-hand replichore and a triple origin construct with *oriX* integrated in the left-hand and *oriZ* in the right-hand replichore. Our data again highlight both replication-transcription conflicts and the replication fork trap as important obstacles to DNA replication and we describe a number of spontaneous large genomic rearrangements which successfully alleviate some of the problems arising from having an additional origin in an ectopic location. But our data reveal additional factors that impact on efficient chromosome duplication, highlighting the complexity of chromosomal architecture.

INTRODUCTION

The ability to accurately duplicate the genetic material and faithfully transmit it into daughter cells is a fundamental necessity of life. An important regulatory step for the initiation of DNA duplication process in all organisms is the assembly of fully functional replisomes at defined origin sequences [1,2]. While eukaryotic cells replicate their genomes from hundreds or thousands of origins [1], the number of initiation sites in bacteria is mostly restricted to a single origin per chromosome (*oriC*) [3,4]. In *Escherichia coli* initiation of DNA replication at *oriC* is tightly controlled by the main initiator protein DnaA, which facilitates recruitment of two replisomes [2,5–7]. These replisomes proceed in opposite directions around the circular chromosome with very high speed and accuracy until they eventually fuse within a specialised termination area opposite the origin (Figure 1A) [8,9]. The area is flanked by 10 primary *ter* sequences *A*–*J*. If bound by Tus protein, these *ter* sites form polar traps that allow forks to enter but not to leave [8,10,11]. The *E. coli* chromosome is thereby divided into two replichores, each being replicated by a single replication fork [8,10–12].

Bacteria can tolerate the integration of a second replication origin or movement of the origin into an ectopic location, but both scenarios cause serious problems. Movement of *oriC* in *Bacillus subtilis* to an ectopic location revealed that forks replicating the chromosome in an orientation opposite to normal was significantly slowed at highly transcribed regions such as the *rrn* operons [13,14], supporting the idea that head-on collisions between replication

and transcription are problematic for ongoing DNA replication [15,16]. Introduction of a second replication origin also appears to be difficult to tolerate. Integration of an IPTG-inducible plasmid origin ~450 kb away from *oriC* was shown to be active, but repressed firing of *oriC* [17].

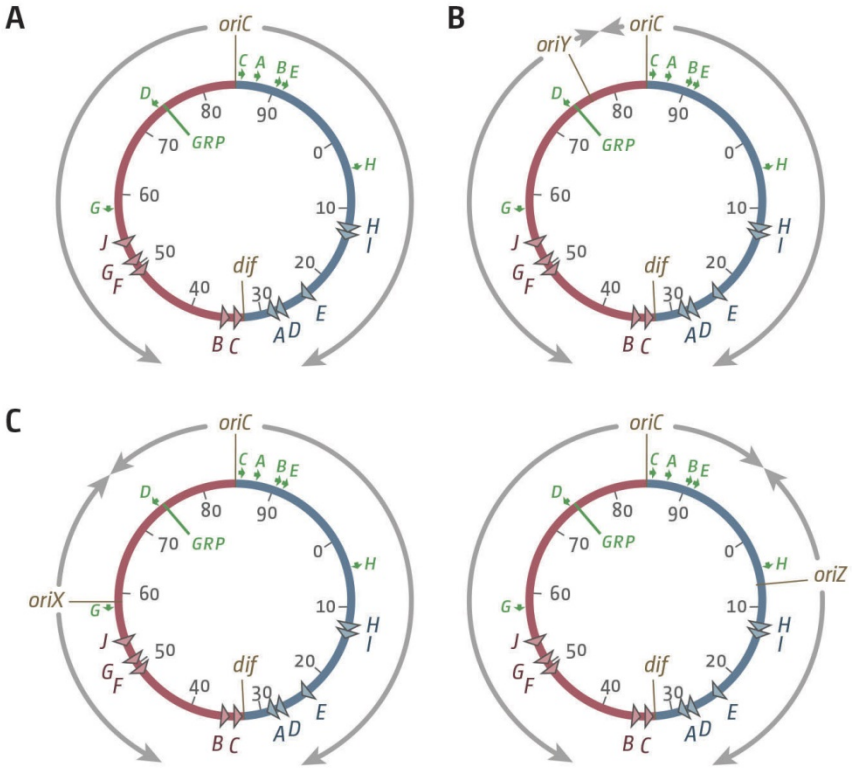


Figure 1. Schematic representation of the replichore arrangement of *E. coli* chromosomes with ectopic replication origins in different locations. **A)** Normal replichore arrangement in *E. coli*. The origin, *oriC*, and the *dif* chromosome dimer resolution site are indicated. *ter* sites are shown by triangles and identified by their corresponding letter ("A" indicates the *terA* site). The numbers represent the minutes of the standard genetic map (0–100 min). Green arrows represent location and direction of transcription of the 7 *rrn* operons *A–E, G* and *H*. The location marked GRP indicates a tight cluster of genes coding for ribosomal proteins, all of which are transcribed co-directionally to replication coming from *oriC*. **B)** Integration site of a 5 kb *oriC* fragment termed *oriY* into *malT* upstream of the *rrnD* operon. **C)** Integration sites of 5 kb *oriC* fragments into *pheA* upstream of the *rrnG* operon, termed *oriX* (this study), and near the *lacZYA* operon, termed *oriZ* [18,19].

In a recent study Wang and colleagues reported the integration of a 5 kb *oriC* fragment called *oriZ* near the *lac* operon at 7.4 min into the *E. coli* chromosome, half way into the right-hand replichore (Figure 1C) [19]. *oriC*$^+$ *oriZ*$^+$ cells grew with doubling times similar to wild type cells and cell biological observations confirmed that both origins fire simultaneously [19]. The authors also observed that in *ΔoriC oriZ*$^+$ cells, in which the chromosome is replicated exclusively from the ectopic origin, the doubling time is only marginally longer than in wild type cells [19], much in contrast to the studies in *B. subtilis* [13,14]. When we re-generated the relevant strains to study their replication dynamics, we found that the doubling time of *ΔoriC oriZ*$^+$ cells was increased from 20 to over 40 min, demonstrating that cells seriously struggle to grow. The replication profiles of these strains revealed two major

obstacles to replication. Firstly, the ectopic *oriZ* disrupts the normal replichore arrangement, with the clockwise replication fork reaching the termination area much quicker than the counter clockwise fork coming from *oriC*. Consequently, the vast majority of forks are blocked by the replication fork trap. Secondly, replication initiated at *oriZ* and traversing the chromosome opposite to normal is also significantly inhibited by the highly transcribed *rrn* operons *rrnH* and *rrnCABE*, all of which are transcribed co-directionally with replication coming from *oriC* [18], in line with the results in *B. subtilis* [13,14]. Our data show that the slow growth of *ΔoriC oriZ*$^+$ cells can be partially suppressed by a) inactivation of the replication fork trap by deletion of *tus* and b) an *rpoB*35* point mutation, which reduces the stability of RNA polymerase-DNA complexes, thereby alleviating conflicts between replication and transcription [18]. However, when we investigated why the original *ΔoriC oriZ*$^+$ construct by Wang and colleagues [19] was growing so quickly, we found a different suppressor mutation altogether: the strain carried a gross chromosomal rearrangement that inverted almost the entire portion of the chromosome that would otherwise have been replicated in the wrong orientation from *oriZ,* including the *rrnCABE* operon cluster, thereby re-aligning replication and transcription [18].

This study describes attempts to integrate ectopic replication origins at two defined locations into the opposite, left-hand replichore. In contrast to *rrn* operons *CABE* and *H* in the right-hand replichore, the left-hand replichore only contains *rrn* operons *D* and *G*, as well as a cluster of genes encoding for ribosomal proteins (Figure 1). We therefore hypothesised that integration of an ectopic origin into the left-hand replichore might be less problematic. However, the results presented suggest the opposite. Integration of an active 5 kb origin fragment termed *oriY* upstream of *rrnD* was not possible. Given that no *rrn* operons would be encountered head-on by replication starting from this location, the inability to integrate a functional origin in this location suggests that multiple factors must contribute towards origin activity. Integration of a functional 5 kb origin fragment termed *oriX* just upstream of *rrnG* into the left-hand replichore was successful, but *ΔoriC oriX*$^+$ cells grew even more slowly than *ΔoriC oriZ*$^+$ cells and rapidly accumulated suppressor mutations, some of which are characterised. Finally, we report the successful construction of *oriC*$^+$ *oriX*$^+$ *oriZ*$^+$ cells. In this triple origin background all origins are active in principle, but both ectopic origins show a reduced activity relative to *oriC*. Our results re-iterate that both the termination area and head-on replication-transcription encounters act as severe obstacles for chromosomal replication if the replichore arrangement is asymmetric. But our inability to integrate a functional *oriY*, the slow growth of *ΔoriC oriX*$^+$ cells and the preference for *oriC* in triple origin cells strongly support the idea that a number of different factors influences origin activity and successful genome duplication in the presence of additional ectopic replication initiation sites.

MATERIAL & METHODS

Bacterial strains and general methods

For *Escherichia coli* K12 strains see Table 1. Strains were constructed via P1*vir* transductions [20] or by single-step gene disruptions [21].

Table 1: *Escherichia coli* K-12 strains

Strain number	Relevant Genotype[a]	Source
General P1 donors		
WX297	AB1157 *oriZ-<kan>*	[19]
RRL190	AB1157 *<kan>-ypet-dnaN*	[19]
RUC1593	DY330 *pheA::oriX-cat*	This study
MG1655 derivatives		
MG1655	F⁻ *rph-1*	[22]
AS1062	*<kan>-ypet-dnaN*	MG1655 × P1.RRL190 to Km^r
JD1181	*ΔlacIZYA pheA::oriX-cat*	TB28 × P1.RUC1593 to Cm^r
JD1187	*ΔlacIZYA pheA::oriX-cat ΔoriC::kan*	JD1181 × P1.RCe576 to Km^r
JD1190	*rpoB*35 ΔlacIZYA pheA::oriX-cat*	N5925 × P1.RUC1593 to Cm^r
JD1197	*rpoB*35 ΔlacIZYA pheA::oriX-cat ΔoriC::kan*	JD1190 × P1.RCe576 to Km^r
JD1203	*ΔlacIZYA pheA::oriX-cat tus1::dhfr*	JD1181 × P1.N6798 to Tm^r
JD1205	*rpoB*35 ΔlacIZYA pheA::oriX-cat tus1::dhfr*	JD1190 × P1.N6798 to Tm^r
JD1208	*ΔlacIZYA pheA::oriX-cat tus1::dhfr ΔoriC::kan*	JD1203 × P1.RCe576 to Km^r
JD1209	*rpoB*35 ΔlacIZYA pheA::oriX-cat tus1::dhfr ΔoriC::kan*	JD1205 × P1.RCe576 to Km^r
JD1332	*ΔlacIZYA pheA::oriX-cat oriZ-<kan>*	JD1181 × P1.WX297 to Km^r
JD1333	*ΔlacIZYA pheA::oriX-cat oriZ-<kan>*	JD1181 × P1.WX297 to Km^r
JD1336	*ΔlacIZYA oriZ-<kan>*	TB28 × P1.WX297 to Km^r
JD1339	*ΔlacIZYA oriZ-<>*	JD1336 × pCP20 to Km^s Ap^s
JD1341	*ΔlacIZYA oriZ-<> pheA::oriX-cat*	JD1339 × P1.RUC1593 to Cm^r
JD1343	*ΔlacIZYA oriZ-<> pheA::oriX-cat ΔoriC::kan*	JD1341 × P1.RCe576 to Km^r
JJ1359	*ΔlacIZYA dam1::kan ΔrecG::apra tus1::dhfr*	[23]
N4560	*ΔrecG265::cat*	[24]
N5925	*rpoB*35 ΔlacIZYA*	[25]
N6798	*ΔrecG265::cat tus1::dhfr*	N4560 × P1.JJ1359 to Tm^r
RCe504	*oriZ-<cat>*	[18]
RCe576	*rpoB*35 oriZ-cat-frt tus1::dhfr ΔoriC::kan*[b]	[18]
RCe749	*oriZ-<cat> <kan>-ypet-dnaN*	RCe504 × P1.AS1062 to Km^r
RCe751	*ΔlacIZYA pheA::oriX-cat <kan>-ypet-dnaN*	JD1181 × P1.AS1062 to Km^r
RCe753	*ΔlacIZYA oriZ-<> pheA::oriX-cat <kan>-ypet-dnaN*	JD1341 × P1.AS1062 to Km^r
TB28	*ΔlacIZYA*	[26]

a – Only the relevant additional genotype of the derivatives is shown. The abbreviations *kan*, *cat* and *dhfr* refer to insertions conferring resistance to kanamycin (Km^r), chloramphenicol (Cm^r) and trimethoprim (Tm^r), respectively. *frt* stands for the 34 bp recognition site of the FLP/*frt* site-directed recombination system.

b – *ΔoriC* refers to a replacement of the entire origin region (754 bp) including DnaA boxes and 13mers as well as the entire *mioC* gene by a kanamycin resistance cassette [23].

Growth media

Luria broth (LB) and agar was modified from Luria and Burrous [27] as follows: 1% tryptone (Bacto™, BD Biosciences), 0.5% yeast extract (Bacto™, BD Biosciences) and 0.05% NaCl (Sigma Aldrich). The pH was adjusted to 7.4. M9 minimal medium (Bacto™, BD Biosciences) contained 15 g/L KH_2PO_4, 64 g/L Na_2HPO_4, 2.5 g/l NaCl and 5.0 g/L NH_4Cl. Before use, $MgSO_4$, $CaCl_2$ and glucose were added from sterile-filtered stock solutions to final concentrations of 2 mM, 0.1 mM and 0.2%, respectively, according to the manufacturer's recommendation. Doubling times of MG1655 in our growth media were 19.3 ± 1.7 min in LB and 68.8 ± 6.2 min in M9 glucose.

Marker frequency analysis by deep sequencing

Marker frequency analysis by deep sequencing was performed as described before [18]. See the Supplementary Methods section for a detailed description. All relevant raw sequencing data can be accessed at the European Nucleotide Archive (http://www.ebi.ac.uk/ena/data/view/PRJEB9476)

LOESS regression

A LOESS regression allows for a simplified visualisation of complex data sets. For a LOESS regression relatively simple models are fitted to defined small subsets of data points in order to develop a function describing the deterministic part of the variation in the data. Weighted least squares are used to fit a low-degree polynomial to a specified percentage of data points. Data points are weighted by a smooth decreasing function of their distance to the smoothed point, giving more weight to points closer to the point whose response is being estimated, while less weight is given to points further away. We used a second order polynomial for local fit, tricube as weight function and set a fraction of data used for smoothing to 10%, which corresponds to a smoothing window around 460 kbp [28]. To account for circularity of the chromosome, periodic boundary conditions were used.

Growth curves

Samples from cultures of a strain grown over night in LB broth were diluted 100-fold in fresh broth and incubated with vigorous aeration at 37°C until A_{600} reached 0.48. The only exception were *ΔoriC oriX*$^+$ backgrounds, for which growth was initiated from a single colony from a streak plate to avoid suppressors formed in the overnight culture outgrowing the slow growing *ΔoriC oriX*$^+$ cells. Upon reaching an A_{600} of 0.48 the culture was diluted 100-fold in pre-warmed fresh broth and grown under identical conditions. Samples were taken every 30

min, diluted to 10^{-7} in M9 minimal medium without added glucose and 10 µl aliquots of each dilution dropped onto LB agar plates. For each dilution series 2 sets of drops were spotted. Colonies were counted after incubation for 18–24 h at 37°C. Mean colony numbers from both spots were calculated and a growth curve plotted. A suitable period where growth was exponential was selected (usually between 60 and 180 min following dilution into fresh LB). For calculation of the doubling time the LINEST function in MS Excel was used to determine linear regression parameters for data points which were calculated from averages per time point of between three and eight independent experiments. The doubling times of strains shown in tables 2 and 3 were carried out in sets. Thus, relevant controls, such as MG1655, *oriC*$^+$ *oriZ*$^+$ and *oriC*$^+$ *oriX*$^+$, were always measured in parallel to the strains of interest, explaining the slight variations of the doubling times of these strains in the respective tables. Doing so allowed us to largely avoid the comparison of doubling times generated under potentially slightly varying conditions.

Mathematical modelling

See Supplementary Methods for a detailed description of the mathematical modelling.

RESULTS

Ectopic replication origins in the left-hand replichore

Previously we investigated replication parameters in strains in which a 5 kb *oriC* fragment called *oriZ* was integrated about 1 Mbp away from the native *oriC* in the right-hand replichore [18,19]. Here we attempted to integrate another copy of the 5 kb *oriC* fragment in two separate locations into the left-hand replichore. Our previous study had identified *rrn* operons *C*, *A*, *B*, *E* and *H* as major obstacles to the progression of replication forks coming from the ectopic origin [18]. We speculated that the opposite replichore might pose less problems, as only 2, rather than 5, *rrn* operons are present (Figure 1). We attempted to integrate one 5 kb *oriC* fragment called *oriY* into the *malT* gene at 76.5 min, which is upstream of *rrnD*. The location allows forks coming from *oriY* to progress without any *rrn* operons in their way (Figure 1B). A second construct termed *oriX* was integrated into *pheA* at 59 min, an integration location that is roughly equivalent to the *oriZ* location in terms of replichore length (Figure 1C). The *pheA* gene is just upstream of *rrnG*. Thus, only *rrn* operon *D* and a cluster of genes coding for ribosomal proteins will be encountered in a direction opposite to normal in *ΔoriC oriX*$^+$ cells (Figure 1).

Both chromosomal integrations resulted in colonies with the correct antibiotic resistance. However, deletion of *oriC* was only possible in *oriC*$^+$ *oriX*$^+$ cells; we failed to generate a *ΔoriC oriY*$^+$ construct. PCR verification of two of the *oriY* constructs demonstrated one partial truncation and one complete loss of the *oriC* core elements (Figure 2), explaining the lack of functionality. A repeat of the chromosomal integration directly into MG1655 again did not result in constructs with a functional *oriY*. We do not currently know what is causing

the inability to integrate *oriY* into the chromosome, given *oriX*, which was amplified from the same template, could be integrated without difficulty.

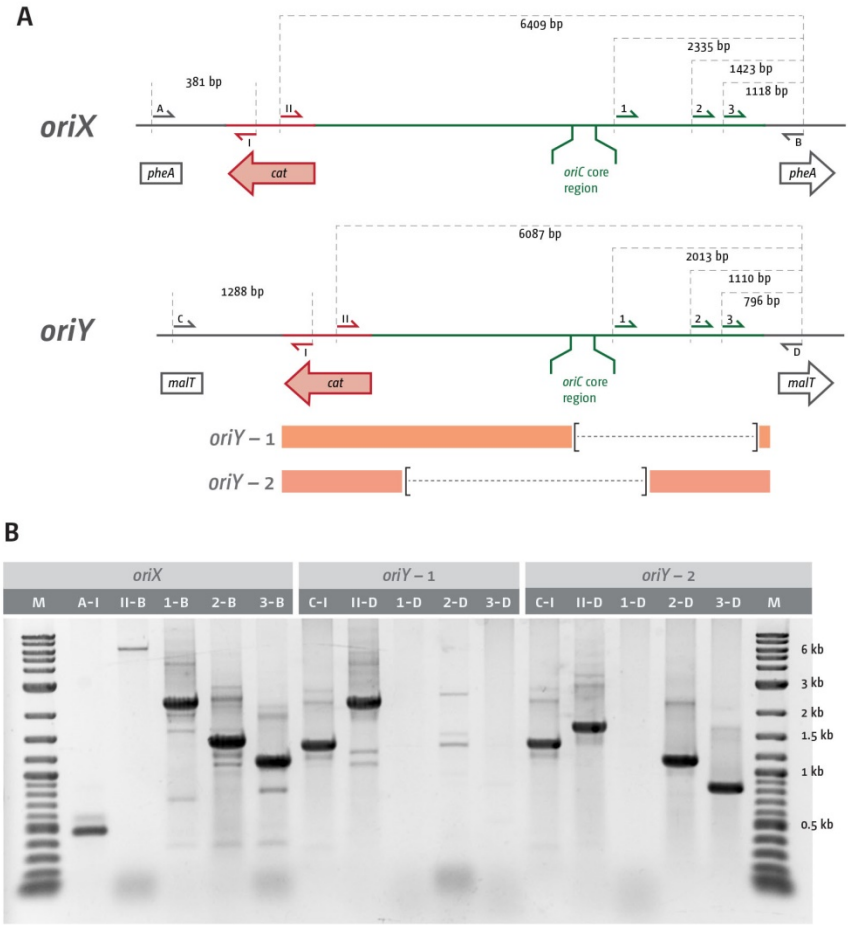


Figure 2. PCR confirmation of *oriX* and *oriY* integration cassettes into the chromosome. **A)** Schematic representation of the integration region following successful integration of *oriX* into *pheA* or *oriY* into *malT*, respectively. Primers are identified according to their position with letters, numbers or roman numerals. Primer binding sites are indicated. The orange bars below the *oriY* scheme indicate the likely regions where truncation has taken place, taking into consideration the overall length of the integrated region as well as the presence and absence of defined primer binding sites as shown in B). The dashed lines represent the approximate sizes of truncations. **B)** Agarose gel electrophoresis of PCRs with primers highlighted in A) on templates in which either *oriX* or *oriY* is integrated into the chromosome. Sizes of relevant marker fragments (2-log kb ladder, NEB) are indicated. The primer combinations used for the individual PCRs are given directly above the relevant lane (primers A and I shown in A) are given as A-I). An inverted gel image is shown for clarity.

oriX is active in double-origin cells

Marker frequency analysis (MFA) was used to investigate the replication profile of *oriC*$^+$ *oriX*$^+$ cells (Figure 3A). Given that all replication profiles of our previous *oriZ* study were generated from cultures grown in LB broth [18], all samples were grown under similar conditions to enable a direct comparison. The replication profile of *oriC*$^+$ *oriX*$^+$ cells showed similar features to the previously obtained *oriC*$^+$ *oriZ*$^+$ profile (Figures 3A). The MFA confirmed that *oriX* was active (Figure 3A ii).

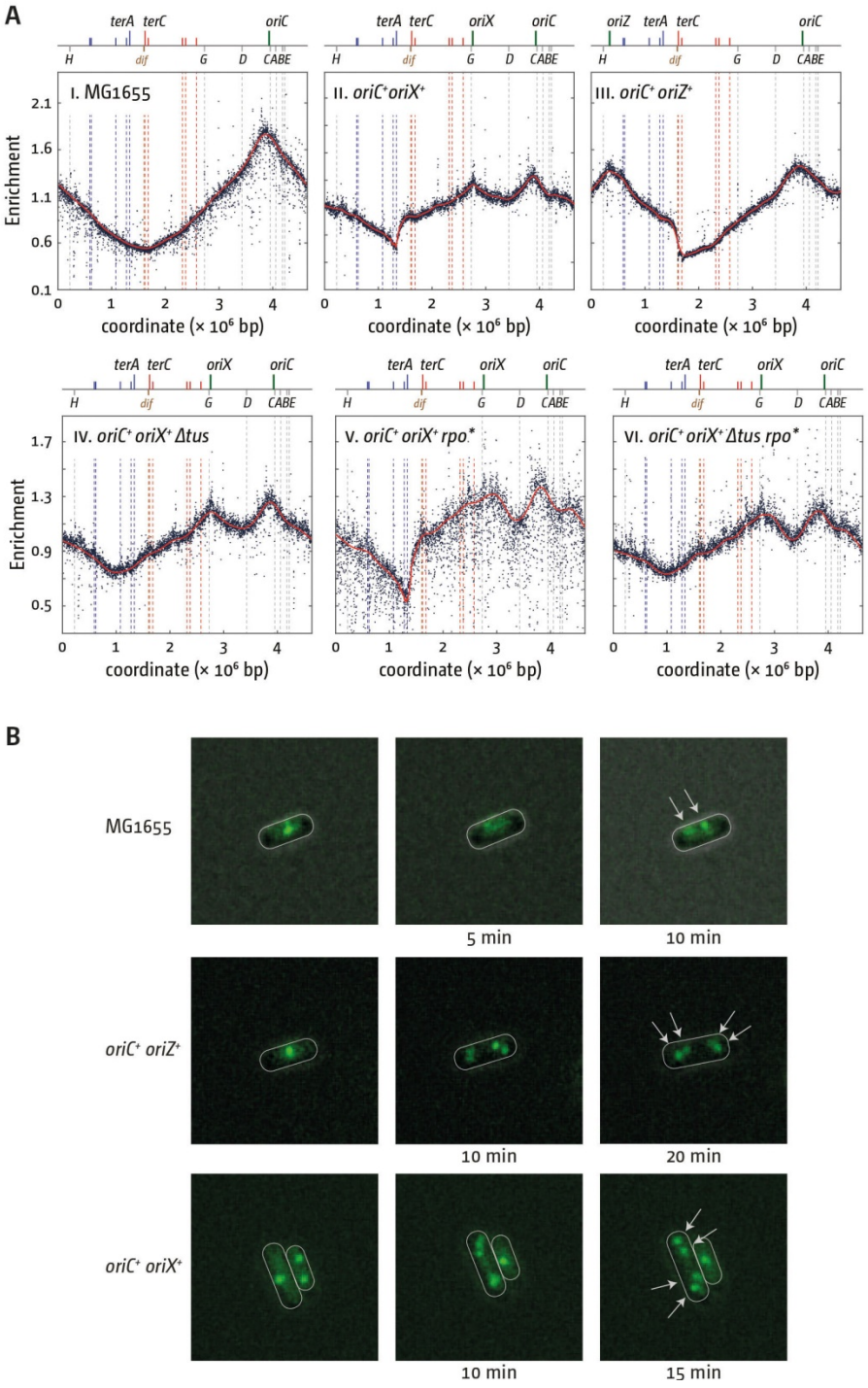


Figure 3. Replication dynamics in *E. coli* cells with one and two replication origins. **A)** Marker frequency analysis of *E. coli oriC*⁺, *oriC*⁺ *oriX*⁺ and *oriC*⁺ *oriZ*⁺ cells and impact of Δ*tus* and an *rpo** point mutation on these cells. The number of reads (normalised against reads for a stationary phase wild type control) is plotted against the chromosomal location. A schematic representation of the *E. coli* chromosome showing positions of *oriC* and *oriX* (green line) and *ter* sites (above) as well as *dif* and *rrn* operons A–E, G and H (below) is shown above the plotted data. The strains used were MG1655 (*oriC*⁺), RCe504 (*oriC*⁺ *oriZ*⁺), JD1181 (*oriC*⁺ *oriX*⁺), JD1203 (*oriC*⁺ *oriX*⁺ Δ*tus*), JD1190 (*oriC*⁺ *oriX*⁺ *rpo**) and JD1205 (*oriC*⁺ *oriX*⁺ Δ*tus rpo**). **B)** Visualisation of replisomes (Ypet-DnaN) in wild type, *oriC*⁺ *oriX*⁺ and *oriC*⁺ *oriZ*⁺ cells. Cells were grown in M9 minimal salts medium with 0.2% glucose and transferred onto a thin agarose pad of the same medium on a microscopy slide (see Material & Methods). Slides were transferred into a chamber heated to 37°C and fluorescent foci in single cells tracked over time. The strains used were AS1062 (*ypet-dnaN*), RCe749 (*oriC*⁺ *oriZ*⁺ *ypet-dnaN*) and RCe751 (*oriC*⁺ *oriX*⁺ *ypet-dnaN*).

There appears to be a minor difference in peak height between *oriC* and *oriX*. Our subsequent analysis has shown that this is caused by the column purification procedure to extract genomic DNA for Deep Sequencing. Insufficient proteolytic digest causes DNA fragments to be lost in areas where protein-DNA complexes are particularly tight or frequent, such as *rrn* operons or *ter*/Tus complexes, as proteins still bound to DNA fragments are eluted from the DNA-binding column (see Supplementary Methods and Suppl. Figure 1). For *oriX* it is the proximity of *rrnG* that causes a mild under-representation of the region, which results in a reduced peak height. After identifying this issue, re-sequencing of an *oriC*$^+$ *oriX*$^+$ construct following phenol-chloroform extraction of genomic DNA demonstrated that both *oriC* and *oriX* are active at similar frequencies (Suppl. Figure 2).

To confirm that both origins are simultaneously active, a strain was used in which the bright YFP derivative YPet was fused to the N-Terminus of the β-sliding clamp, encoded by the *dnaN* gene, as reported [19]. To avoid the complexity of overlapping rounds of DNA replication, cells were grown in M9 minimal medium with 0.2% glucose (called M9 hereafter; see Material & Methods). Time-lapse microscopy of otherwise wild type cells shows that under these conditions replisomes are disassembled upon completion of synthesis before replication is initiated at the segregated copies of *oriC* (Figure 3B). Time-lapse analysis of both *oriC*$^+$ *oriZ*$^+$ and *oriC*$^+$ *oriX*$^+$ cells showed that both origins are active, as shown before for *oriC*$^+$ *oriZ*$^+$ cells [19], ruling out that either *oriX*/*oriZ* or *oriC* fire independently but with similar frequencies.

Termination and replication-transcription conflicts in double-origin cells

Replication initiated at *oriX* and proceeding counter clockwise will reach the termination area much earlier than forks coming from *oriC* and consequently forks will be blocked at the *terA*/Tus complex, the first *ter*/Tus complex encountered in blocking orientation, which results in the clearly visible step of the replication profile at *terA* (Figure 3A ii). A similar step is observed in *oriC*$^+$ *oriZ*$^+$ cells at *terC*/*B* (Figure 3A iii) [18]. Deletion of *tus* in *oriC*$^+$ *oriX*$^+$ cells enabled replication forks to proceed into the opposite replichore, resulting in a symmetrical replication profile (Figure 3A iv). The arithmetic mid-point between *oriC* and *oriX* is at position 1.010 Mbp, close to the measured low point of the LOESS regression at 0.991 Mbp (Suppl. Table 1). Thus, even if leaving the termination area in a direction opposite to normal, forks appear to proceed with similar speed. In line with this, the introduction of an *rpo** point mutation which decreases the stability of transcribing RNAP complexes [29], did not significantly change the location of the low point of the replication profile (Figure 3A v), suggesting that problems associated with replication-transcription encounters must be similar for both replication machineries.

Doubling times of all *oriC*$^+$ *oriX*$^+$ constructs followed trends that were similar to our previous observations in *oriC*$^+$ *oriZ*$^+$ cells (Table 2 and Figure 4). Introduction of *oriX* mildly slowed the doubling time, indicating that integration of a second replication origin interferes

in some way with the fast growth observed in wild type cells. An *rpo** point mutation was shown before to slow growth [18] and consequently a slower doubling time is seen in *oriC*$^+$ *oriX*$^+$ *rpo** cells (Table 2 and Figure 4). A *tus* deletion had little effect, but a combination of *Δtus* and *rpo** resulted in the slowest doubling time (Table 2 and Figure 4).

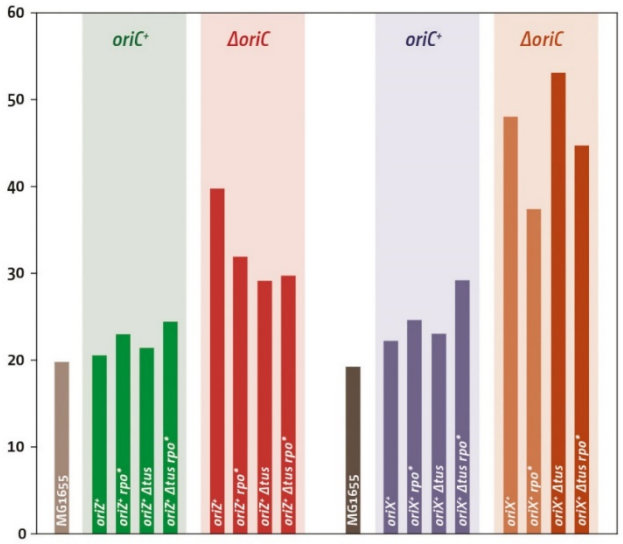


Figure 4. Comparison of doubling times of *oriZ*$^+$ constructs, as reported in [18], and *oriX*$^+$ constructs (this study). The presence or absence of *oriC* is highlighted above each group of strains. The ectopic origin and all other genotype details are identified for each strain individually. The two *ΔoriC oriX*$^+$ constructs identified by a lighter colour contained large chromosomal inversions (see main text for details). All doubling times were determined by measuring viable titres of cultures grown in LB broth (see Material and Methods for details).

Table 2: Doubling times of *E. coli* strains with an ectopic replication origin in the left replichore

Strain background	doubling time [min]	SD	r²	Doubling time *oriZ* constructs[a]
MG1655	19.3	± 1.7	0.983	19.9
oriC$^+$ *oriX*$^+$	22.3	± 1.2	0.981	20.6
ΔoriC oriX$^+$	48.1	± 5.6	0.969	39.8
oriC$^+$ *oriX*$^+$ *Δtus*	23.1	± 0.7	0.985	21.5
oriC$^+$ *oriX*$^+$ *rpoB*35*	24.7	± 1.5	0.986	23.1
oriC$^+$ *oriX*$^+$ *Δtus rpoB*35*	29.3	± 1.9	0.993	24.5
ΔoriC oriX$^+$ *Δtus*	53.2	± 9.1	0.977	29.2
ΔoriC oriX$^+$ *rpoB*35*	37.5	± 8.4	0.980	32.0
ΔoriC oriX$^+$ *Δtus rpoB*35*	44.8	± 9.2	0.99	29.8

a – doubling times as reported in [18].

Deletion of *oriC* in double-origin mutants triggers chromosomal rearrangements

In *ΔoriC oriX*$^+$ cells only *rrnD* is replicated in an orientation opposite to normal, together with a cluster of ~30 genes encoding for ribosomal proteins. However, *ΔoriC oriX*$^+$ cells had a growth rate even slower than that of *ΔoriC oriZ*$^+$ cells (Table 2 and Figure 4) [18] and

rapidly accumulated fast growing suppressor mutations (Figure 5A). Given our experience of suppressor accumulation in *ΔoriC oriZ*$^{+}$ cells we were vigilant for spontaneous suppressor mutations arising whilst generating *ΔoriC oriX*$^{+}$ constructs. Nevertheless, our *ΔoriC oriX*$^{+}$ construct contained a gross chromosomal rearrangement (GCR), inverting an ~820 kb fragment of the chromosome that spans from IS5 at 575 kb to IS5 at 1394 kb (Figure 5B i & ia; Suppl. Figure 3 for PCR verification of the inversion). This inversion spans ***all*** restrictive *ter* sites (*terA*, *D*, *E*, *H* and *I*) and flips them into permissive orientation, thereby allowing forks to leave the termination area. While the previously reported inversion that re-aligned replication and transcription in *ΔoriC oriZ*$^{+}$ cells acted as a very efficient suppressor of the slow growth phenotype [18], the *ΔoriC oriX*$^{+}$ construct containing the inverted *ter* sites (*ΔoriC oriX*^{inv}) grew slowly (Table 2 and Figure 4), suggesting that additional effects must interfere with efficient chromosome duplication. We suspect that *ΔoriC oriX*$^{+}$ cells without the GCR have an even longer doubling time or might potentially be inviable.

The doubling time of *ΔoriC oriX*$^{+}$ *Δtus* cells was roughly comparable to that of our *ΔoriC oriX*$^{+}$ construct carrying the GCR, in line with the replication fork trap not being active in both backgrounds (Table 2 and Figure 4). The doubling time of the *ΔoriC oriX*$^{+}$ *Δtus* construct was markedly longer than that of the corresponding *ΔoriC oriZ*$^{+}$ *Δtus* construct (Table 2 and Figure 4), and the replication profile of *ΔoriC oriX*$^{+}$ *Δtus* cells (Figure 5B ii) revealed a discontinuity that indicates a duplication of a 175 kb stretch spanning the *rrn* operons *A* and *B*. This GCR turned out to be a spontaneous mutation in the culture grown for the preparation of genomic DNA, but not in our stock culture, as a second replication profile showed no GCR (Suppl. Figure 4). This suggests that the measured doubling time (Table 2 and Figure 4) was correctly determined.

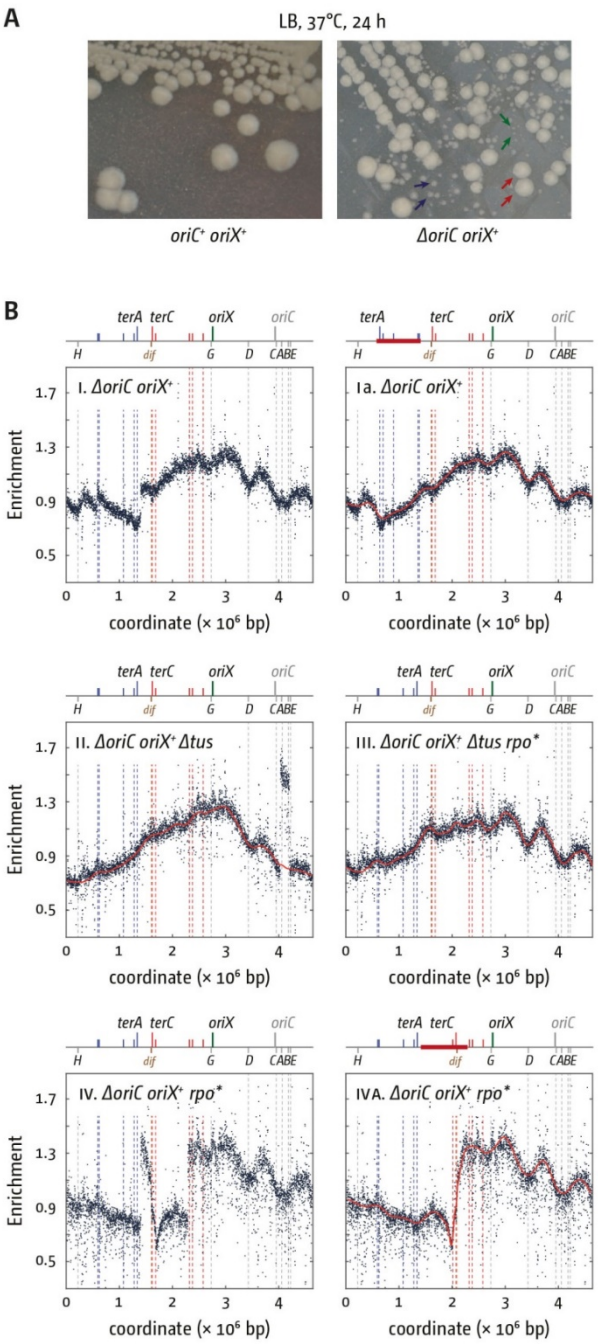


Figure 5. Growth and replication profiles of *E. coli* cells replicating from a single ectopic replication origin. **A)** Large colony variants due to accumulation of suppressor mutations in Δ*oriC oriX*⁺ cells. Shown is a streak to single colonies of an overnight culture of both constructs. While an *oriC*⁺ *oriX*⁺ strain shows largely uniform colony sizes with only some variation due to colony density, a Δ*oriC oriX*⁺ constructs shows small, medium and large colonies, as highlighted by green, blue and red arrows, respectively. The strains used were JD1181 (*oriC*⁺ *oriX*⁺) and JD1187 (Δ*oriC oriX*⁺). **B)** Replication profiles of *E. coli* cells with a single ectopic replication origin. Shown is the marker frequency analysis of *E. coli* Δ*oriC oriX*⁺, Δ*oriC oriX*⁺ Δ*tus*, Δ*oriC oriX*⁺ *rpo** and Δ*oriC oriX*⁺ Δ*tus rpo** cells. The number of reads (normalised against reads for a stationary phase wild type control) is plotted against the chromosomal location. A schematic representation of the *E. coli* chromosome showing positions of *oriC* and *oriZ* (green line) and *ter* sites (above) as well as *dif* and *rrn* operons *A–E, G* and *H* (below) is shown above the plotted data. Clear discontinuities of the profiles can be seen in panels i, ii and iv. For panels i and iv these are due to large inversions, as highlighted by the continuous replication profile that results if the area highlighted in red in the schematic representation of the chromosome is inverted. The strains used were JD1187 (Δ*oriC oriX*⁺), JD1208 (Δ*oriC oriX*⁺ Δ*tus*), JD1197 (Δ*oriC oriX*⁺ *rpo**) and JD1209 (Δ*oriC oriX*⁺ Δ*tus rpo**).

To determine the impact of replication-transcription conflicts that occur when part of the chromosome is replicated in an orientation opposite to normal a *ΔoriC oriX*$^{+}$ *rpo** construct was generated. This construct indeed showed a faster doubling time (Table 2 and Figure 4), but contained yet another GCR. An 895 kb section of the chromosome spanning from IS5 at 1394 kb to IS5 at 2288 kb was inverted (Figure 5B iv & iva; see Suppl. Figure 3 for PCR verification of the inversion). In this case the GCR was observed in two independent MFAs, suggesting that it has arisen during the construction process. Its presence prevents a detailed analysis. However, the doubling time of the *ΔoriC oriX*$^{+}$ *rpo** construct carrying the inversion is faster than those of *ΔoriC oriX*^{inv} and *ΔoriC oriX*$^{+}$ *Δtus* (Table 2), suggesting that the *rpo** mutation still improves growth. Indeed, introduction of an *rpo** point mutation into *ΔoriC oriX*$^{+}$ *Δtus* cells resulted in a decrease of the doubling time (Table 2), in line with the idea that replication-transcription conflicts contribute to the slow growth phenotype of *ΔoriC oriX Δtus* cells. The *ΔoriC oriX*$^{+}$ *Δtus rpo** construct is the only construct without GCRs, similar to *ΔoriC oriZ*$^{+}$ *Δtus rpo** cells in which suppressor accumulation is markedly reduced [18]. However, the growth rate of *ΔoriC oriX*$^{+}$ *Δtus rpo** cells is still substantially slower than that of the equivalent *ΔoriC oriZ*$^{+}$ *Δtus rpo** construct (Table 2 and Figure 4), further supporting the idea that a number of factors influence the doubling time in *oriX*$^{+}$ cells.

Replication initiation in cells with a triple-origin chromosome

We wanted to investigate whether an *E. coli* chromosome with three active origins could be constructed. In *oriC*$^{+}$ *oriX*$^{+}$ *oriZ*$^{+}$ cells defined areas would be replicated opposite to normal, thereby causing some difficulties, but replication should be less asymmetric than in double origin cells. Construction of an *oriC*$^{+}$ *oriX*$^{+}$ *oriZ*$^{+}$ construct was easily achieved. However, the doubling time of this construct was longer than that of both wild type and double origin cells (Table 3) and the replication profile revealed a surprising skew in origin usage (Figure 6).

Table 3: Doubling times of *E. coli* strains with two ectopic replication origins

Strain background	doubling time [min]	SD	r²
MG1655	19.6	± 1.0	0.999
oriC^{+} *oriZ*$^{+}$	21.0	± 0.8	0.997
oriC^{+} *oriX*$^{+}$	21.8	± 0.8	0.996
oriC^{+} *oriX*$^{+}$ *oriZ*$^{+}$	22.7	± 2.5	0.994
ΔoriC oriX^{+} *oriZ*$^{+}$	35.3	± 2.6	0.990

oriC showed the highest peak height, while the peak heights of both *oriZ* and *oriX* was reduced (Figure 6 iii). As replication profiles only give an indication of origin usage within a population of cells, time-lapse fluorescence microscopy of *oriC*$^{+}$ *oriX*$^{+}$ *oriZ*$^{+}$ cells carrying YPet-DnaN was used to investigate whether there are cells in which all three origins can be active. While the signal in double origin cells produced defined foci (Figure 3B), the signal in

triple origin cells was less defined. In addition, the close proximity of multiple and less defined foci made differentiation with conventional fluorescence microscopy very difficult. Nevertheless, in some cells three separate foci were observed, suggesting that all three origins can be active at least in a fraction of cells (Figure 6B). Given the resolution limit of conventional fluorescence microscopy and the fact that the β-sliding clamp remains bound to DNA for some time after the replisome has passed [30,31], we did not attempt a detailed analysis of foci dynamics in cells, as this is unlikely to result in meaningful data. However, foci numbers in snap shots of cells in exponential phase grown in M9 minimal medium with 0.2% glucose were analysed. Fluorescent DnaN foci per cell were then counted (Figure 7A). Overall only a minor increase in the number of DnaN foci per cell was observed both in *oriC*$^+$ *oriZ*$^+$ and *oriC*$^+$ *oriX*$^+$ cells, despite the fact that time-lapse analysis shows clearly that both origins are active (Figure 3B). We believe the main reason for this is the short presence of multiple replisomes. Upon initiation of replication one replisome coming from *oriC* will replicate a relatively short stretch of 500 kb before it is met by a replisome coming from the ectopic origin. If replication proceeds with the reported 650–1000 nt · s^{-1} [32], forks will fuse after 10–12 min and disassemble, leaving two forks that move in the opposite directions, the same number as in wild type cells. Thus, in asynchronously growing cultures only a small fraction of cells will show an increased number of replisomes, which, together with the limited resolution, explains the very moderate shift in foci numbers.

One interesting feature of triple origin cells is the increase in cells with no foci, while both *oriC*$^+$ *oriX*$^+$ and *oriC*$^+$ *oriZ*$^+$ cells show a decrease in comparison to wild type cells. One explanation for this effect might be a limitation of initiation of DNA replication in triple origin cells. It was reported before that multiple chromosomal locations including the *datA* locus bind the DnaA initiator protein with high affinity [33]. Upon initiation of chromosome replication the duplication of the origin, these regions will act as a sink for DnaA, thereby reducing the concentration of free DnaA protein in the cell [34], which limits initiation of replication [35,36].

Levels of DnaA are clearly high enough to allow simultaneous initiation at two independent copies of the origin (Figure 3A and B) [18,19]. However, a third copy might cause the concentration of free DnaA to drop below the threshold level for initiation for longer, thereby limiting initiation of replication and thus leading to an increased fraction of cells with 0 foci. This effect might also explain why triple origin cells grow more slowly than both double origin constructs (Table 3 and Figure 7B). To test whether this was the case, a low copy number plasmid carrying a copy of *dnaA* under its native promoter was introduced into these cells and the doubling times measured. An increased *dnaA* copy number caused only minor reductions of the doubling time of double origin cells, but triple origin cells show a marked reduction, in line with the idea that the concentration of free DnaA becomes limiting (Figure 7B and Suppl. Table 2).

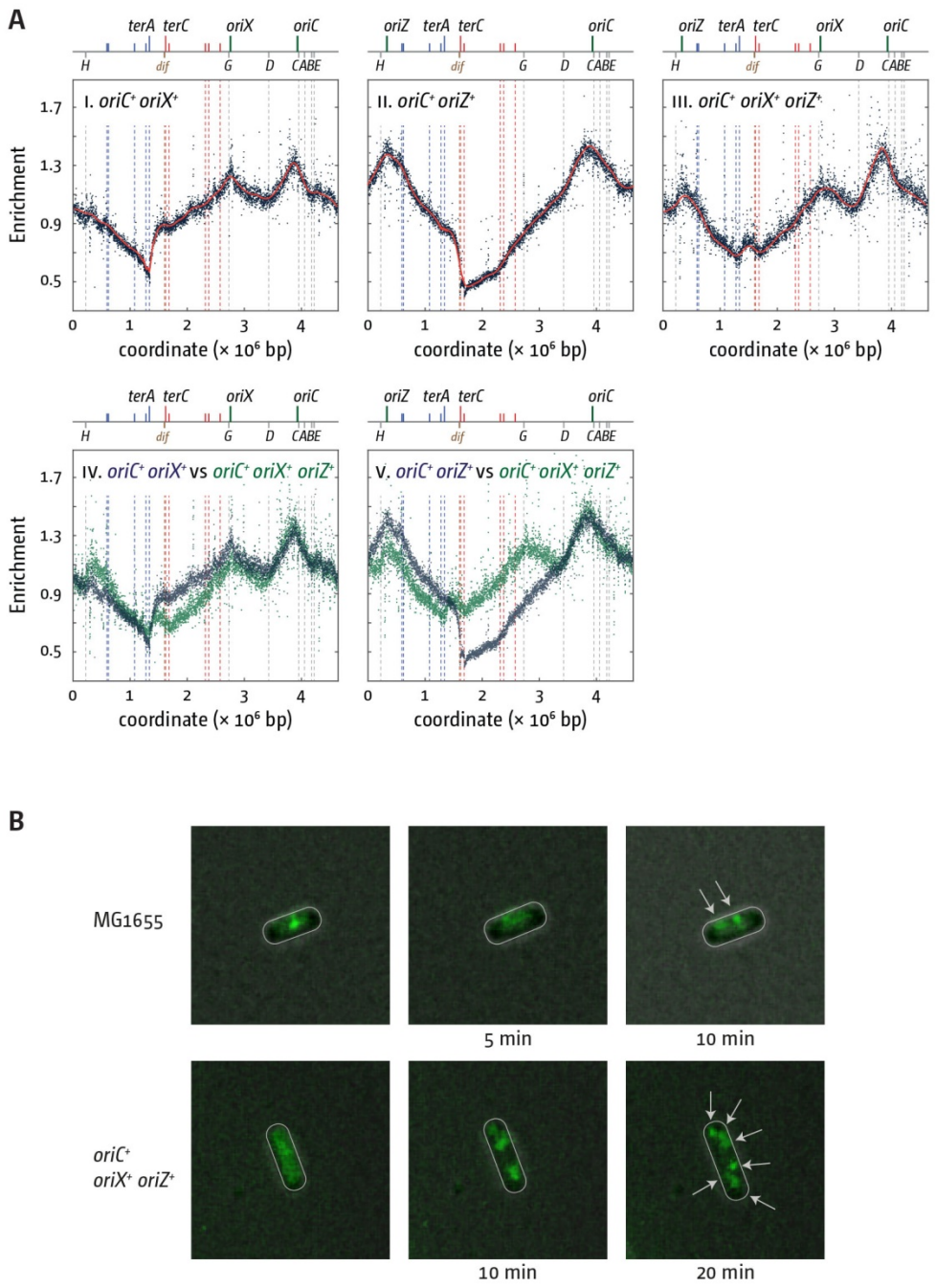


Figure 6. Replication dynamics in *E. coli* cells with one and two ectopic replication origins. **A)** Marker frequency analysis in *E. coli oriC⁺ oriZ⁺*, *oriC⁺ oriX⁺ and oriC⁺ oriX⁺ oriZ⁺* cells. The number of reads (normalised against reads for a stationary phase wild type control) is plotted against the chromosomal location. A schematic representation of the *E. coli* chromosome showing positions of *oriC*, *oriX* and *oriZ* (green lines) and *ter* sites (all above) as well as *dif* and *rrn* operons *A–E*, *G* and *H* (all below) is shown above the plotted data. The strains used were JD1181 (*oriC⁺ oriX⁺*), RCe504 (*oriC⁺ oriZ⁺*) and JD1333 (*oriC⁺ oriX⁺ oriZ⁺*). **B)** Visualisation of replisomes (Ypet-DnaN) in wild type and *oriC⁺ oriX⁺ oriZ⁺* triple origin cells. Cells were grown in M9 minimal salts medium with 0.2% glucose and transferred onto a thin agarose pad of the same medium on a microscopy slide (see Material & Methods). Slides were transferred into a chamber heated to 37°C and fluorescent foci in single cells tracked over time. The strains used were AS1062 (*ypet-dnaN*) and RCe753 (*oriC⁺ oriX⁺ oriZ⁺ ypet-dnaN*).

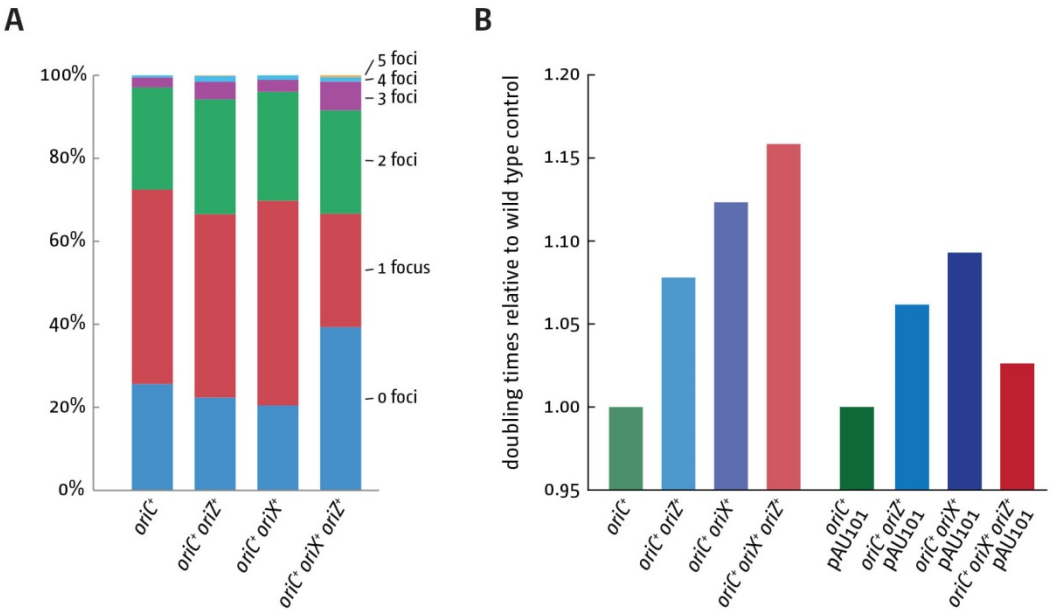


Figure 7. Replisome numbers and doubling times of cells with one and two ectopic replication origins. **A)** Replisome numbers (YPet-DnaN) in wild type, *oriC*⁺ *oriX*⁺, *oriC*⁺ *oriZ*⁺ and *oriC*⁺ *oriX*⁺ *oriZ*⁺ cells. A minimum of 300 cells from at least 3 independent experiments were analysed per strain. Shown are the average focus counts per strain and focus class. The strains used were AS1062 (*ypet-dnaN*), RCe749 (*oriC*⁺ *oriZ*⁺ *ypet-dnaN*), RCe751 (*oriC*⁺ *oriX*⁺ *ypet-dnaN*) and RCe753 (*oriC*⁺ *oriX*⁺ *oriZ*⁺ *ypet-dnaN*). **B)** Doubling times of *E. coli* cells with one or two ectopic replication origins in the presence and absence of an additional copy of the dnaA gene expressed from a low copy number plasmid from its native promoter. All doubling times were determined by measuring viable titres of cultures grown in LB broth (see Material and Methods for details). Changes in doubling times relative to wild type cells are shown due to the fact that the presence of ampicillin necessary for plasmid selection causes a mild change in doubling times (see Suppl. Table 2). The strains used were MG1655, RCe504 (*oriC*⁺ *oriZ*⁺), JD1181 (*oriC*⁺ *oriX*⁺) and JD1333 (*oriC*⁺ *oriX*⁺ *oriZ*⁺) in the presence or absence of plasmid pAU101 (see Supplementary Methods), as indicated.

Finally we wanted to investigate growth of a *ΔoriC oriX*⁺ *oriZ*⁺ construct. A *ΔoriC oriX*⁺ *oriZ*⁺ construct has a symmetrical replichore arrangement, but forks coming both from *oriX* and *oriZ* will still replicate ¼ of the chromosome in an orientation opposite to normal, which would be expected to impose problems. In line with this assumption the deletion of *oriC* increased the doubling time to 35.3 min (Table 3). However, the doubling time of *ΔoriC oriX*⁺ *oriZ*⁺ cells is still significantly quicker than that of *ΔoriC oriX*⁺ cells, suggesting that the presence of *oriZ* alleviates some of the problems that occur in *ΔoriC oriX*⁺ cells.

DISCUSSION

Previously we investigated replication dynamics in cells in which an ectopic origin termed *oriZ* was integrated in the right-hand replichore [18,19]. In this study we attempted to integrate ectopic replication origins at different locations in the left-hand replichore. We hypothesised that replication-transcription conflicts should be less severe, as the left-hand replichore contains less highly transcribed *rrn* operons (Figure 1). We were surprised to find that the attempted integration of *oriY* at 3.55 Mbp into the chromosome, a position where no highly-transcribed *rrn* operons are encountered head-on, only resulted in constructs in

which the *oriC* core sequences were truncated (Figure 2), despite the use of PCR products with the correct length. The truncations differed in all constructs analysed, suggesting that they are spontaneous mutations. If so this indicates strongly that integration of an active origin in this precise location is toxic, while the general integration of sequences such as the antibiotic resistance marker is not. This result rules out that inactivation of *malT* itself is harmful to cells for some reason or that the integration of this fragment somehow activates a cryptic gene that might be toxic for cells. Indeed, it was reported before that integration of an ectopic replication origin resulted in silencing of the native *oriC* [17], supporting the idea that the activity of two origins in close proximity might cause problems for cells.

In contrast, integration of *oriX* into *pheA* was unproblematic and replication profiles as well as fluorescence microscopy analysis confirmed that in *oriC*$^+$ *oriX*$^+$ cells both origins are active and fire with similar frequencies (Figure 3; Suppl. Figure 2), as observed for *oriC*$^+$ *oriZ*$^+$ cells [18,19] (Figure 3).

Termination and replication-transcription conflicts in double-origin strains

The features of the replication profile of *oriC*$^+$ *oriX*$^+$ cells were similar to the replication profiles of *oriC*$^+$ *oriZ*$^+$ cells [8,18]. The innermost *ter* sites *terA* and *terD* stop synthesis coming from *oriX* efficiently, causing a marked asymmetry in the termination area (Figure 3). The impact of *ter*/Tus complexes is highlighted in particular by the ~800 kb inversion found when we attempted to generate a *ΔoriC oriX*$^+$ construct. This inversion flipped all *ter* sites of the left-hand replichore into the permissive orientation for replication coming from *oriX*, thereby effectively inactivating the replication fork trap in this replichore (Figure 5). Thus, the situation in *ΔoriC oriX*$^{+inv}$ cells should be similar to the situation in *ΔoriC oriX*$^+$ *Δtus* cells, and indeed, *ΔoriC oriX*$^{+inv}$ and *ΔoriC oriX*$^+$ *Δtus* cells had similar doubling times (Table 2 and Figure 4). Since no "clean" *ΔoriC oriX*$^+$ construct was generated, we currently do not know whether the inactivation of *tus* acts as a suppressor of the slow growth phenotype of *ΔoriC oriX*$^+$ cells. However, it is likely that the doubling time of *ΔoriC oriX*$^+$ cells is even longer. In this case both the deletion of *tus* and the inversion of all blocking *ter* sites act indeed as a suppressor mutation of the slow growth phenotype of *ΔoriC oriX*$^+$ cells, as observed in *ΔoriC oriZ*$^+$ cells (Table 2 and Figure 4) [18].

Our results show that the growth rate of *ΔoriC oriX*$^+$ *Δtus* cells is considerably slower than the equivalent *ΔoriC oriZ*$^+$ *Δtus* construct (Table 2 and Figure 4), suggesting that replication in *ΔoriC oriX*$^+$ *Δtus* cells has to deal with other serious problems that do not apply in the same way to *ΔoriC oriZ*$^+$ *Δtus* cells. One contributing factor might be head-on replication-transcription encounters, and the doubling time of *ΔoriC oriX*$^+$ *Δtus rpo** cells is indeed reduced in comparison to *ΔoriC oriX*$^+$ *Δtus* cells (Table 2 and Figure 4). Given that an *rpo** point mutation itself slows the doubling time of wild type cells [18], the real effect is likely to be even stronger than the difference immediately obvious from the direct comparison of *ΔoriC oriX*$^+$ *Δtus* cells with and without *rpo**. However, the fact that the

doubling time of *ΔoriC oriX*$^+$ *Δtus rpo** cells is significantly longer than that of *ΔoriC oriZ*$^+$ *Δtus rpo** cells (Table 2 and Figure 4) further supports the idea that additional factors must interfere with successful and efficient chromosome duplication in *ΔoriC oriX*$^+$ cells.

Large chromosomal rearrangements in double-origin cells

A clue as to which additional factors might interfere with DNA replication in *ΔoriC oriX*$^+$ cells might come from a spontaneous rearrangement observed in one of our *ΔoriC oriX*$^+$ *Δtus* cultures, duplicating the chromosomal stretch containing *rrn* operons *A* and *B* (Figure 5B ii). The location of important genetic elements relative to the origins and the resulting gene dosage effect was described before [37]. The *rrn* operons *CABE* and *D* are all located in close proximity to the replication origin, ensuring an increased copy number in fast growing cells (Figure 3A) [37]. In contrast, shifting the origin from its original to the *oriX* location results in a much reduced copy number especially of the *rrn* operons *CABE* and *H* (Figure 5B i). This effect is specific to *oriX* due to its distance to all *rrn* operons with the exception of *rrnG*. The location of *oriZ* is in close proximity to *rrn* operons *H* and the *rrnCABE* cluster (Figure 1), providing a potential explanation why *ΔoriC oriZ*$^+$ cells struggle less. In addition, the inversion found in *ΔoriC oriZ*$^+$ cells not only re-aligns replication and transcription, but also brings the *rrnCABE* cluster in close proximity of *oriZ* [18], explaining perhaps why this particular inversion is such an efficient suppressor of the slow growth phenotype despite a persisting replication asymmetry. It is tempting to speculate that *ΔoriC oriZ*$^+$ *Δtus* cells containing the duplication of *rrnA* and *rrnB* will be able to grow faster. However, as this duplication was spontaneously acquired in a culture for genomic DNA extraction and only revealed after sequencing, it was not possible to measure whether it conferred a growth advantage. Indeed, other effects might contribute. It was shown before that deletion of *rrn* operons affects growth rate of cells only moderately [38,39]. However, in a recent study a duplication of a similar location was observed as a suppressor of the severe growth defect of cells lacking the DnaA regulatory inactivator Hda [40]. The suppression of the slow growth phenotype of *Δhda* cells was found to be the increased gene dosage for DNA polymerase I [40]. This or other similar effects might be important contributors in *oriX* cells.

The large number of GCRs observed as part of our studies fits well with previous reports of a surprising number of rearrangements in a limited set of *E. coli* samples, including a duplication of the *rrn* operons *A*, *B* and *E* [41], highlighting a surprising degree of plasticity of the *E. coli* chromosome. Rearrangements and especially duplications are among the most frequent mutational events [41,42]. But unless they confer an immediate advantage they will be rapidly lost because of a fitness cost [41]. Given the slow growth of *ΔoriC oriZ*$^+$ cells and the robust suppression by the inversion, the isolation of the GCR observed is not much of a surprise, as it will out-grow the original construct very rapidly. We assume that a similar argument can be made for the GCR observed in our *ΔoriC oriX*$^+$ construct (Figure 5). Perhaps the biggest surprise is the inversion observed in *ΔoriC oriX*$^+$ *rpo** cells. An 895 kb section of

the chromosome spanning from IS5 at 1394 kb to IS5 at 2288 kb was inverted (Figure 5B iv & iv a; see Suppl. Figure 3 for PCR verification of the inversion). This inversion not only brings the *ter* sites *C* and *B* in close proximity of *oriX*, but also switches them to the restrictive orientation, forcing the replication fork coming from *oriX* travelling in the normal orientation to stop after 650 kb. The remaining 4000 kb of the chromosome have to be replicated by the clockwise replication fork. If this inversion acts as a suppressor mutation then it must alleviate a yet unidentified replication stress, but the replication profile gives little clue as to what this stress might be. However, the doubling time of the *ΔoriC oriX*$^{+}$ *rpo** construct carrying the inversion is quicker than the doubling time of *ΔoriC oriX*^{inv} and *ΔoriC oriX tus* (Table 2), suggesting that the *rpo** mutation does indeed improve growth despite the effect of the highly asymmetric replichore arrangement.

It is noteworthy that two of the inversions found in this study have specifically arisen at IS5 elements, which provide large stretches (~1.2 kb) of homology. These IS elements allow for relatively frequent large chromosomal rearrangements to occur that clearly can efficiently alleviate problems to replication and other cellular processes. Indeed, it was shown that the systematic deletion of all IS elements caused a robust genetic stabilisation, with a 75% decrease of the mutation rate determined in this particular study [43], demonstrating their contribution towards the observed plasticity of the genome.

Replication in cells with three functional replication origins

The replication profiles of our triple origin construct provides further evidence of how finely balanced the replication parameters of the *E. coli* chromosome are. While our fluorescence microscopy studies show that all three origins can be active in some of the cells (Figure 6B), the replication profile revealed that the peak height of both ectopic origins was significantly reduced (Figure 6). This suggests that all three origins being simultaneously active is probably a rare event. It is likely that in a fraction of cells only two of the three origins might be active, one of which almost always is the native *oriC*. In *oriC*$^{+}$ *oriZ*$^{+}$ and *oriC*$^{+}$ *oriX*$^{+}$ cells both the ectopic and the native origin fire with similar frequency (Figure 3) [18,19], suggesting that both are equivalent. Apparently this changes in a triple origin background, even though the reason for this effect is not known. The reduction of the doubling time of triple origin cells in which an additional copy of *dnaA* is introduced via a low copy number plasmid (Figure 7B) suggests that three copies of the origin per cell generate an environment where, at least in some cells, the threshold level of DnaA necessary for efficient origin initiation is not reached for some time. This causes a delay of initiation of all origins in a fraction of cells, which explains the increased level of cells in which no replisomes are observed (Figure 7A). Thus, our data are in line with the idea that a delay of origin firing contributes to the slow doubling time of triple origin cells.

Nevertheless, if all origins were equivalent there should be an equal reduction of peak heights of all three origins, which was not observed. The *oriC* peak is significantly higher,

demonstrating that the *oriC* sequence in its native location has the highest capacity for being active. Indeed, in bacteria chromosomes with a single origin are the norm [3], despite the fact that the resulting long replichores require replication machineries with very high speed and accuracy in comparison to DNA synthesis in eukaryotic cells. It was suggested that the genes flanking the origin sequence might influence origin activity [44], explaining why cells carrying a 5 kb *oriC* region stretch as developed in the Sherratt lab [19] is active, whereas smaller fragments are not [44]. It is possible that an even larger fragment of the chromosome is required for full functionality, which might explain the reduced activity of both *oriX* and *oriZ* in our *oriC*$^+$ *oriX*$^+$ *oriZ*$^+$ construct (Figure 6). However, the toxicity of the 5 kb origin fragment integrated into the *malT* gene strongly argues that this assumption is too simple, as there appear to be strong effects relating to the relative position of multiple origins to each other, the precise location of an origin within the cell or the combination of multiple effects.

We were intrigued to find what looks like a peak of over-replication within the termination area. Similar peaks were reported in cells lacking RecG helicase [23], RNase HI [45,46], and other proteins [8,47,48]. We have postulated that the fusion of two replisomes in the termination area results in intermediates which require processing by proteins such as RecG helicase and 3' exonucleases [8,23,49–52], the absence of which results in substantial amounts of over-replication in the termination area. However, all the above proteins are fully functional in our triple origin construct, making it unlikely that the peak is a similar type of over-replication. In fact, the peak can be fully explained if replication is initiated at two of the three origins in a significant fraction of cells. In *oriC*$^+$ *oriZ*$^+$ cells marker frequency is high throughout the termination area, with a marked decrease at *terC/B* (Figure 3A iii). In *oriC*$^+$ *oriX*$^+$ the opposite is the case. Marker frequency is again high throughout the termination area, with a marked decrease at *terA/D* (Figure 3A ii). If in triple origin cells a significant fraction of cells only uses two origins, as the replication profile of triple origin cells suggest, then the replication profile of triple origin cells should be formed by the superposition of the two profiles of *oriC*$^+$ *oriX*$^+$ and *oriC*$^+$ *oriZ*$^+$ cells (Figure 8A).

In both the marker frequency is high in the middle of the termination area, while the areas around *terC/B* and *terA/D* should be reduced because of the marked decrease in one fraction of cells (Figure 8A). We exploited mathematical modelling of whole genome replication [53] (see Supplementary Methods) to predict the replication profile within a population of cells where either *oriC* and *oriX* or *oriC* and *oriZ* are active. In our modelling we assume a constant fork speed once forks are established. The periodicity of origin firing is estimated from our experimental data. For simplicity, *ter*/Tus complexes were treated as a hard stop to replication. While the resulting modelled replication profile lacks the complexity of our data sets (Figure 8B), it fits overall well with the population-based replication profile and shows a clear peak in the termination area, as predicted. This supports the idea that this peak is indeed caused by the presence of defined fractions within the overall population, rather than actual over-replication of the termination area.

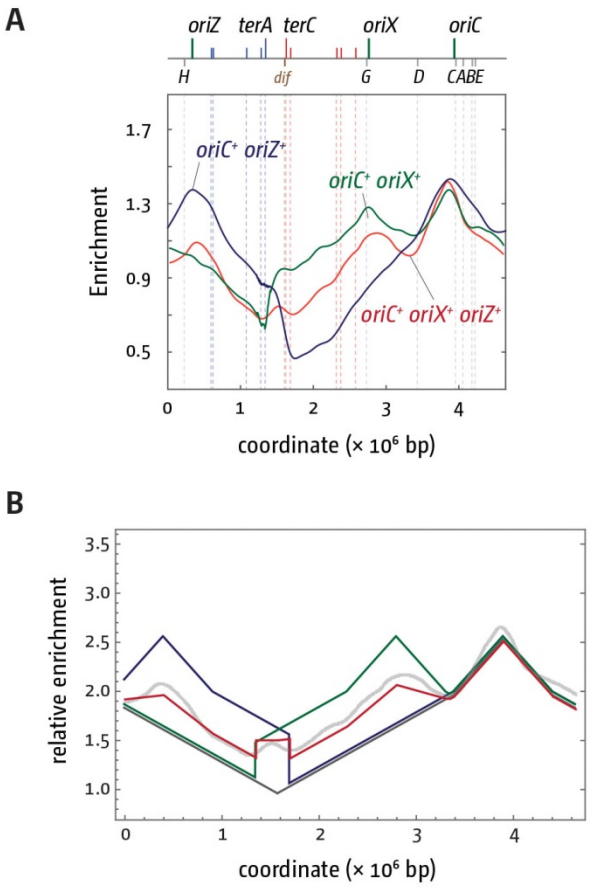


Figure 8. Comparative analysis of replication profiles of *E. coli* cells with two and three replication origins. **A)** Shown is a combination of the LOESS regression profiles for *oriC*⁺ *oriZ*⁺ (blue), *oriC*⁺ *oriX*⁺ (green) and *oriC*⁺ *oriX*⁺ *oriZ*⁺ (red) cells, as shown in Figure 6. **B)** Mathematical model fitting for *oriC*⁺ *oriZ*⁺ (blue), *oriC*⁺ *oriX*⁺ (green) and *oriC*⁺ *oriX*⁺ *oriZ*⁺ (red) cells. The modelling for *oriC*⁺ *oriX*⁺ *oriZ*⁺ (red) assumes that *oriC* and *oriX* are active in 50% of cells, *oriC* and *oriZ* are active in 40% of cells, based on the peak heights in the replication profiles. The LOESS regression curve for *oriC*⁺ *oriX*⁺ *oriZ* ⁺cells is shown in light grey, as shown in Figure 6. See text for further details.

As the replication profiles of cells lacking RecG helicase or 3' exonucleases have been generated from a similar population-based approach [23,45,48], it could be suggested that the peaks observed might be resulting from a similar superposition of different populations. Indeed, it was recently shown that the sharp loss of sequences corresponding to the terminus area in the replication profile of a *recB* mutant strain stems only from a defined fraction of cells [54]. However, the presence of synthesis in the termination area was confirmed using different experimental approaches [52] and we were able to demonstrate that cells lacking RecG helicase can tolerate the inactivation of *oriC* as long as the termination area is inactivated by deletion of *tus* and replication-transcription encounters are alleviated by the presence of an *rpo** point mutation [23,45]. Thus, there is no doubt that extra synthesis is indeed initiated in the termination area of cells lacking RecG. However, use of the rapidly emerging single-cell approaches [55] will enable an even more refined approach to these aspects of replication and chromosome dynamics.

ACCESSION NUMBERS

All relevant raw sequencing data can be accessed at the European Nucleotide Archive (http://www.ebi.ac.uk/ena/data/view/PRJEB19883)

SUPPLEMENTARY MATERIAL

Supplementary Material is available online.

FUNDING

This work was supported by Research Grants BB/K015729/1 and BB/N014995/1 from the Biotechnology and Biological Sciences Research Council to CJR.

Conflict of interest: *none*

ACKNOWLEDGEMENTS

The authors wish to thank Sarah Midgley-Smith and Anastasia Georgievskaya for critical reading of the manuscript.

REFERENCES

1. Leonard, A. C.; Méchali, M. DNA Replication Origins. *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, a010116, doi:10.1101/cshperspect.a010116.
2. Mott, M. L.; Berger, J. M. DNA replication initiation: mechanisms and regulation in bacteria. *Nat. Rev. Microbiol.* **2007**, *5*, 343–354, doi:10.1038/nrmicro1640.
3. Gao, F. Bacteria may have multiple replication origins. *Front. Microbiol.* **2015**, *6*, doi:10.3389/fmicb.2015.00324.
4. Gao, F.; Zhang, C.-T. Ori-Finder: A web-based system for finding oriC s in unannotated bacterial genomes. *BMC Bioinformatics* **2008**, *9*, 79, doi:10.1186/1471-2105-9-79.
5. Jameson, K. H.; Wilkinson, A. J. Control of Initiation of DNA Replication in Bacillus subtilis and Escherichia coli. *Genes* **2017**, *8*, 22, doi:10.3390/genes8010022.
6. Messer, W. The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. *FEMS Microbiol. Rev.* **2002**, *26*, 355–374.
7. Skarstad, K.; Katayama, T. Regulating DNA replication in bacteria. *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, a012922, doi:10.1101/cshperspect.a012922.
8. Dimude, J. U.; Midgley-Smith, S. L.; Stein, M.; Rudolph, C. J. Replication Termination: Containing Fork Fusion-Mediated Pathologies in Escherichia coli. *Genes* **2016**, *7*, doi:10.3390/genes7080040.
9. Reyes-Lamothe, R.; Wang, X.; Sherratt, D. Escherichia coli and its chromosome. *Trends Microbiol.* **2008**, *16*, 238–245, doi:10.1016/j.tim.2008.02.003.
10. Duggin, I. G.; Wake, R. G.; Bell, S. D.; Hill, T. M. The replication fork trap and termination of chromosome replication. *Mol. Microbiol.* **2008**, *70*, 1323–1333, doi:10.1111/j.1365-2958.2008.06500.x.
11. Neylon, C.; Kralicek, A. V.; Hill, T. M.; Dixon, N. E. Replication termination in Escherichia coli: structure and antihelicase activity of the Tus-Ter complex. *Microbiol. Mol. Biol. Rev. MMBR* **2005**, *69*, 501–526, doi:10.1128/MMBR.69.3.501-526.2005.
12. Reyes-Lamothe, R.; Nicolas, E.; Sherratt, D. J. Chromosome replication and segregation in bacteria. *Annu. Rev. Genet.* **2012**, *46*, 121–143, doi:10.1146/annurev-genet-110711-155421.
13. Srivatsan, A.; Tehranchi, A.; MacAlpine, D. M.; Wang, J. D. Co-Orientation of Replication and Transcription Preserves Genome Integrity. *PLoS Genet* **2010**, *6*, e1000810, doi:10.1371/journal.pgen.1000810.

14. Wang, J. D.; Berkmen, M. B.; Grossman, A. D. Genome-wide coorientation of replication and transcription reduces adverse effects on replication in Bacillus subtilis. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 5608–5613, doi:10.1073/pnas.0608999104.
15. McGlynn, P.; Savery, N. J.; Dillingham, M. S. The conflict between DNA replication and transcription. *Mol. Microbiol.* **2012**, *85*, 12–20, doi:10.1111/j.1365-2958.2012.08102.x.
16. Merrikh, H.; Zhang, Y.; Grossman, A. D.; Wang, J. D. Replication-transcription conflicts in bacteria. *Nat. Rev. Microbiol.* **2012**, *10*, 449–458, doi:10.1038/nrmicro2800.
17. Kouzminova, E. A.; Kuzminov, A. Patterns of chromosomal fragmentation due to uracil-DNA incorporation reveal a novel mechanism of replication-dependent double-stranded breaks. *Mol. Microbiol.* **2008**, *68*, 202–215, doi:10.1111/j.1365-2958.2008.06149.x.
18. Ivanova, D.; Taylor, T.; Smith, S. L.; Dimude, J. U.; Upton, A. L.; Mehrjouy, M. M.; Skovgaard, O.; Sherratt, D. J.; Retkute, R.; Rudolph, C. J. Shaping the landscape of the Escherichia coli chromosome: replication-transcription encounters in cells with an ectopic replication origin. *Nucleic Acids Res.* **2015**, *43*, 7865–7877, doi:10.1093/nar/gkv704.
19. Wang, X.; Lesterlin, C.; Reyes-Lamothe, R.; Ball, G.; Sherratt, D. J. Replication and segregation of an Escherichia coli chromosome with two replication origins. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, E243-250, doi:10.1073/pnas.1100874108.
20. Thomason, L. C.; Costantino, N.; Court, D. L. E. coli genome manipulation by P1 transduction. *Curr. Protoc. Mol. Biol. Ed. Frederick M Ausubel Al* **2007**, *Chapter 1*, Unit 1.17, doi:10.1002/0471142727.mb0117s79.
21. Datsenko, K. A.; Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 6640–6645, doi:10.1073/pnas.120163297.
22. Bachmann, B J Derivations and Genotypes of Some Mutant Derivatives of Escherichia coli K-12. In *Escherichia coli and Salmonella Cellular and Molecular Biology*; ASM Press, 1996.
23. Rudolph, C. J.; Upton, A. L.; Stockum, A.; Nieduszynski, C. A.; Lloyd, R. G. Avoiding chromosome pathology when replication forks collide. *Nature* **2013**, *500*, 608–611, doi:10.1038/nature12312.
24. Meddows, T. R.; Savory, A. P.; Lloyd, R. G. RecG helicase promotes DNA double-strand break repair. *Mol. Microbiol.* **2004**, *52*, 119–132, doi:10.1111/j.1365-2958.2003.03970.x.
25. Guy, C. P.; Atkinson, J.; Gupta, M. K.; Mahdi, A. A.; Gwynn, E. J.; Rudolph, C. J.; Moon, P. B.; van Knippenberg, I. C.; Cadman, C. J.; Dillingham, M. S.; Lloyd, R. G.; McGlynn, P. Rep provides a second motor at the replisome to promote duplication of protein-bound DNA. *Mol. Cell* **2009**, *36*, 654–666, doi:10.1016/j.molcel.2009.11.009.
26. Bernhardt, T. G.; de Boer, P. A. J. The Escherichia coli amidase AmiC is a periplasmic septal ring component exported via the twin-arginine transport pathway. *Mol. Microbiol.* **2003**, *48*, 1171–1182.
27. Luria, S. E.; Burrous, J. W. Hybridization between Escherichia coli and Shigella. *J. Bacteriol.* **1957**, *74*, 461–476.
28. Cleveland WS Robust Locally Weighted Regression and Smoothing Scatterplots. *J. Am. Stat. Assoc.* **1979**, *74*, 829–836, doi:10.1080/01621459.1979.10481038.
29. Rudolph, C. J.; Dhillon, P.; Moore, T.; Lloyd, R. G. Avoiding and resolving conflicts between DNA replication and transcription. *DNA Repair* **2007**, *6*, 981–993, doi:10.1016/j.dnarep.2007.02.017.
30. Reyes-Lamothe, R.; Sherratt, D. J.; Leake, M. C. Stoichiometry and architecture of active DNA replication machinery in Escherichia coli. *Science* **2010**, *328*, 498–501, doi:10.1126/science.1185757.
31. Moolman, M. C.; Krishnan, S. T.; Kerssemakers, J. W. J.; van den Berg, A.; Tulinski, P.; Depken, M.; Reyes-Lamothe, R.; Sherratt, D. J.; Dekker, N. H. Slow unloading leads to DNA-bound β2-sliding clamp accumulation in live Escherichia coli cells. *Nat. Commun.* **2014**, *5*, 5820, doi:10.1038/ncomms6820.
32. Pham, T. M.; Tan, K. W.; Sakumura, Y.; Okumura, K.; Maki, H.; Akiyama, M. T. A single-molecule approach to DNA replication in Escherichia coli cells demonstrated that DNA polymerase III is a major determinant of fork speed. *Mol. Microbiol.* **2013**, *90*, 584–596, doi:10.1111/mmi.12386.
33. Roth, A.; Messer, W. High-affinity binding sites for the initiator protein DnaA on the chromosome of Escherichia coli. *Mol. Microbiol.* **1998**, *28*, 395–401.
34. Boye, E.; Løbner-Olesen, A.; Skarstad, K. Limiting DNA replication to once and only once. *EMBO Rep.* **2000**, *1*, 479–483, doi:10.1093/embo-reports/kvd116.
35. Kitagawa, R.; Ozaki, T.; Moriya, S.; Ogawa, T. Negative control of replication initiation by a novel chromosomal locus exhibiting exceptional affinity for Escherichia coli DnaA protein. *Genes Dev.* **1998**, *12*, 3032–3043.
36. Løbner-Olesen, A.; Skarstad, K.; Hansen, F. G.; von Meyenburg, K.; Boye, E. The DnaA protein determines the initiation mass of Escherichia coli K-12. *Cell* **1989**, *57*, 881–889.

37. Jin, D. J.; Cagliero, C.; Zhou, Y. N. Growth rate regulation in Escherichia coli. *FEMS Microbiol. Rev.* **2012**, *36*, 269–287, doi:10.1111/j.1574-6976.2011.00279.x.

38. Asai, T.; Condon, C.; Voulgaris, J.; Zaporojets, D.; Shen, B.; Al-Omar, M.; Squires, C.; Squires, C. L. Construction and initial characterization of Escherichia coli strains with few or no intact chromosomal rRNA operons. *J. Bacteriol.* **1999**, *181*, 3803–3809.

39. Quan, S.; Skovgaard, O.; McLaughlin, R. E.; Buurman, E. T.; Squires, C. L. Markerless Escherichia coli rrn Deletion Strains for Genetic Determination of Ribosomal Binding Sites. *G3 Genes Genomes Genet.* **2015**, *5*, 2555–2557, doi:10.1534/g3.115.022301.

40. Babu, V. M. P.; Itsko, M.; Baxter, J. C.; Schaaper, R. M.; Sutton, M. D. Insufficient levels of the nrdAB-encoded ribonucleotide reductase underlie the severe growth defect of the Δhda E. coli strain. *Mol. Microbiol.* **2017**, *104*, 377–399, doi:10.1111/mmi.13632.

41. Skovgaard, O.; Bak, M.; Løbner-Olesen, A.; Tommerup, N. Genome-wide detection of chromosomal rearrangements, indels, and mutations in circular chromosomes by short read sequencing. *Genome Res.* **2011**, *21*, 1388–1393, doi:10.1101/gr.117416.110.

42. Reams, A. B.; Roth, J. R. Mechanisms of gene duplication and amplification. *Cold Spring Harb. Perspect. Biol.* **2015**, *7*, a016592, doi:10.1101/cshperspect.a016592.

43. Umenhoffer, K.; Draskovits, G.; Nyerges, Á.; Karcagi, I.; Bogos, B.; Tímár, E.; Csörgő, B.; Herczeg, R.; Nagy, I.; Fehér, T.; Pál, C.; Pósfai, G. Genome-Wide Abolishment of Mobile Genetic Elements Using Genome Shuffling and CRISPR/Cas-Assisted MAGE Allows the Efficient Stabilization of a Bacterial Chassis. *ACS Synth. Biol.* **2017**, doi:10.1021/acssynbio.6b00378.

44. Milbredt, S.; Farmani, N.; Sobetzko, P.; Waldminghaus, T. DNA Replication in Engineered Escherichia coli Genomes with Extra Replication Origins. *ACS Synth. Biol.* **2016**, *5*, 1167–1176, doi:10.1021/acssynbio.6b00064.

45. Dimude, J. U.; Stockum, A.; Midgley-Smith, S. L.; Upton, A. L.; Foster, H. A.; Khan, A.; Saunders, N. J.; Retkute, R.; Rudolph, C. J. The Consequences of Replicating in the Wrong Orientation: Bacterial Chromosome Duplication without an Active Replication Origin. *mBio* **2015**, *6*, doi:10.1128/mBio.01294-15.

46. Maduike, N. Z.; Tehranchi, A. K.; Wang, J. D.; Kreuzer, K. N. Replication of the Escherichia coli chromosome in RNase HI-deficient cells: multiple initiation regions and fork dynamics. *Mol. Microbiol.* **2014**, *91*, 39–56, doi:10.1111/mmi.12440.

47. Markovitz, A. A new in vivo termination function for DNA polymerase I of Escherichia coli K12. *Mol. Microbiol.* **2005**, *55*, 1867–1882, doi:10.1111/j.1365-2958.2005.04513.x.

48. Wendel, B. M.; Courcelle, C. T.; Courcelle, J. Completion of DNA replication in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 16454–16459, doi:10.1073/pnas.1415025111.

49. Lloyd, R. G.; Rudolph, C. J. 25 years on and no end in sight: a perspective on the role of RecG protein. *Curr. Genet.* **2016**, *62*, 827–840, doi:10.1007/s00294-016-0589-z.

50. Rudolph, C. J.; Upton, A. L.; Briggs, G. S.; Lloyd, R. G. Is RecG a general guardian of the bacterial genome? *DNA Repair* **2010**, *9*, 210–223, doi:10.1016/j.dnarep.2009.12.014.

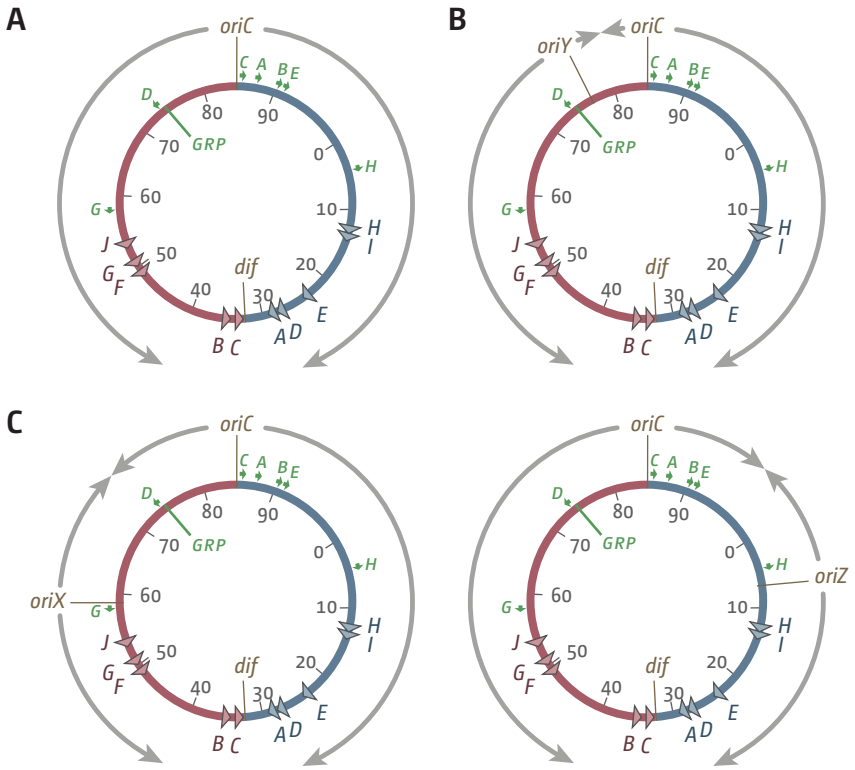
51. Rudolph, C. J.; Mahdi, A. A.; Upton, A. L.; Lloyd, R. G. RecG protein and single-strand DNA exonucleases avoid cell lethality associated with PriA helicase activity in Escherichia coli. *Genetics* **2010**, *186*, 473–492, doi:10.1534/genetics.110.120691.

52. Rudolph, C. J.; Upton, A. L.; Lloyd, R. G. Replication fork collisions cause pathological chromosomal amplification in cells lacking RecG DNA translocase. *Mol. Microbiol.* **2009**, *74*, 940–955, doi:10.1111/j.1365-2958.2009.06909.x.

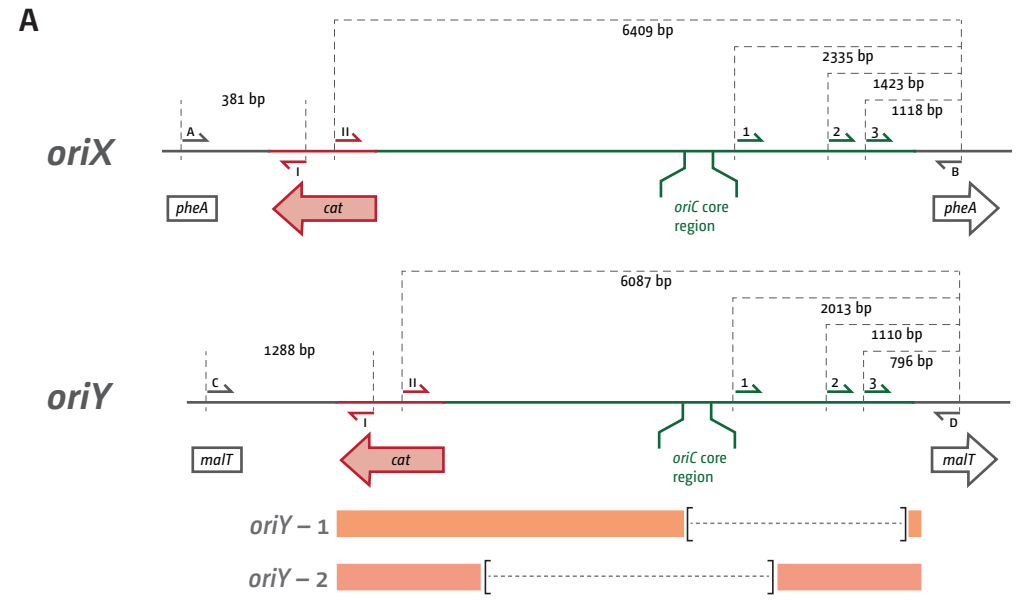
53. Retkute, R.; Nieduszynski, C. A.; de Moura, A. Mathematical modeling of genome replication. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **2012**, *86*, 031916.

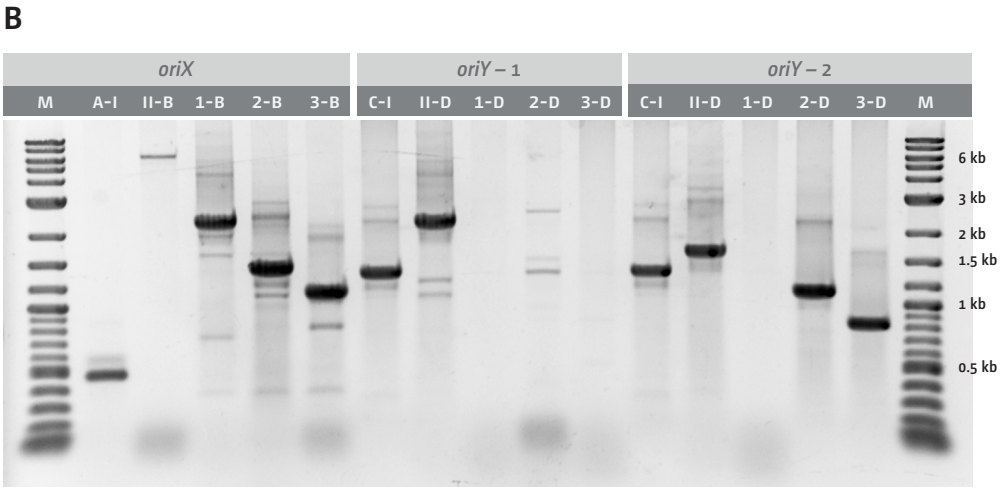
54. Sinha, A. K.; Durand, A.; Desfontaines, J.-M.; Iurchenko, I.; Auger, H.; Leach, D. R. F.; Barre, F.-X.; Michel, B. Division-induced DNA double strand breaks in the chromosome terminus region of Escherichia coli lacking RecBCD DNA repair enzyme. *PLoS Genet.* **2017**, *13*, e1006895, doi:10.1371/journal.pgen.1006895.

55. Macaulay, I. C.; Voet, T. Single Cell Genomics: Advances and Future Perspectives. *PLOS Genet.* **2014**, *10*, e1004126, doi:10.1371/journal.pgen.1004126.

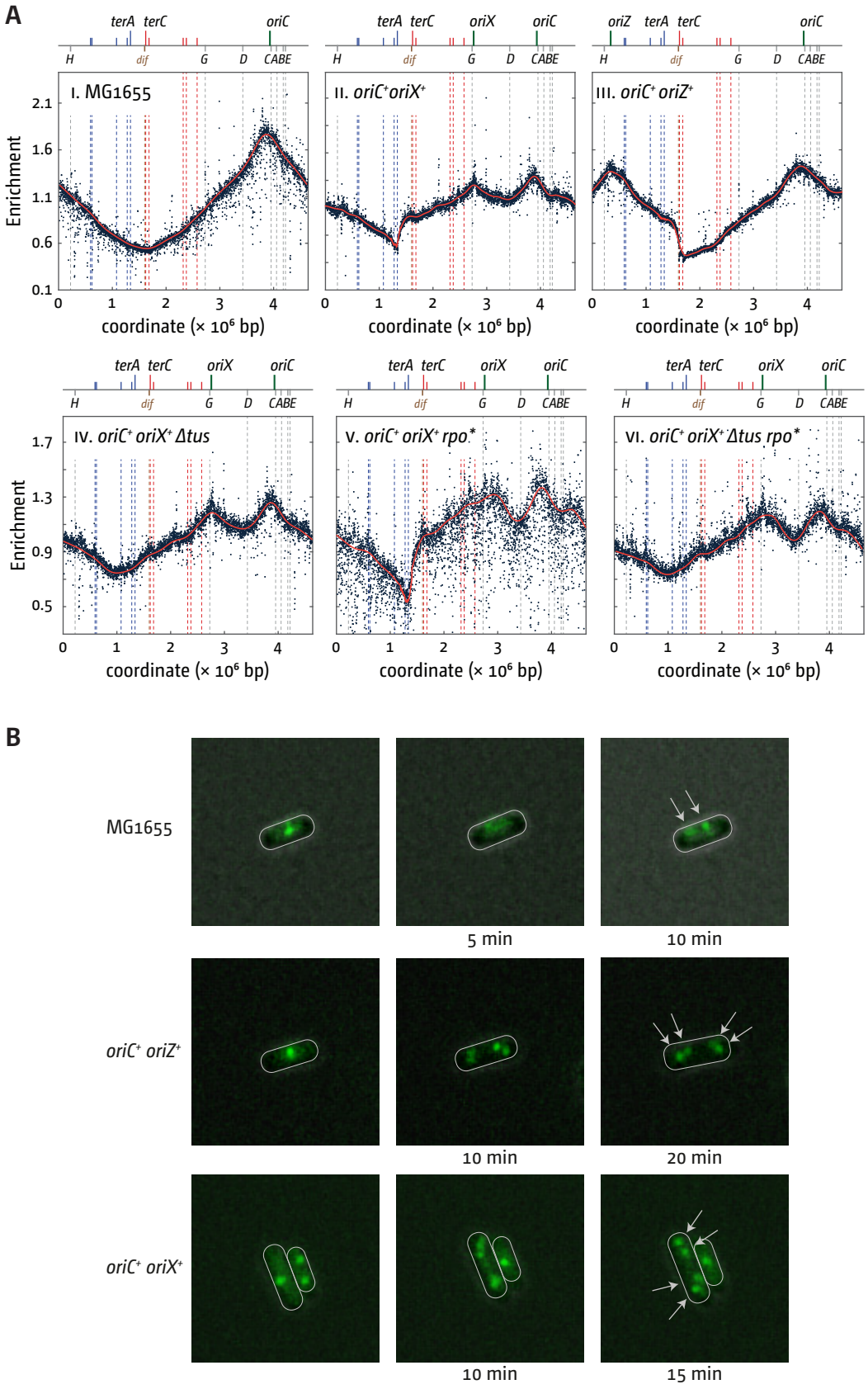


Dimude *et al.* Figure 1

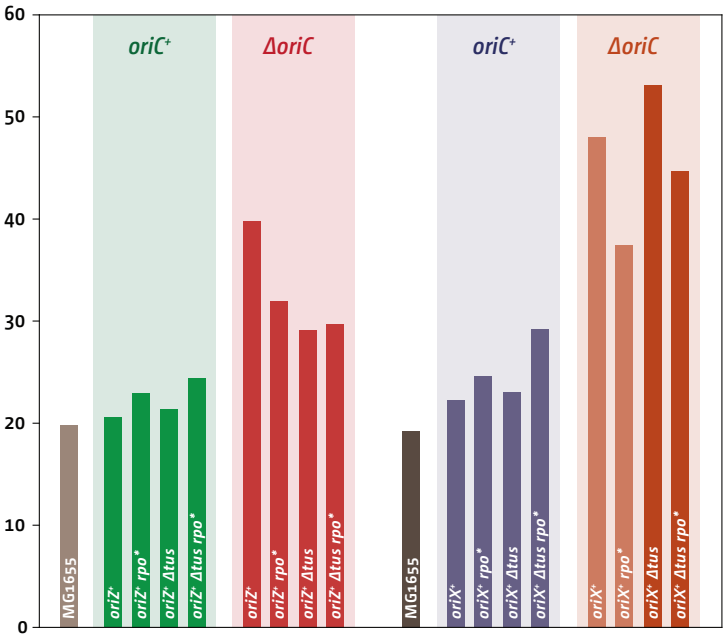




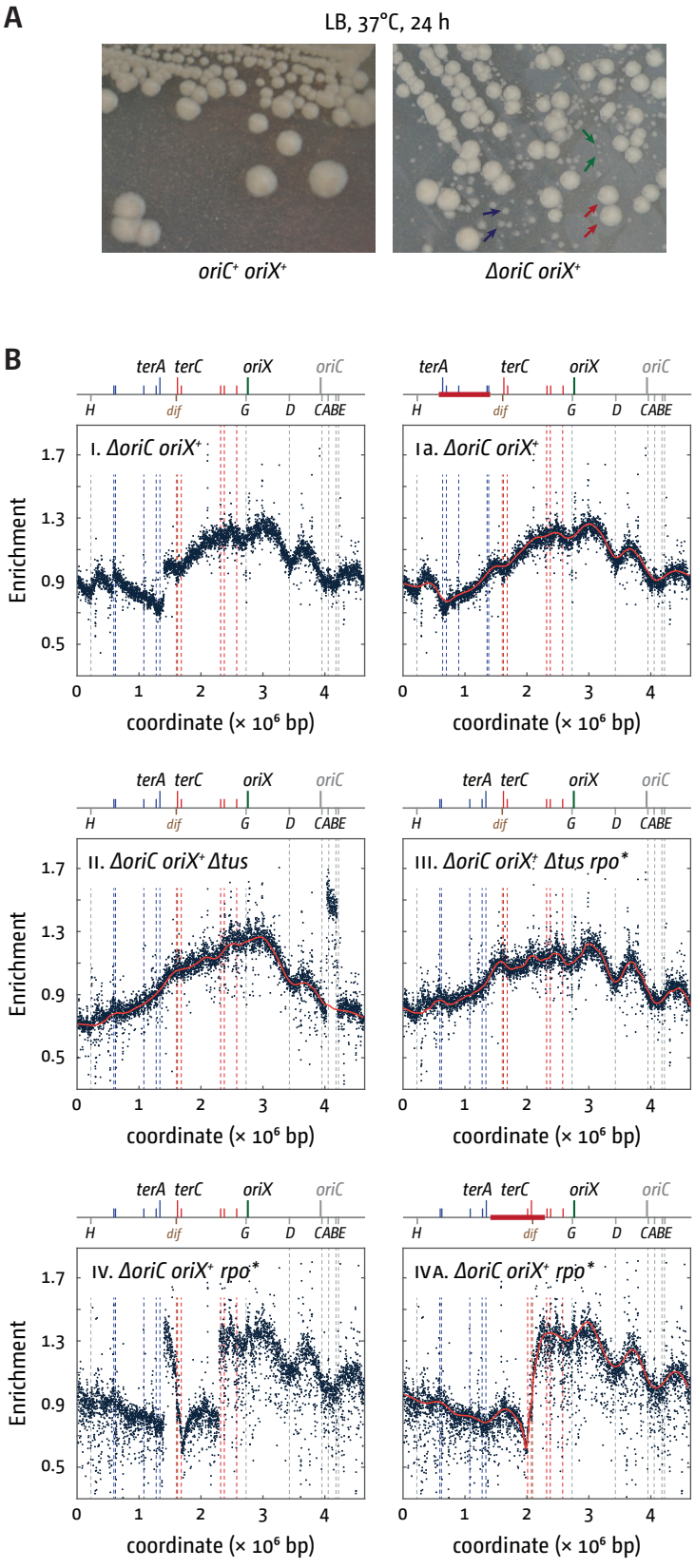
Dimude *et al.* Figure 2



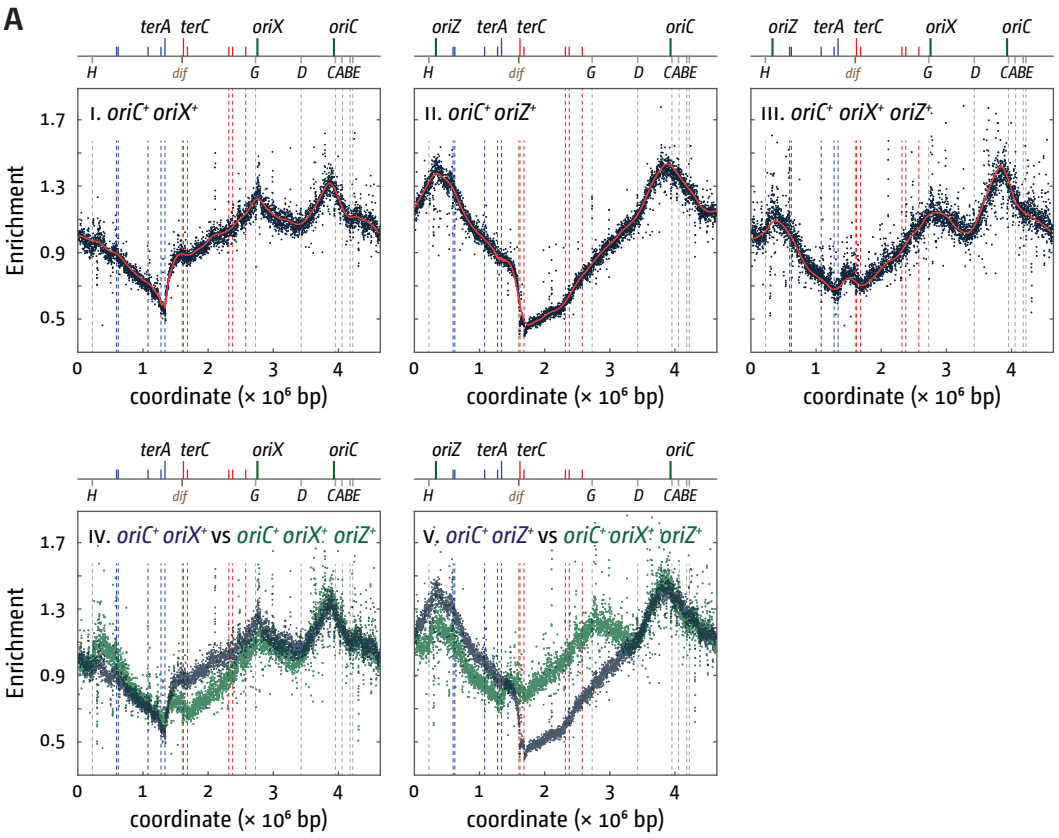
Dimude *et al.* Figure 3

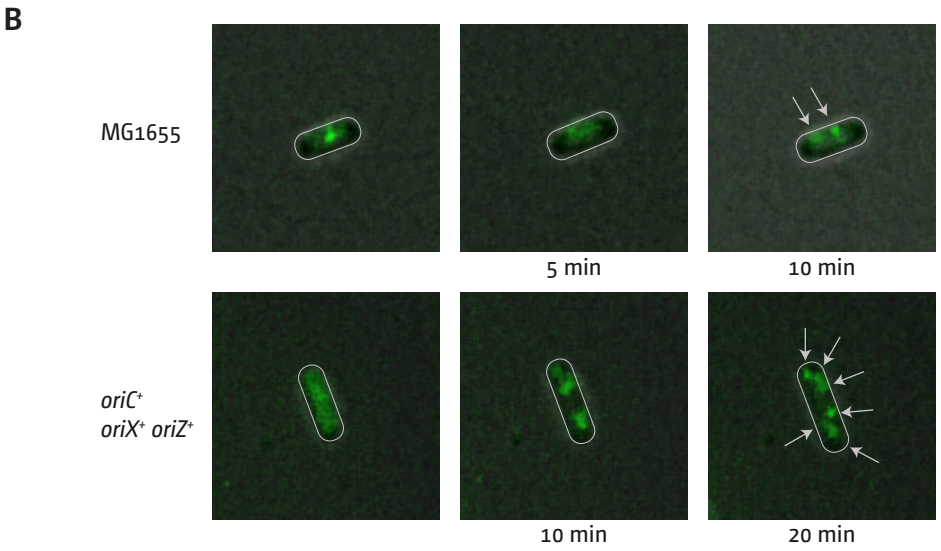


Dimude *et al.* Figure 4

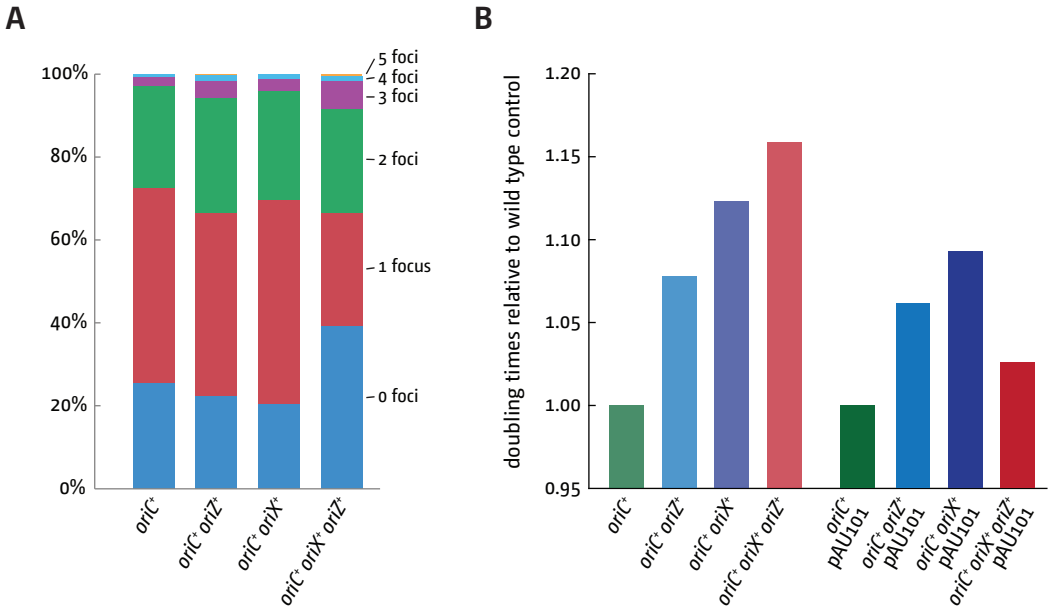


Dimude *et al.* Figure 5

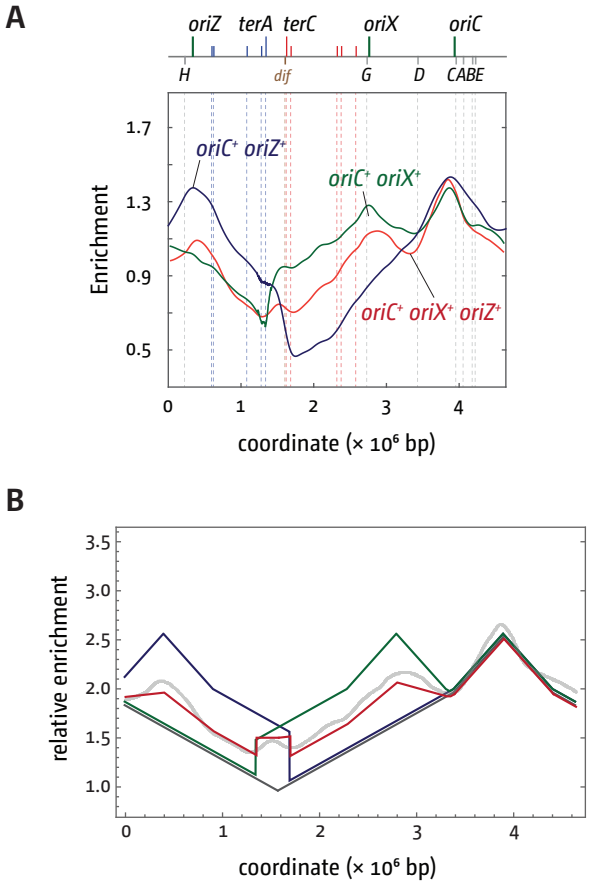




Dimude *et al.* Figure 6



Dimude *et al.* Figure 7



Dimude *et al.* Figure 8

Supplementary Information

Origins left, right and centre: increasing the number of initiation sites in the *Escherichia coli* chromosome

Juachi U. Dimude[1], Monja Stein[1], Ewa E. Andrzejewska[1], Mohammad S. Khalifa[1], Alexandra A. Gajdosova[1], Renata Retkute[2], Ole Skovgaard[3] and Christian J. Rudolph[1,*]

*Corresponding author: christian.rudolph@brunel.ac.uk

[1]Division of Biosciences, College of Health and Life Sciences,
Brunel University London, Uxbridge, UB8 3PH, UK

[2]School of Life Science, The University of Warwick,
Gibbet Hill Campus, Coventry, CV4 7AL, UK

[3]Department of Science, Systems and Models, Roskilde University,
DK-4000 Roskilde, Denmark

SUPPLEMENTARY MATERIAL AND METHODS

Plasmids

Plasmid pAU101 is a derivative of pRC7 [1] carrying the coding sequence for *dnaA*$^+$ including its native promoter. The *dnaA* region was PCR amplified from MG1655 using 5' and 3' primers incorporating *Apa*I sites. The PCR product was cloned into the *Apa*I site within *lacI*^{q} to give pAU101. The coding sequence inserted is transcribed in the same orientation as the disrupted *lacI*^{q} gene. pAU101 fully complements the temperature sensitivity of the *dnaA46* temperature sensitive strains.

Marker frequency analysis by deep sequencing

Samples from cultures of a strain grown over night in LB broth were diluted 100-fold in fresh broth and incubated with vigorous aeration until an A_{600} reached 0.48 at 37°C. The only exceptions were all *ΔoriC oriX* backgrounds, for which growth was initiated from a single colony from a streak plate to avoid suppressors formed in the overnight culture outgrowing the slow growing *ΔoriC oriX* derivatives. All cultures were then diluted a second time 100-fold in pre-warmed fresh broth and grown again until an A_{600} of 0.48 was reached. Samples from these exponential phase cultures were flash-frozen in liquid nitrogen at this point for subsequent DNA extraction. Growth curves were recorded using the same procedure (see below), demonstrating that cultures grown to an A_{600} of 0.48 did not show any sign of transition into stationary phase. For wild type incubation of the remaining culture was continued until several hours after the culture had saturated and showed no further increase in the A_{600}. A further sample (stationary phase) was frozen at this point. For all samples shown in the main Figures of this work DNA was then extracted using the GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich), using a 30 min proteinase K digest at 55°C, as indicated in the manufacturer protocol (see below for limitations of this procedure). Marker frequency analysis was performed using Illumina HiSeq 2500 sequencing (fast run) to measure sequence copy number. FastQC was used for a basic metric of quality control in the raw data. Bowtie2 was used to align the sequence reads to the reference. Samtools was used to calculate the enrichment of uniquely mapped sequence tags in 1 kb windows for an exponentially growing (replicating) sample relative to a non-replicating stationary phase wild type sample to correct for differences in read depth across the genome and to allow presentation of the data as a marker frequency, as described previously [2–4].

For presentation of the data as a marker frequency replication profile the raw read counts for each construct was divided by the average of all read counts across the entire genome to correct for the somewhat different absolute numbers of aligned reads in the various samples. The normalised read count values for each exponentially growing sample were then divided by the corresponding normalised read count value from a stationary (non-replicating) sample. This division "cleans" the raw data significantly, because data points which are

outliers caused by technical aspects (precise sequence environment interfering with library preparation or similar issues) will be similarly distorted both in the exponential and the stationary samples. However, while true in principle, we have observed that there can be variations specifically in these noisy data points even within a single batch of samples processed in parallel. If the absolute sequence reads of the genome fragments causing the noisy data points in a sample are underrepresented in comparison to the same fragments in the stationary phase sample, then the division process described above causes all of these data points to skew below the position of the neighbouring data points. In contrast, if the absolute sequence reads of the fragments are higher than the sequence reads in the stationary control, then the same division process causes all of these data points to skew above the position of the neighbouring data points. An example of this effect can be seen in Figures 3A and 3B. While the sample in panel iii shows no skew, indicating that noise both in the exponential sample and the stationary sample are of a similar level, the sample in panel v shows a clear skew of all noisy data points below the level of neighbouring data points, while the sample in panel vi shows a skew above the level of neighbouring data points. We do not currently know what is causing these variations even though we have run extensive tests to try to identify their cause. From our tests we suspect that a combination of factors including quality of genomic DNA preparation and library generation contributes to this effect. Whatever the reason, these problems affect mostly the noise and do not obscure the general trend of the bulk of the data points.

We have by now identified another effect that is specifically caused by the quality of the genomic DNA. The genomic DNA extraction via the GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich) requires a 30 min proteolytic digest with proteinase K. This digestion step is not sufficient to fully remove all proteins in the sample. As a consequence, some partially digested or undigested proteins remain bound to DNA fragments. As part of the following column purification procedure, these proteins including the bound DNA are removed. This causes areas of the chromosome in which proteins are tightly bound (*ter*/Tus complexes are one example) or which are very frequently bound by proteins (highly transcribed areas such as the *rrn* operons) to be under-represented in the genomic DNA preparation, leading to small dips in the profile. Examples can be seen in Figure 5. In Figure 5B panel i/ia clear dips of the profile can be seen at all *rrn* operons and at some of the *ter* sites. As shown in Suppl. Figure 1, these dips are much reduced if the proteolytic digest of the samples is extended to 2 h.

Mathematical modelling

We used the DNA replication modelling described in Retkute *et al.* [5]. Our modelling has the following assumptions: (i) the length of the chromosome is normalised by half of its length with *oriC* positioned at x = 0 (i.e. the length unit is the distance between *oriC* and *ter*); (ii) the replication time unit is defined as time required for full replication of the half of the

chromosome (C period of the bacterial cell cycle); (iii) fork velocity is constant and equal to 1 time unit per length unit; (vi) the age of the genome is defined from one fork termination to the next; (v) the time at which new initiation events occur is *s* (the periodicity of initiation) and it is defined with respect to the previous replication initiation event; (vi) all origins activate at the same time and with the same periodicity *s*. Supplementary Figure 5A shows a spatiotemporal representation of the replication program for a hypothetical chromosome with two replication origins positioned at x = 0 and x = 0.5. Each new round of replication starts while the previous replication round is still ongoing, so there are four copies of newly replicated genetic material. Given the age distribution of genomes [6] (shown in Supplementary Figure 5B), the mean number of copies is calculated as an integral over all ages of age distribution multiplied by the number of copies at a particular position and a particular age (shown in Supplementary Figure 5C). Then different compositions (percentages of genomes with one, two or three active origins) were set as parameters. Supplementary Figure 5D shows an illustrative example with 25% genomes firing one origin and 75% firing both origins. Parameters were fitted by minimising a mean squared error (MSE) between model predicted values, F_i, and experimental data, d_i:

$$MSE = \sqrt{\frac{\sum_{i=1}^{n}(d_i - a\,F_i)^2}{n}}$$

with a scaling factor *a* fitted as one of the parameters, along with periodicity of initiation and percentage of genomes.

In the case of asynchronious initiation (shown in Supplementary Figure 5E), there would be differences in comparison to the synchronous initiation with a fraction of cells firing one origin. A comparison of profiles for synchronous initiation (blue curve) and asynchronous initiation (dashed magenta curve) is shown in Supplementary Figure 5F.

SUPPLEMENTARY REFERENCES

1. Bernhardt, T. G.; de Boer, P. A. J. The Escherichia coli amidase AmiC is a periplasmic septal ring component exported via the twin-arginine transport pathway. *Mol. Microbiol.* **2003**, *48*, 1171–1182.
2. Ivanova, D.; Taylor, T.; Smith, S. L.; Dimude, J. U.; Upton, A. L.; Mehrjouy, M. M.; Skovgaard, O.; Sherratt, D. J.; Retkute, R.; Rudolph, C. J. Shaping the landscape of the Escherichia coli chromosome: replication-transcription encounters in cells with an ectopic replication origin. *Nucleic Acids Res.* **2015**, *43*, 7865–7877, doi:10.1093/nar/gkv704.
3. Müller, C. A.; Hawkins, M.; Retkute, R.; Malla, S.; Wilson, R.; Blythe, M. J.; Nakato, R.; Komata, M.; Shirahige, K.; de Moura, A. P. S.; Nieduszynski, C. A. The dynamics of genome replication using deep sequencing. *Nucleic Acids Res.* **2014**, *42*, e3, doi:10.1093/nar/gkt878.
4. Skovgaard, O.; Bak, M.; Løbner-Olesen, A.; Tommerup, N. Genome-wide detection of chromosomal rearrangements, indels, and mutations in circular chromosomes by short read sequencing. *Genome Res.* **2011**, *21*, 1388–1393, doi:10.1101/gr.117416.110.
5. Retkute, R.; Nieduszynski, C. A.; de Moura, A. Mathematical modeling of genome replication. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **2012**, *86*, 031916.
6. Sueoka, N.; Yoshikawa, H. The chromosome of Bacillus subtilis. I. Theory of marker frequency analysis. *Genetics* **1965**, *52*, 747–757.

SUPPLEMENTARY TABLES

Supplementary Table 1: Replication profile minima established by LOESS regression of the replication profiles of *E. coli* strains with one and two replication origins

Strain background	Location of terminus-proximal LOESS minima [Mbp]	Location of *oriC*–*oriX* LOESS minima [Mbp]	Arithmetic mid points [Mbp]
MG1655	1.627	n/a	1.603
oriC⁺ *oriX*⁺	1.322	3.3925	1.010; 3.330
oriC⁺ *oriX*⁺ *Δtus*	0.991	3.348	1.010; 3.330
oriC⁺ *oriX*⁺ *rpoB*35*	1.3175	3.373	1.010; 3.330
oriC⁺ *oriX*⁺ *Δtus rpoB*35*	0.967	3.360	1.010; 3.330
ΔoriC oriX⁺ *Δtus*	0.292	n/a	0.4159
ΔoriC oriX⁺ *rpoB*35*	1.9675	n/a	0.4159
ΔoriC oriX⁺ *Δtus rpoB*35*	0.2658	n/a	0.4159
ΔoriC oriX⁺	0.658	n/a	0.4159

Supplementary Table 2: Effect of increased *dnaA* gene dosage on the doubling times in cells with one and two ectopic replication origins

Strain background	Doubling time [min]	SD	r²
MG1655[a]	19.6	± 1.0	0.999
oriC⁺ *oriX*⁺ [a]	21.0	± 0.8	0.997
oriC⁺ *oriZ*⁺ [a]	21.8	± 0.8	0.996
oriC⁺ *oriX*⁺ *oriZ*⁺ [a]	22.7	± 2.5	0.994
MG1655 pAU101	27.1	± 2.4	0.985
oriC⁺ *oriX*⁺ pAU101	28.5	± 2.9	0.995
oriC⁺ *oriZ*⁺ pAU101	29.7	± 1.7	0.992
oriC⁺ *oriX*⁺ *oriZ*⁺ pAU101	27.7	± 1.9	0.995

a – data for constructs without *dnaA* plasmid pAU101 as in Table 3, for comparison. For details of pAU101 see Supplementary Material and Methods.

SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure 1. Marker frequency analysis and sample quality of *E. coli ΔoriC oriX*⁺ *Δtus rpo** cells following short (30 min) and extended (120 min) de-proteinisation via proteolytic digest using proteinase K. The numbers of reads (normalised against reads for a stationary phase wild type control) are plotted against the chromosomal location. A schematic representation of the *E. coli* chromosome showing positions of *oriC* and *oriX* and *ter* sites (above) as well as *dif* and *rrn* operons *A*–*E*, *G* and *H* (below) is shown above the plotted data. The strain used was JD1209 (*ΔoriC oriX*⁺ *Δtus rpo**).

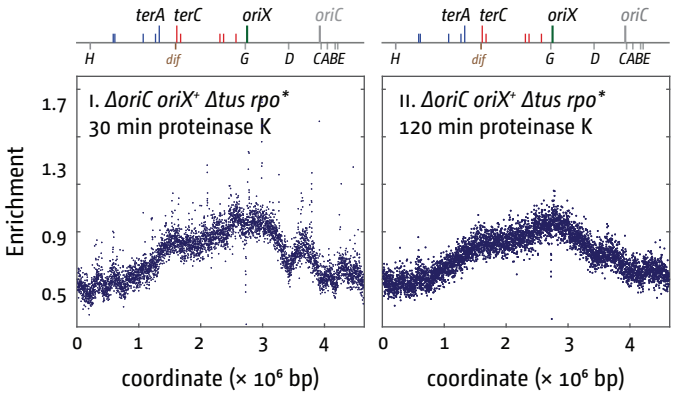
Supplementary Figure 2. Marker frequency analysis of *E. coli oriC*$^{+}$ *oriX*$^{+}$ cells following phenol-chloroform extraction of genomic DNA. The numbers of reads (normalised against reads for a stationary phase wild type control) are plotted against the chromosomal location. A schematic representation of the *E. coli* chromosome showing positions of *oriC* and *oriX* (green line) and *ter* sites (above) as well as *dif* and *rrn* operons *A–E*, *G* and *H* (below) is shown above the plotted data.

Supplementary Figure 3. PCR verification of chromosomal inversions. **A)** Schematic representation of primer binding sites, inversion locations and the relocation of primer binding sites following specific inversion events. The schematic showing the inversion between IS5 elements at location 575 kb and 1394 kb is shaded in red, the schematic showing the inversion between IS5 elements at 1394 kb and 2288 kb is shaded blue. The wild type situation is shaded in yellow. Primers have a single letter identifier, which is shown in bold if the binding site is relocated due to an inversion event to highlight their changed position. Location of primer binding sites are not to scale. All expected PCR products are between 3 and 6.5 kb in length. **B)** Agarose gel electrophoresis of PCRs with primer combinations probing for the wild type sequence and chromosomal DNA templates for a wild type control (yellow), the *ΔoriC oriX* background carrying the inversion at IS5 elements at 575 kb and 1394 kb (red) as well as the *ΔoriC oriX rpo** background that carries an inversion at IS5 elements at 1394 kb and 2288 kb. Primer combinations as shown in A are given above each lane. The size of the PCR product for a specific primer combination is indicated by a grey arrow. The + or – indicates whether a PCR product is expected with the template used. Primer combination a & b did not give a PCR product in any PCR attempted. However, PCR products for both primers a and b are obtained if paired with different secondary primers, suggesting that it is the specific combination of a & b that fails to produce a PCR product. An inverted gel image is shown for clarity. **C)** Agarose gel electrophoresis of PCRs with primer combinations probing for both inversions and chromosomal DNA templates for a wild type control (yellow), the *ΔoriC oriX* background carrying the inversion at IS5 elements at 575 kb and 1394 kb (red) as well as the *ΔoriC oriX rpo** background that carries an inversion at IS5 elements at 1394 kb and 2288 kb. Primer combinations as shown in A are given above each lane, with a + or – indicating whether a PCR product is expected. An inverted gel image is shown for clarity. All primers that span flanks following both inversion events show a PCR product, confirming both inversion events identified in our replication profiles.

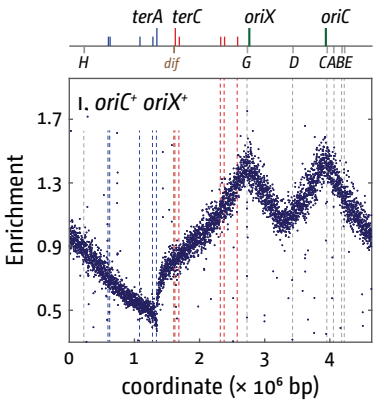
Supplementary Figure 4. Replication profiles of *E. coli* cells with synthesis starting at ectopic replication origins only. **A–B)** Marker frequency analysis of *E. coli ΔoriC oriX*$^{+}$ derivatives. The numbers of reads are normalised against reads for a non-growing stationary phase wild type control and then plotted against the chromosomal location. In this particular run the noise observed comes from an increased overall level of noise of the entire sequencing run. This is made worse by the fact that the stationary wild type control was particularly

affected by the noise, which introduces this noise into all other samples due to the normalisation. A schematic representation of the *E. coli* chromosome showing positions of *oriC* and *oriX* (green line) and *ter* sites (above) as well as *dif* and *rrn* operons *A–E*, *G* and *H* is shown above the plotted data. Inverted regions are highlighted by a red box. Replication profiles in A are obtained from independent experiments, with independently generated chromosomal DNA, library generation and sequencing runs. Replication profiles in B are reproduced from Figure 5 for comparison. The direct comparison of the *ΔoriC oriX*$^{+}$ *Δtus* replication profile from the first and second run shows a duplication of the *rrnA–B* region present only in the second run, even though cultures for the preparation of genomic DNA were prepared from the same frozen stock (highlighted in red in B and in grey in A).

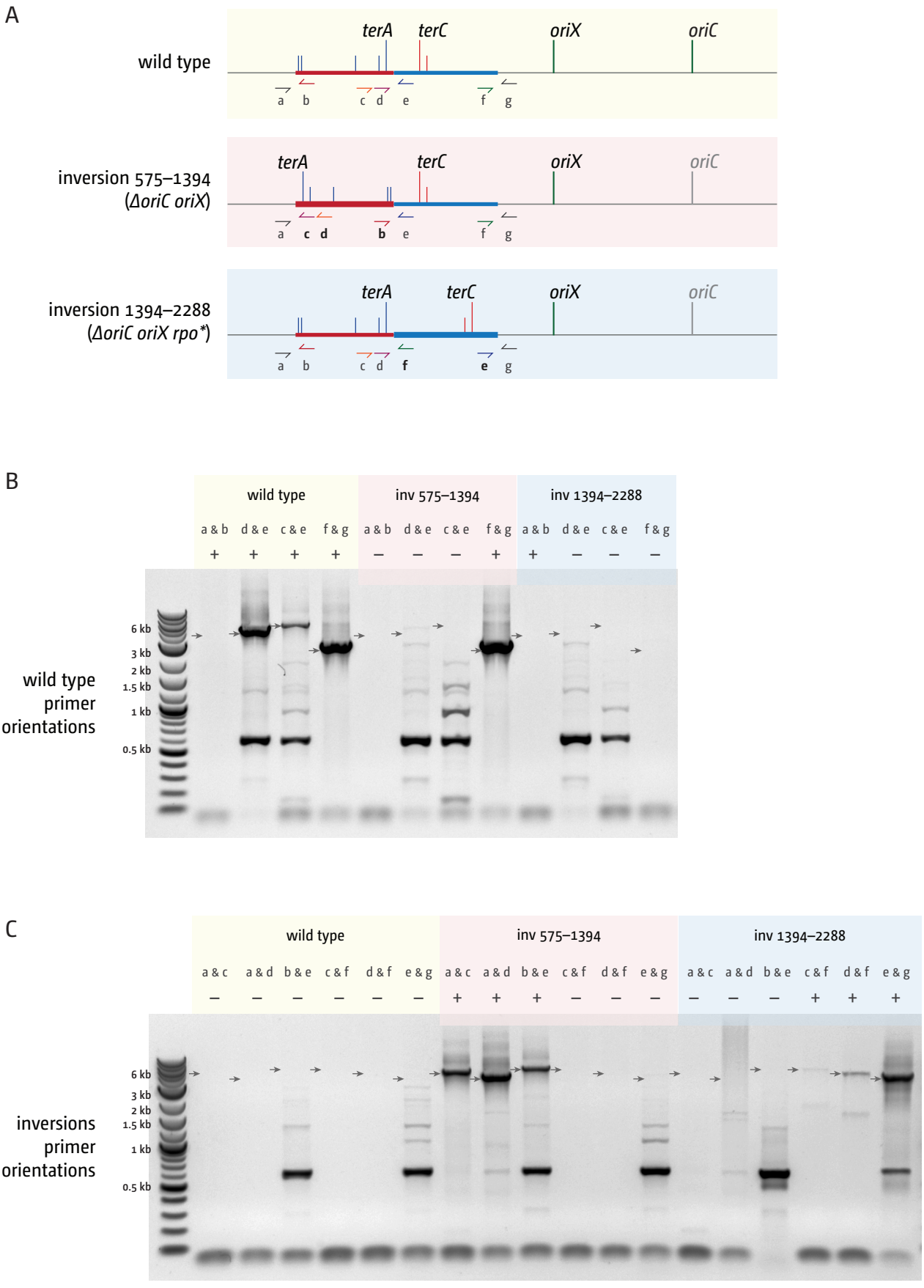
Supplementary Figure 5. Mathematical modelling of chromosomal replication in *E. coli* with one or multiple origins. **A)** Spatiotemporal representation of a replication program for two origins positioned at $x = 0$ and $x = 0.5$. The tops of each inverted red triangle indicate the initiation of replication. Number of genome copies are 1 (white), 2 (yellow) or 4 (red). The difference between two initiation events establishes the periodicity *s*. **B)** Age distribution. **C)** Mean number of copies. **D)** Inferring population composition: overall profile (blue) is a result of 25% of genomes with only origin at $x = 0$ active (red) and 75% of genomes having both origins active. **E)** Spatiotemporal representation of the replication program for two asynchronously initiating origins. **F)** Mean number of copies for synchronous initiation with 25% of cells firing one origin and 75% firing two origins (blue), and asynchronous initiation with 100% of cells firing two origins but at different times (magenta). **G)** Overlay of model predictions for synchronous (blue) *versus* asynchronous (magenta) initiations and LOESS data of the replication profile of an *oriC*$^{+}$ *oriX* strain. Asynchronous initiation predicts a shift of the termination point to the left, while a shift to the right is observed in our experimental data.



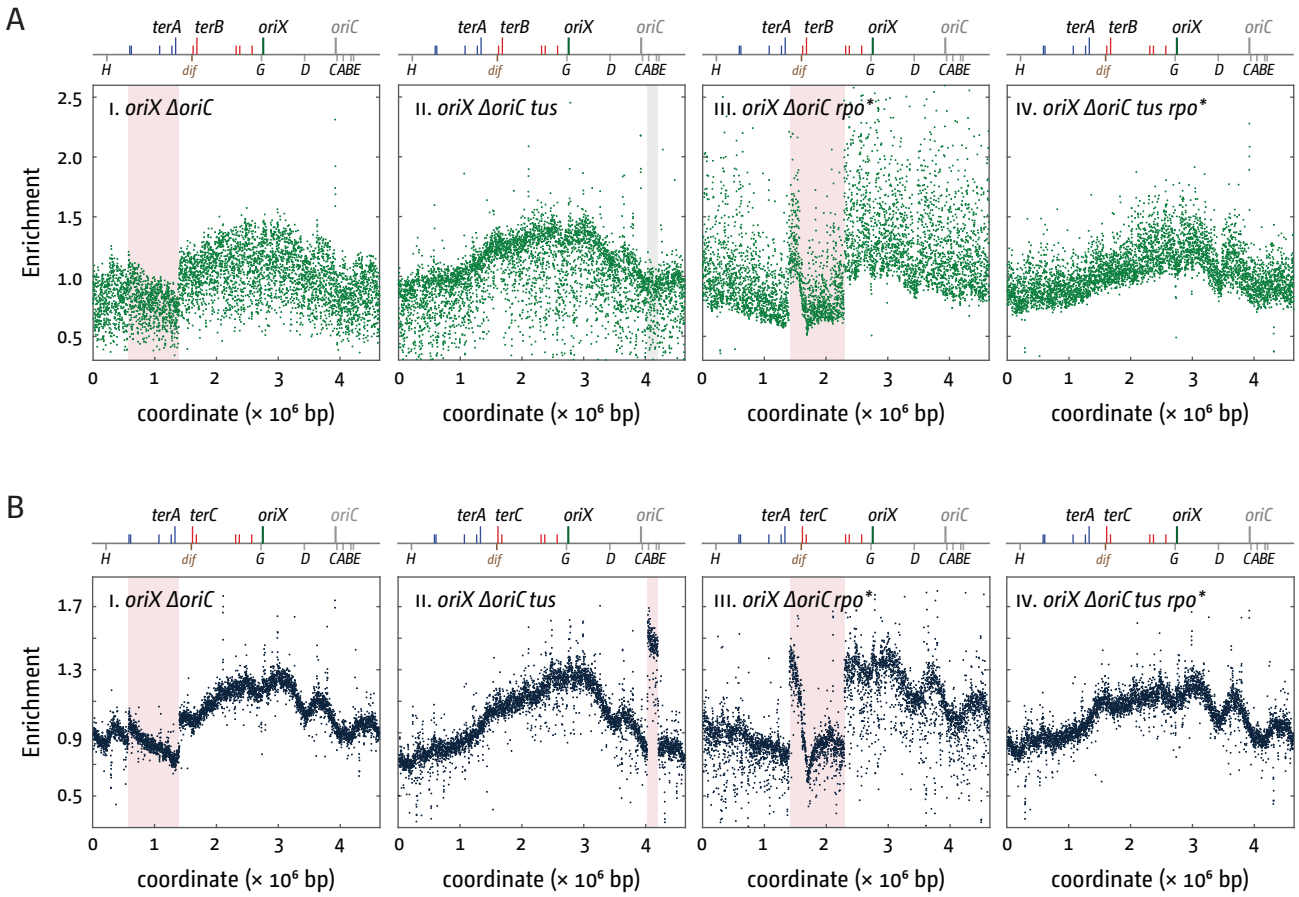
Dimude *et al.* Figure S1



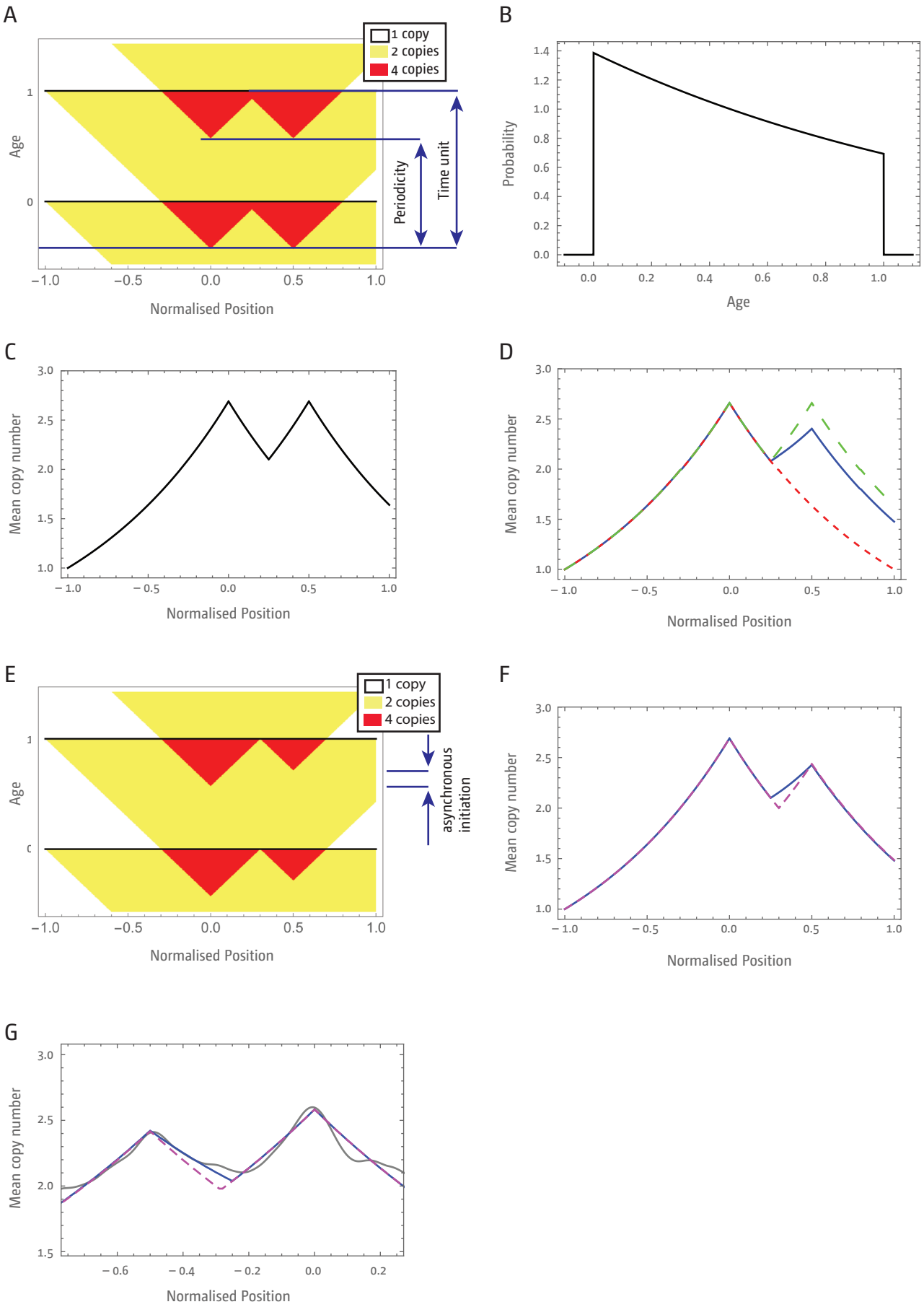
Dimude *et al.* Figure S2



Dimude *et al.* Figure S3



Dimude *et al.* Figure S4



Dimude *et al.* Figure S5