Speed fluctuations of bacterial replisomes

Deepak Bhat[a], Samuel Hauf[b], Charles Plessy[c], Yohei Yokobayashi[b], and Simone Pigolotti[a,1]

[a]Biological Complexity Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa 904-0495, Japan
[b]Nucleic Acid Chemistry and Engineering Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa 904-0495, Japan
[c]Genomics and Regulatory Systems Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa 904-0495, Japan
[1]Corresponding author: simone.pigolotti@oist.jp

October 15, 2021

Abstract

Replisomes are multi-protein complexes that replicate genomes with remarkable speed and accuracy. Despite their importance, the dynamics of replisomes along the genome is poorly characterised, especially in vivo. In this paper, we link the replisome dynamics with the DNA abundance distribution measured in an exponentially growing bacterial population. Our approach permits to accurately infer the replisome dynamics along the genome from deep sequencing measurements. As an application, we experimentally measured the DNA abundance distribution in *Escherichia coli* populations growing at different temperatures. We find that the average replisome speed increases nearly five-fold between 17^{o}C and 37^{o}C. Further, we observe wave-like variations of the replisome speed along the genome. These variations are correlated with previously observed variations of the mutation rate along the genome. We interpret this correlation as a speed–error trade-off and discuss its possible dynamical origin. Our approach has the potential to elucidate replication dynamics in *E. coli* mutants and in other bacterial species.

Every cell must copy its genome in order to reproduce. This task is carried out by large protein complexes called replisomes. Each replisome separates the two DNA strands and synthesizes a complementary copy of each of them, thereby forming a Y–shaped DNA junction called a replication fork. The speed and accuracy of replisomes is impressive [1]. They proceed at several hundreds to one thousand base pairs per second [2, 3], with an inaccuracy of about one mis-incorporated monomer every 10 billion base pairs [4]. In bacteria, two replisomes initiate replication at a well-defined origin site

on the circular genome, progress in opposite directions and complete replication upon encountering each other in a terminal region.

The initiation and the completion of DNA replication conventionally delimit the three stages of the bacterial cell cycle [5, 6]. The first stage, B, spans the period from cell birth until initiation of DNA replication. The second stage, C, encompasses the time needed for replication. The last phase, D, begins at the end of DNA replication and concludes with cell division. While it is established that DNA replication and the cell cycle must be coordinated, their precise relation has been a puzzle for decades [7]. A classic study by Cooper and Helmstetter [8] finds that, upon modifying the growth rate by changing the nutrient composition in *Escherichia coli*, the durations of stages C and D remain constant at about 40 min and 20 min, respectively. This means that the replisome speed must be unaffected by the nutrient composition, at least on average. When the cell division time is shorter than one hour, DNA replication is initiated in a previous generation. This implies that, in fast growth conditions, multiple pairs of replisomes simultaneously replicate the same genome [9]. Tuning the growth rate by changing the temperature has a radically different effect on bacterial physiology. For example, in vivo [10] and in vitro [11] studies show that the speed of replisomes is affected in this case.

Detailed information about replisome dynamics, such as whether their speed varies along the genome, is harder to obtain [2]. One approach for gaining this information is to measure the DNA abundance distribution, i.e. the frequency of DNA fragments along the genome in an exponentially growing cell population. In fact, the frequency of these fragments in the population depends on the proportions of synthesizing genomes of different lengths, which in turn depend on the replisome dynamics. Previous studies have measured the DNA abundance distribution, but focused on qualitative analysis of the observed variations in knockout mutants [12, 13, 14, 15, 16].

In this paper, we formulate a theory that permits to infer the dynamics of replisomes from the DNA abundance distribution. We apply our theory to elucidate replisome dynamics in *E. coli*. To this aim, we experimentally measured the DNA abundance distribution of *E. coli* growing at different temperatures between 17°C and 37°C using high throughput sequencing. Our theory reveals that the average speed of replisomes exhibits an Arrhenius dependence on the temperature, with an almost five-fold variation in the range we considered. Moreover, the speed of replisomes varies along the genome in a seemingly periodic and highly repeatable fashion around this average value. We find that this pattern is highly correlated with previously observed wave-like variations of the single base pair mutation rate along the bacterial genome [17, 18]. We discuss possible common causes for these two patterns.

Results

Distribution of genome types

We consider a population of bacteria that grow exponentially in a steady, nutrient-rich environment. Each cell in the growing population can encompass three types of genomes, see Fig. 1a and Fig. 1b: (i) one template genome, i.e., the genome that the cell inherited at its birth. (ii) incomplete genomes, i.e., genomes which are being

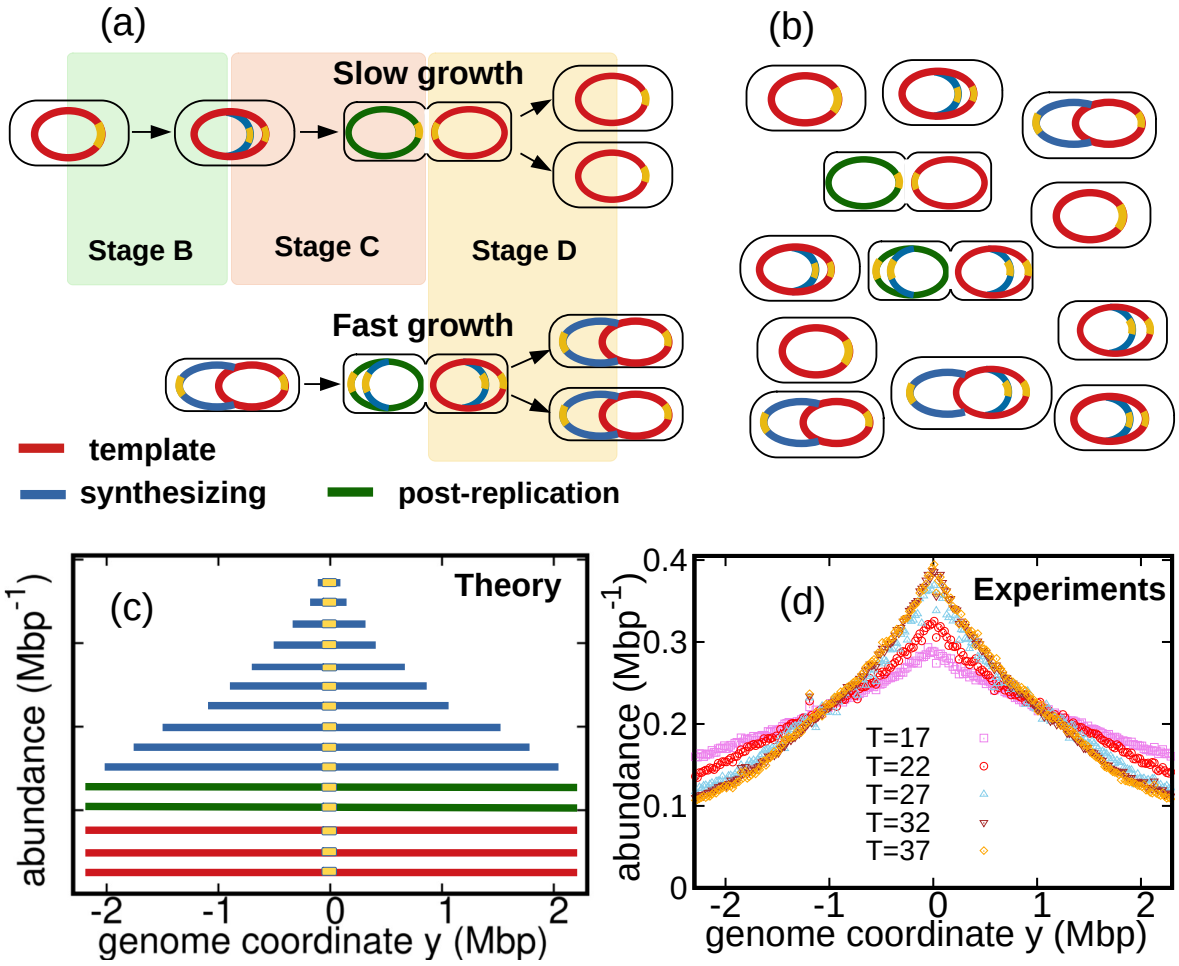

Figure 1: Genome types and DNA abundance distribution in an exponentially growing population. (a) Cell cycle: In slow growth conditions (top panel), newborn cells contain a template (stage B, red). As cell cycle progresses, replisomes begin synthesis of a new genome (stage C, blue) from the origin on the template (yellow spot). When replication terminates, cells contain the original template and a post-replication genome (stage D, green). Upon subsequent cell division, the post replication genome becomes the template for the newborn cell. In fast growth conditions (bottom panel), newborn cells acquire a template which is already undergoing synthesis. (b) Composition of genomes in an exponentially growing population of cells. Each cell may contain a different number of genomes, depending on its stage in the cell cycle and growth conditions. (c) Sketch of the DNA abundance distribution as a function of the genome coordinate. All three types of genomes contribute to the DNA abundance distribution. Because of incomplete genomes, the DNA abundance is largest at the origin and smallest at the terminal region (i.e., towards the periphery). (d) Experimental DNA abundance distribution at different temperatures.

synthesized. (iii) post-replication genomes that will be passed to new cells and become their templates.

In nutrient-rich conditions, bacteria replicate their genome in parallel, so that the numbers of incomplete genomes and post-replication genome per cell are variable, see Fig. 1b. The classic Cooper-Helmstetter model [8] describes the dynamics of these genomes in a given cell through generations. We adopt a different approach and focus on the abundance dynamics of the three types of genomes in the whole population. We call $N_T(t)$, $N_S(t)$, $N_P(t)$ the total number of template genomes, incomplete (synthesizing) genomes, and post-replication genomes, respectively, that are present in the population as a function of time t. Our aim is to quantify the relative fractions of these three types of genomes.

The total number of genomes is $N(t) = N_T(t) + N_S(t) + N_P(t)$. Since each cell contains exactly one template, the total number of cells is equal to $N_T(t)$. The genome numbers evolve as

$$\frac{d}{dt}N_T(t) = \alpha N_P \tag{1}$$

$$\frac{d}{dt}N_S(t) = kN - \beta N_S \tag{2}$$

$$\frac{d}{dt}N_P(t) = \beta N_S - \alpha N_P, \tag{3}$$

where α is the cell division rate per post-replication genome, k is the rate per genome at which forks are fired in the population, and β is the rate of completion of DNA replication per incomplete genome. It follows from Eq. (1) that, in steady growth, the total number of genomes grows exponentially at rate k, $N(t) \propto \exp(kt)$. In this exponential regime, the steady fractions of the three genome types are constant:

$$\frac{N_T(t)}{N(t)} = \frac{\alpha\beta}{(k+\beta)(k+\alpha)} \tag{4}$$

$$\frac{N_S(t)}{N(t)} = \frac{k}{(k+\beta)} \tag{5}$$

$$\frac{N_P(t)}{N(t)} = \frac{\beta k}{(k+\beta)(k+\alpha)}. \tag{6}$$

The ratio N/N_T can be interpreted as the average number of genomes per cell. Since this ratio is constant, the fork firing rate k can also be identified as the exponential growth rate of the number of cells.

DNA abundance distribution

The DNA abundance distribution $\mathcal{A}(y)$ is the probability that a small DNA fragment randomly picked from the population originates from genome position y, see Fig. 1c. We define the genome coordinate $y \in [-L/2, L/2]$, where $y = 0$ corresponds to the origin of replication and L is the genome length. Templates and post-replication genomes yield a uniform contribution to the distribution $\mathcal{A}(y)$ (red and green in Fig. 1c). In contrast, incomplete genomes contribute in a way that depends on the replisome positions along the genomes (blue in Fig. 1c). Our experiments permit to measure the distribution

$\mathcal{A}(y)$ with high accuracy, see Fig. 1d and Methods. Our goal is to use this experimental information to infer the distribution of replisomes, and thereby their dynamics.

To this aim, we now analyze the incomplete genomes in detail. We call x_1 and x_2 the portions of a given incomplete genome copied by the two replisomes at a given time, with $0 \leq x_1, x_2 \leq L$. Replication initiates at $x_1 = x_2 = 0$ and completes once the replisomes meet each other, i.e. $x_1 + x_2 = L$. In steady exponential growth, the copied portions x_1, x_2 of an incomplete genome are characterized by a stationary probability distribution $p^{\rm st}(x_1, x_2)$. This distribution depends on the replisome dynamics, and is directly linked with the DNA abundance distribution $\mathcal{A}(y)$, see Methods.

Replisome dynamics

The last ingredient of our theory is the dynamics of replisomes. We assume that each replisome is characterized by a speed $v(x)$ which depends, in principle, on the replisome position x (be it x_1 or x_2) and on a diffusion coefficient D representing random fluctuations of the speed. Close to thermodynamic equilibrium, the diffusion coefficient D can be estimated by the Stokes-Einstein relation [19]. However, since replisomes are driven far from equilibrium by the consumption of dNTPs, their diffusion coefficient could deviate from this estimate. In the absence of fluctuations ($D = 0$), each of the two replisomes copies exactly half of the genome, whereas for $D > 0$ their meeting point is characterized by a certain degree of uncertainty.

For given choices of $v(x)$ and D, our theory permits to compute $p^{\rm st}(x_1, x_2)$ and thereby the DNA abundance distribution $\mathcal{A}(y)$, see Methods. Therefore, by experimentally measuring the DNA abundance distribution we can test our hypotheses on the speed function $v(x)$ and the diffusion coefficient D.

Constant speed model

We first consider a scenario in which replisomes progress at a constant speed $\bar{v}$ and without fluctuations, $D = 0$. We find that, in this case, the DNA abundance distribution is expressed by

$$\mathcal{A}(y) = \frac{k}{2\bar{v}[1 - e^{-kL/2\bar{v}}]} e^{-k|y|/\bar{v}}. \tag{7}$$

We fit this distribution to the experimental data using maximum likelihood, see Fig. 2a. The speed v is the only fitting parameter as we independently measure the exponential growth rate k from the optical density, see Supporting Figure S1.

We find that the speed increases nearly five fold with temperature in the range we considered, see Fig. 2b. This temperature dependence appears to follow an Arrhenius law, see Fig. 2b. This behavior resembles that of the growth rate. The effective activation energy characterizing the cell cycle is larger than that characterizing the replisome speed, see Fig. 2c, possibly due the large number of molecular reactions involved in the cell cycle. The data point at 17^oC appears to deviate from the Arrhenius law for both the temperature and the growth rate [20], see Fig 2c.

In fast growth conditions (i.e., high temperature), multiple replisomes synthesize DNA in parallel inside each cell. In the temperature range we studied, we predict that

the number of replisomes per complete genome increases from two to almost six, see Fig. 2d.

The average DNA content per cell $\mathcal{C}$ is the product of the average genome length ℓ times the average number of genomes per cell N/N_T. Using Eq. (13) for the average genome length and Eq. (6) for the average number of genomes per cell, we obtain that the average DNA content per cell is expressed by

$$\mathcal{C} = \frac{2\bar{v}}{k}\frac{k+\alpha}{\alpha}\left[e^{kL/(2\bar{v})} - 1\right]. \tag{8}$$

Since we assumed constant speed and $D = 0$, the duration of the replication cycle is equal to $L/(2\bar{v})$ and does not fluctuate. The classic Cooper-Helmstetter model [8] predicts the DNA content per cell assuming constant durations of stages B, C, and D of the cell cycle. Indeed, if we make the same assumptions, we find that the prediction of Eq. (8) becomes equivalent to that of the Cooper-Helmstetter model (see SI).

Equation (8) gives us a chance to discuss DNA-protein homeostasis, i.e. the notion that the ratio between the amount of DNA and the amount of proteins per cell should be maintained approximately constant at varying the growth rate [21]. If the growth rate is varied by changing the nutrient composition, $\bar{v}$ remains constant, and Eq. (8) predicts an approximately exponential growth of $\mathcal{C}$ with k, which is consistent with observations [8]. In this case, the Schaechter–Maaloe–Kjeldgaard growth law states that the cell size grows exponentially with k as well [22], thereby ensuring DNA-protein homeostasis. In the case of varying temperature, we find that $\bar{v}$ and k present a similar dependence on T (see Fig. 2c), so that their ratio and thereby $\mathcal{C}$ weakly depends on k (see Supporting Figure S5). Our result is consistent with the observation that, at varying temperature, the cell size remains approximately constant as well [23, 24].

Oscillating speed model

The assumption of constant speed leads to a rather good fit of our DNA abundance data. However, the precision of our sequencing data permits us to appreciate systematic deviations from the model predictions under the constant speed hypothesis, see Fig. 3(a-e). These deviations appear as regular oscillations as a function of the genome coordinate. They are highly repeatable (see Supporting Figure S6, S7, and S8) and approximately symmetric with respect to the origin of replication. Our analysis reveals similar oscillations in experimental data from a previous study [15], see Supporting Figure S9. Taken together, these evidences support that this phenomenon is very robust.

To account for this observation, we introduce a more refined model in which the replisome speed oscillates along the genome:

$$v(x) = \bar{v}[1 + \delta\cos(\omega x + \phi)], \tag{9}$$

where δ represent the relative amplitude of oscillations; ω their angular frequency along the genome; and ϕ their initial phase. We also take into account random speed fluctuations in this case, $D \geq 0$. By fitting the DNA abundance, we estimate the parameters $\bar{v}$, δ, ω, ϕ, and D, see Fig. 3(f-j) and Table 1.

Our fitted speed oscillations are reminiscent of a previously observed wave-like pattern in the mutation rate along the genome of different bacterial species [18, 17]. For

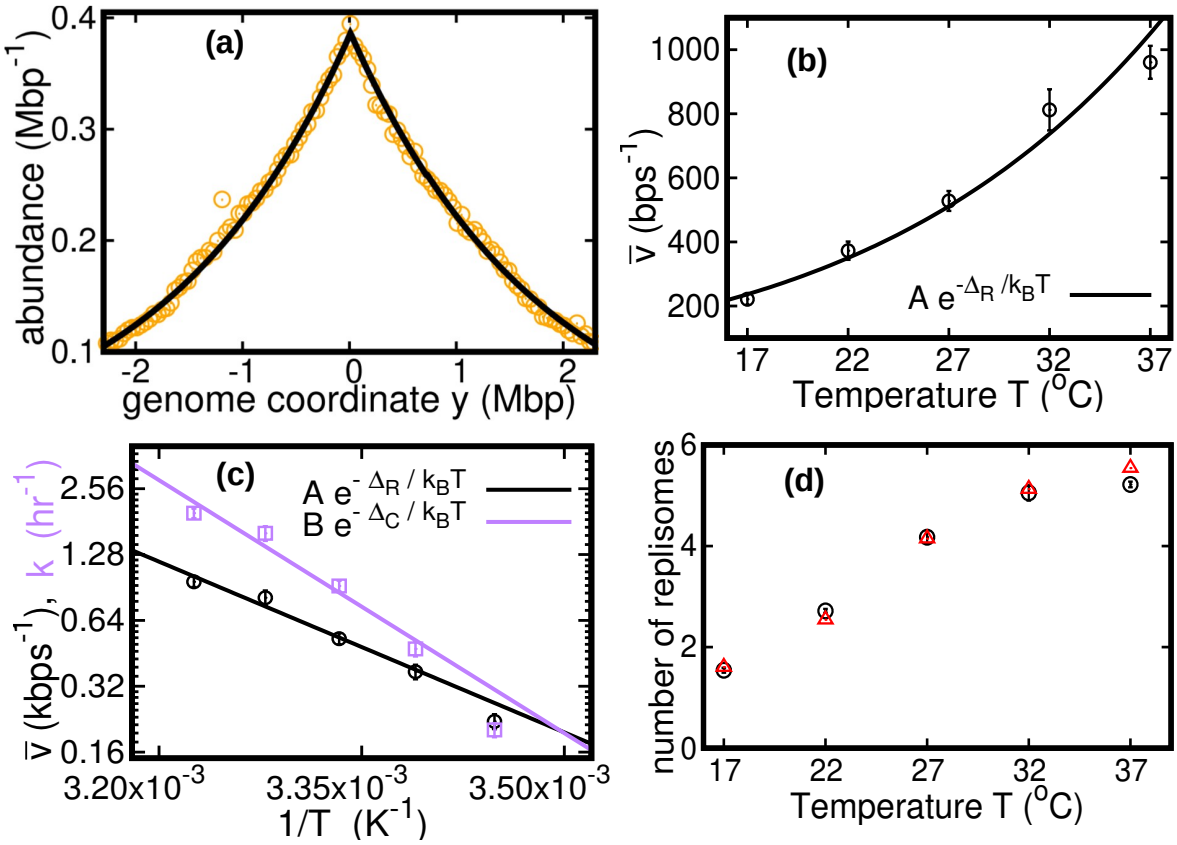

Figure 2: Results of the constant speed model. (a) DNA abundance distribution for $T = 37^{\circ}$C. Orange circles represent experimental data. The solid black line is the prediction of our model assuming constant speed and $D = 0$. Fits are performed using a maximum likelihood method, see SI for details. The quality of fits for replicates and other temperatures is comparable, see Supporting Figures S2, S3, S4. In particular, fits of replicates yield similar values of the speed $\bar{v}$, whereas the speed appreciably depends on temperature. (b) Replisome speed as a function of temperature. Error bars represent sample-to-sample variations. (c) Comparison of the temperature-dependence of speed and growth rate (see Supporting Figure S1). The solid curves are fits of Arrhenius laws to the data. The fitted parameters are $A = (2.5 \pm 5.3) \times 10^8 \text{bps}^{-1}$, $\Delta_R = (50 \pm 5) \text{kJmol}^{-1}$, $B = (6.0 \pm 24.9) \times 10^{12} \text{hr}^{-1}$ and $\Delta_C = (74 \pm 10) \text{kJmol}^{-1}$. We exclude the data point for $T = 17^{\circ}$C in both fits. (d) Estimated number of replisomes per complete genome at different temperature. The red triangles represents the estimate from Eq. (17) in which we use the expression of β for the constant speed model, Eq. (20). The black circles are the estimates from Eq. (18).

T	$\bar{v}$	δ	ω	ϕ	D
17	246±33	0.22±0.13	0.7±0.5	3.3±1.0	0.39±0.43
22	351±30	0.20±0.06	2.7±0.6	3.4±0.6	0.81±1.18
27	541±30	0.18±0.03	4.7±0.1	2.1±0.1	0.35±0.49
32	821±66	0.11±0.04	5.5±0.2	1.5±0.1	1.15±1.23
37	970±51	0.17±0.03	4.3±0.2	3.0±0.2	2.90±2.48

Table 1: Parameters of the oscillatory speed model. Temperatures are expressed in ^{o}C, $\bar{v}$ in bps^{-1}, ω in $\mathrm{rad\,Mbp}^{-1}$, ϕ in radians and D in $\mathrm{kbp}^2\mathrm{s}^{-1}$. Reported values are average over experimental replicates. Error bars represent sample-to-sample variations. The average speed estimated from the oscillatory speed model and constant speed model are comparable, see Supporting Table S1.

a quantitative comparison, we analyze this pattern in a mutant *E. coli* strain lacking DNA mismatch repair [17]. We find that the oscillations in mutation rate and in speed are highly correlated, see Fig. 4a. The mutation rate appears approximately in phase with the speed, meaning that regions where replisomes proceed at higher speed are characterized by a higher mutation rate. This observation strongly suggests that the two phenomena have a common cause.

It was argued that the oscillations in the mutation rate could originate from a systematic process related with the cell cycle [17]. We test this hypothesis for the speed oscillations. Assuming approximately constant cell division times, we estimate the cell division time as $\tau = (\ln 2)/k$. If the speed of replisomes was coupled to a factor oscillating with period τ, this would cause spatial oscillation of speed with angular frequency $\omega = 2\pi/(\bar{v}\tau) = 2\pi k/[(\ln 2)\bar{v}]$. This prediction qualitatively agrees with our fitted values of ω, see Fig. 4b. Since k is also equal to the fork firing rate per genome, an alternative interpretation is that the oscillations are caused by competition among replisomes for nucleotides or other molecules required for replication.

If the wave-like pattern were caused by competition among replisomes, one would expect either a minimum of the speed every time a new fork is fired ($\phi = \pi$) or the speed to start decreasing when a new fork is fired ($\phi = \pi/2$). Our fitted values of the phase ϕ are compatible with this range, see Fig. 4c.

Our results show that the diffusion coefficient D is quite small. For about one third of our experimental realizations at each temperature, our fitted value of D is not significant according to the Akaike information criterion (see Fig. 4d and Supporting Figure S11). For comparison, we estimate the equilibrium diffusion constant of replisomes in the cytoplasm from the Stokes-Einstein relation as $D_{\mathrm{SE}} \approx 6\mathrm{kbp}^2\mathrm{s}^{-1}$ (see SI), of the same order of magnitude of our fitted values, see Table 1 and Fig. 4. Taken together, these results suggest that, despite its fast speed, the magnitude of fluctuations of replisome position are remarkably similar to the equilibrium case.

The diffusion coefficient determines the uncertainty about the genome site where the two replisomes meet. In the absence of diffusion ($D = 0$), replisomes would always meet at the mid point on the circular genome. For $D > 0$, we estimate the typical size l_{D} of the region in which the two replisomes meet as follows. Since the fitted values of δ and D are both small, we approximate the replication time as $\tau_C \approx L/(2\bar{v})$. In this time, the accumulated uncertainty due to diffusion is equal to $l_{\mathrm{D}} \approx 2\sqrt{2D\tau_D}$. From our estimated diffusion coefficients and average velocities, we obtain values of l_{D} on the

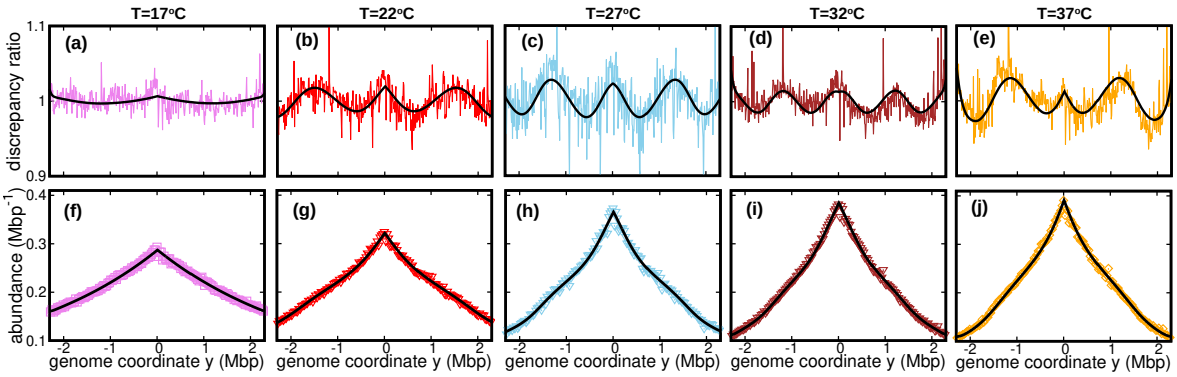

Figure 3: Deviations from the predictions of the constant speed model indicate that the replisome speed oscillates along the genome. (a-e) Colored lines: ratios of the experimental DNA abundance over the corresponding prediction assuming constant speed and $D = 0$. These ratios exhibit a wave-like pattern. The solid black lines represent the ratios of the predictions assuming oscillatory speed, Eq. (9) and $D \geq 0$, over constant speed and $D = 0$. Corresponding plots for replicates and other temperatures are presented in Supporting Figures S6, S7, S8. (f-j) Experimental DNA abundance distribution at different temperatures. The solid black lines are the fits of the oscillatory speed model. Tests based on the Akaike information criterion show that the oscillatory speed model should be chosen over the constant speed model for all the replicates at different temperatures, see Supporting Figure S10. The fitted parameters are reported in Table 1.

order of $100 - 200$kbp, depending on temperature.

Discussion

In this paper, we developed a theory to infer the dynamics of replisomes from the DNA abundance distribution in a growing bacterial population. We tested our method by measuring the DNA abundance distribution of growing *E. coli* populations at different temperatures. We found that the dependence of the average speed on the temperature is well described by an Arrhenius law, similar to that governing the population growth rate. Our analysis shows that this result is compatible with DNA-protein homeostasis, given that the average cell size hardly varies with temperature [23, 24].

Our approach reveals a wave-like oscillation of the replisome speed along the *E. coli* genome. The relative amplitude of these oscillations ranges from 10% to 20% of the average replisome speed. A quantitatively similar pattern was observed in the bacterial mutation rate along the DNA of an *E. coli* mutant strain [17] and in other bacterial species [18]. This similarity strongly suggests that the two phenomena have a common dynamical origin. In particular, we suggest that their link could be the trade-off between accuracy and speed that characterizes DNA polymerases [25, 26, 27]. Because of this tradeoff, any mechanism increasing the speed of a polymerase is expected to reduce its accuracy as well. Our analysis of the wave length of these oscillations supports that this pattern originates from a process synchronized with the cell cycle [18], whose activity alters the replisome function. An alternative explanation is that the oscillations

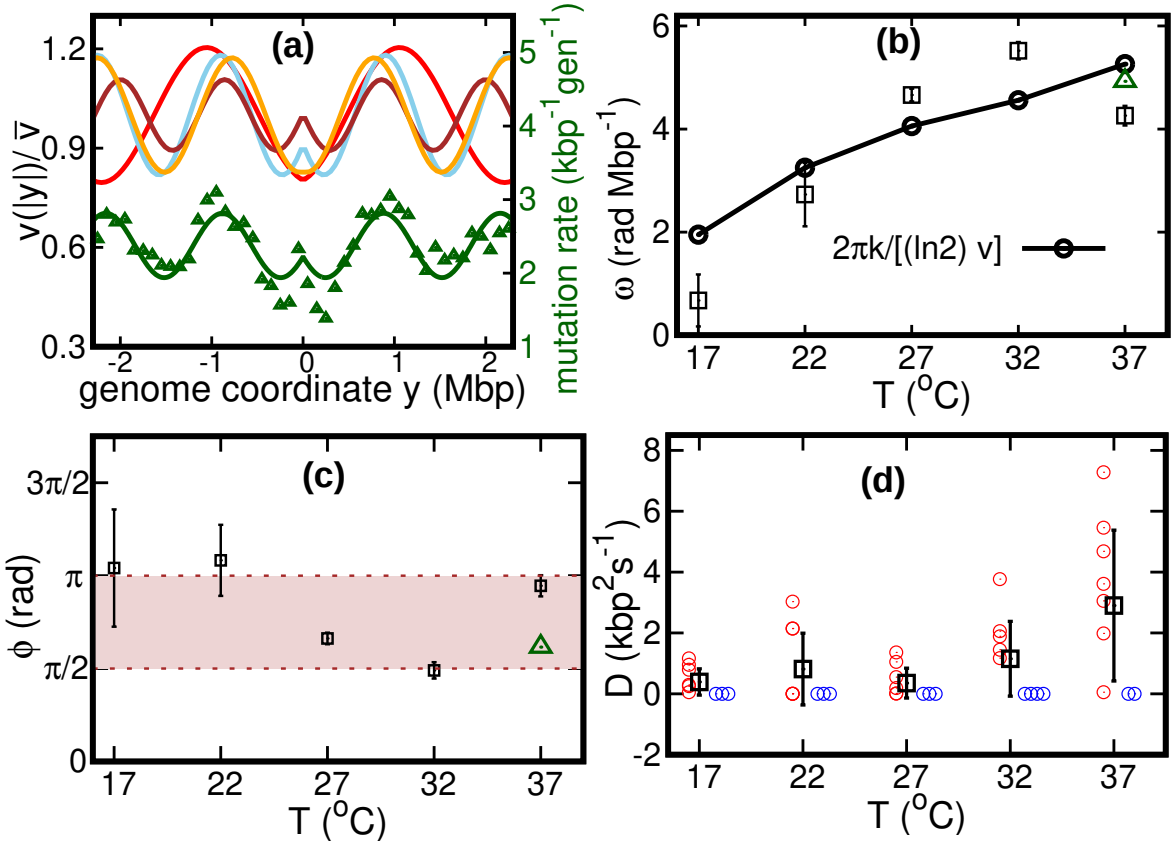

Figure 4: Results of the oscillatory speed model. (a) Solid lines: relative speeds $v(|y|)/\bar{v}$ along the genome (Red: $T = 22$°C, sky blue: $T = 27$°C, brown: $T = 32$°C, and orange: $T = 37$°C). We omitted the curve for $T = 17$°C as the oscillations are less evident in this case. The wave-like pattern of the speed is quantitatively similar to the variations of the mutation rate along the genome (green triangles, from [17]; Pearson correlation coefficients between speed and mutation rate: $r_{22C} = 0.40$; $r_{27C} = 0.84$; $r_{32C} = 0.80$; and $r_{37C} = 0.69$). The mutation rate is defined as the number of base pair substitutions per generation per kilo base pairs. The solid green line is a fit to the mutation rate data with a function that has same form as in (9). The fit parameters are $\bar{v} = 2.4\text{kbp}^{-1}\text{gen}^{-1}$, $\delta = 0.18$, $\omega = 4.9\text{rad}\,\text{Mbp}^{-1}$ and $\phi = 1.93\text{rad}$. (b) Temperature dependence of angular frequency of oscillation ω. (c) phase ϕ. Green triangles in (b) and (c) represent the angular frequency and phase, respectively, from the fit to the mutation rate data with Eq. (9). (d) Diffusion coefficient D. Circles represent individual fitted values of diffusion coefficients. Blue circles represent cases in which the fitted value of D is either zero or not significant (see SI). This occurs in 2 out of 9 cases for 37°C and 3 out of 9 cases for the other temperatures.

originate from competition among replisomes for shared resources.

Beside these regular and repeatable fluctuations, our analysis shows that random fluctuations of replisome speed are quite small, leading to an uncertainty of about $100 - 200kpb$ on the location of the replisome meeting point. In bacteria, the terminal region of replication is flanked by two groups of termination (Ter) sites having opposite orientation. Ter sites are binding site for the Tus protein and permit passage of replication forks in one direction only [3], so that the two groups effectively trap the two forks in the terminal region [28]. Out of the ten Ter sequences in *E. coli*, only two of them (TerB and TerC) are within $100 - 200kpb$ of the point diametrically opposite to the origin. These two sequences have the same orientation. Our result therefore implies that most Ter sequences are usually not needed to localize the replisome meeting point. This prediction is consistent with previous observations that the phenotypes of Tus-*E. coli* mutants [29] or mutants lacking Ter sequences [28] do not appear distinct from that of the wild type.

Quantitative modeling of the DNA abundance distribution has the potential to shed light on aspects of replisome dynamics beyond those explored in this paper. For example, it was observed that knockout of proteins involved in completion of DNA replication leads to either over-expression or under-expression of DNA in the terminal region [12, 13, 30]. Incorporating the role of these proteins into our model will permit to validate possible explanations for these patterns. More in general, our approach is simple and general enough to be readily applied to other bacterial species, to unravel common principles and differences in their DNA replication dynamics

Materials and Methods

Cultivation

E. coli MG1655 was cultured in LB medium supplemented with 50mM MOPS pH 7.2 and 0.2% glucose. Overnight cultures grown at 37°C were diluted into fresh medium and grown until reaching an OD600 of about 1.0 at the target temperature. These cultures were used to inoculate 50 ml medium at the desired temperature in 500ml Erlenmeyer flasks with baffles at a target OD of 0.01. Cultivation was always performed with shaking at 250 rpm. OD was determined with a NanoDrop One in cuvette mode.

DNA extraction

E. coli cultures (1.4 ml) were harvested by centrifugation at 21000g for 20 seconds after reaching an OD of around 0.5 (mid-exponential phase, see Supporting Figure S1). Cells were kept growing for at least 45 doubling times (as measured in exponential phase) to reach stationary phase. Samples of 0.2 ml from the stationary phase cultures grown at 17^oC, 27^oC, and 37^oC were harvested for DNA extraction. The pellets were immediately frozen at -80°C until DNA extraction. DNA was extracted in parallel using Genomic DNA Purification Kit from Thermo Fisher Scientific.

Sequencing

We sequenced three samples in the exponential phase for each temperature and three stationary samples at three different temperatures. DNA samples were sheared by ultrasound using Covaris AFA technology. Libraries were prepared using the Illumina DNA PCR-Free Library Prep Kit. Sequencing was performed on a Novaseq6000 using paired-end 150 bp reads.

Alignment and bias elimination

We aligned reads from each sample using Bowtie2 2.3.4.1 [31], using the MG1655 genome as a reference. We calculated the frequency of reads as a function of the genome coordinate with bin size 10kbp. To attenuate bias, we divided the frequency at each genome coordinate in a sample from the exponential phase by the frequency of the corresponding bin in a stationary sample [12, 15]. We alternatively used all of our three stationary samples to correct the bias of each sample in the exponential phase. Therefore, after bias elimination, we effectively have $3 \times 3 = 9$ different DNA abundance curves in the exponential phase at each temperature. See SI for details.

DNA abundance from the replisome distribution

In this section we link the DNA abundance distribution $\mathcal{A}(y)$ with the replisome position distribution $p^{\rm st}(x_1, x_2)$. We first introduce the probability $\mathcal{P}(y)$ that a randomly chosen genome (either complete or incomplete) in the population includes the genome location y. The integral of $\mathcal{P}(y)$ is equal to the average genome length ℓ in the population

$$\ell = \int_{-L/2}^{L/2} \mathcal{P}(y) dy. \tag{10}$$

It follows from Eq. (5) that a randomly chosen genome is complete with probability $(1 - N_S/N) = \beta/(k+\beta)$. For incomplete genomes, we take into account that either of the replisomes can have copied position y. This argument leads to the expression

$$\mathcal{P}(y) = \frac{k}{\beta + k} \left[\int_{|y|}^{L} dx_1 \int_{0}^{L-x_1} dx_2 \; p^{\rm st}(x_1, x_2) + \int_{L-|y|}^{L} dx_2 \int_{0}^{L-x_2} dx_1 \; p^{\rm st}(x_1, x_2) \right] + \frac{\beta}{k+\beta}. \tag{11}$$

The DNA abundance distribution $\mathcal{A}(y)$ is proportional to $\mathcal{P}(y)$, up to a normalization constant:

$$\mathcal{A}(y) = \frac{\mathcal{P}(y)}{\int_{-L/2}^{L/2} \mathcal{P}(y') dy'}. \tag{12}$$

Combining Eq. (10), Eq. (12), and the fact that $\mathcal{P}(0) = 1$, we obtain a simple relation between the DNA abundance distribution and the average genome length:

$$\ell = \mathcal{A}(0)^{-1}. \tag{13}$$

Dynamics of incomplete genomes

We call $n_S(x_1, x_2; t)$ the number density of incomplete genomes at time t with replisome positions at x_1 and x_2. By definition $\int_0^L dx_1 \int_0^{L-x_1} dx_2 n_S(x_1, x_2; t) = N_S(t)$. We assume that this number density evolves according to

$$\frac{\partial}{\partial t} n_S(x_1, x_2; t) = -\vec{\nabla} \cdot [\vec{v} n_S] + D\nabla^2 n_s, \tag{14}$$

where $\vec{\nabla} = (\partial/\partial_{x_1}, \partial/\partial_{x_2})$ and $\vec{v} = (v(x_1), v(x_2))$. We now introduce the normalized probability $p(x_1, x_2; t) = n_S(x_1, x_2; t)/N_S(t)$. By substituting this definition into (14), we obtain

$$\frac{\partial}{\partial t} p(x_1, x_2; t) = -\vec{\nabla} \cdot [\vec{v} p] + D\nabla^2 p - kp. \tag{15}$$

The last term in (15) is a dilution term that accounts for the exponentially increase in newborn cells. The stationary distribution $p^{\text{st}}(x_1, x_2)$ is a time-independent solution of Eq. (15).

Because of replication completion, the line $x_1 + x_2 = L$ is an absorbing state for the dynamics described by Eq. (15). Equation (15) must be consistent with Eq. (2), which describes the dynamics of incomplete genomes regardless of the coordinates of their replisomes. This implies that the rate β at which replication completes must equal to the probability flux through the absorbing boundary:

$$\beta = \int_{x_1+x_2=L} \vec{J} \cdot \hat{n} \; dl, \tag{16}$$

where we introduce the probability current $\vec{J}(x_1, x_2) = \vec{v}p - D\vec{\nabla}p$, the unit vector $\hat{n} = (1/\sqrt{2}, 1/\sqrt{2})$, and the infinitesimal line increment dl along the absorbing boundary. Similarly, the probability flux entering the system at $(x_1, x_2) = (0, 0)$ should match the rate of replication initiation as given by (2).

Average number of replisomes per complete genome

We estimate the average number of replisomes per complete genome $\mathcal{N}$ in two alternative ways. On the one hand, using Eqs. (4)-(6) we find that

$$\mathcal{N} = \frac{2N_S}{N_P + N_T} = \frac{2k}{\beta}. \tag{17}$$

On the other hand, it can be seen in Fig. 2d that the fraction of complete genome in the population is equal to the ratio $\mathcal{A}(L/2)/\mathcal{A}(0)$ between the DNA abundance at the terminal and at the origin. It follows that

$$\mathcal{N} = \frac{2[\mathcal{A}(0) - \mathcal{A}(L/2)]}{\mathcal{A}(L/2)}. \tag{18}$$

Constant speed

We focus on the scenario with constant speed and $D = 0$. In this case, the steady solution of Eq. (15) is given by

$$p^{\text{st}}(x_1, x_2) = \frac{k e^{-\frac{k}{2\bar{v}}(x_1+x_2)}}{\bar{v}\left(1 - e^{-kL/(2\bar{v})}\right)} \delta(x_1 - x_2). \tag{19}$$

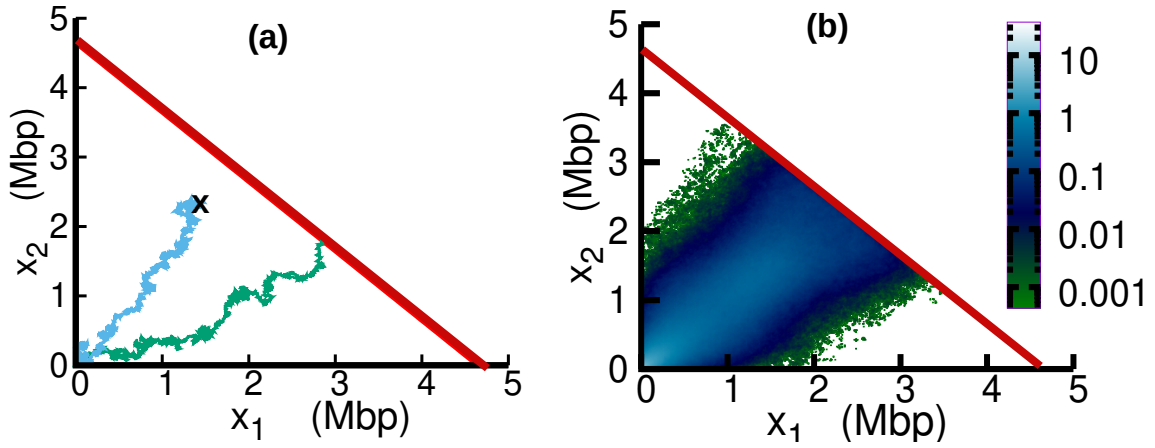

Figure 5: Replisome dynamics in the (x_1, x_2) plane. (a) Two different trajectories demonstrate two different types of resetting events in our simulations. The trajectories are reset to $x_1 = 0, x_2 = 0$ when either the two replisomes meet (green trajectory) at the absorbing boundary (solid red line) or when a new initiation occurs (sky blue). (b) Replisome position distribution $p^{\rm st}(x_1, x_2)$ in the steady state. In both panels, parameters are $\bar{v} = 973\text{bps}^{-1}$, $\delta = 0.19$, $\omega = 4\text{rad}\,\text{Mbp}^{-1}$, $\phi = 3.1$ and $D = 55\text{kbp}^2\text{s}^{-1}$. These parameters are on the order of those fitted from experiments (see Table 1), except for D which is chosen to be larger for illustration purposes.

The rate at which replication completes is equal to

$$\beta = \frac{k e^{-kL/(2\bar{v})}}{1 - e^{-kL/(2\bar{v})}}. \tag{20}$$

Substituting Eqs. (19) and (20) into Eq. (11), we obtain

$$\mathcal{P}(y) = e^{-k|y|/\bar{v}}, \tag{21}$$

from which Eq. (7) follows by normalizing, see Eq. (12).

Stochastic simulations

In the case of oscillating speed and $D > 0$, we compute the stationary solution of Eq. (15) using numerical simulations. To this aim, we interpret Eq. (15) as describing a stochastic process subject to stochastic resetting [32]. Specifically, we evolve trajectories according to a system of stochastic differential equations:

$$\begin{aligned} \frac{d}{dt}x_1(t) &= v(x_1) + \sqrt{2D}\,\xi_1(t) \\ \frac{d}{dt}x_2(t) &= v(x_2) + \sqrt{2D}\,\xi_2(t)\,, \end{aligned} \tag{22}$$

where $\xi_1(t)$ and $\xi_2(t)$ are white noise sources satisfying $\langle \xi_1(t) \rangle = \langle \xi_2(t) \rangle = 0$, $\langle \xi_1(t)\xi_1(t') \rangle = \langle \xi_2(t)\xi_2(t') \rangle = \delta(t - t')$, and $\langle \xi_1(t)\xi_2(t') \rangle = 0$. In addition to the dynamics described by Eqs. (22), with a stochastic rate k, trajectories are reset to the origin, $x_1 = x_2 = 0$ (blue trajectory in Fig. 5a). Since the boundary $x_1 + x_2 = L$ is an absorbing state,

trajectories that reach this boundary are also reset to the origin (green trajectory in Fig. 5a). The probability distribution associated with this dynamics evolves according to Eq. (15). We simulate this stochastic dynamics to estimate the stationary distribution $p^{\text{st}}(x_1, x_2)$ in a computationally efficient way, see Fig. 5b. We also estimate from the same simulations the parameter β as the empirical rate at which the absorbing boundary is reached, see Eq. (16).

References

[1] Tania A Baker and Stephen P Bell. Polymerases and the replisome: machines within machines. *Cell*, 92(3):295–305, 1998.

[2] Tuan Minh Pham, Kang Wei Tan, Yuichi Sakumura, Katsuzumi Okumura, Hisaji Maki, and Masahiro Tatsumi Akiyama. A single-molecule approach to dna replication in escherichia coli cells demonstrated that dna polymerase iii is a major determinant of fork speed. *Molecular Microbiology*, 90(3):584–596, 2013.

[3] Mohamed M Elshenawy, Slobodan Jergic, Zhi-Qiang Xu, Mohamed A Sobhy, Masateru Takahashi, Aaron J Oakley, Nicholas E Dixon, and Samir M Hamdan. Replisome speed determines the efficiency of the tus- ter replication termination barrier. *Nature*, 525(7569):394–398, 2015.

[4] R.M. Schaaper. Base selection, proofreading, and mismatch repair during dna replication in escherichia coli. *Journal of Biological Chemistry*, 268(32):23762–23765, 1993.

[5] Liselot Dewachter, Natalie Verstraeten, Maarten Fauvart, and Jan Michiels. An integrative view of cell cycle control in Escherichia coli. *FEMS Microbiology Reviews*, 42(2):116–136, 01 2018.

[6] J.D. Wang and P.A. Levin. Metabolism, cell growth and the bacterial cell cycle. *Nature Reviews Microbiology*, 7(11):822–827, 2009.

[7] Lisa Willis and Kerwyn Casey Huang. Sizing up the bacterial cell cycle. *Nature Reviews Microbiology*, 15(10):606–620, 2017.

[8] S Cooper and C E Helmstetter. Chromosome replication and the division cycle of escherichia coli b/r. *Journal of molecular biology*, 31(3):519–40, Feb 1968.

[9] Solveig Fossum, Elliott Crooke, and Kirsten Skarstad. Organization of sister origins and replisomes during multifork dna replication in escherichia coli. *The EMBO journal*, 26(21):4514–22, Oct 2007.

[10] O Pierucci. Chromosome replication and cell division in escherichia coli at various temperatures of growth. *Journal of bacteriology*, 109(2):848–854, 1972.

[11] Nina Y. Yao, Roxana E. Georgescu, Jeff Finkelstein, and Michael E. O'Donnell. Single-molecule analysis reveals that the lagging strand increases replisome processivity but slows replication fork progression. *Proceedings of the National Academy of Sciences*, 106(32):13236–13241, 2009.

[12] Brian M. Wendel, Charmain T. Courcelle, and Justin Courcelle. Completion of dna replication in escherichia coli. *Proceedings of the National Academy of Sciences*, 111(46):16454–16459, 2014.

[13] Brian M. Wendel, Jessica M. Cole, Charmain T. Courcelle, and Justin Courcelle. Sbcc-sbcd and exoi process convergent forks to complete chromosome replication. *Proceedings of the National Academy of Sciences*, 115(2):349–354, 2018.

[14] Christian J Rudolph, Amy L Upton, Anna Stockum, Conrad A Nieduszynski, and Robert G Lloyd. Avoiding chromosome pathology when replication forks collide. *Nature*, 500(7464):608–11, Aug 2013.

[15] Sarah L Midgley-Smith, Juachi U Dimude, Toni Taylor, Nicole M Forrester, Amy L Upton, Robert G Lloyd, and Christian J Rudolph. Chromosomal over-replication in Escherichia coli recG cells is triggered by replication fork fusion and amplified if replichore symmetry is disturbed. *Nucleic Acids Research*, 46(15):7701–7715, 06 2018.

[16] Sarah L Midgley-Smith, Juachi U Dimude, and Christian J Rudolph. A role for 3' exonucleases at the final stages of chromosome duplication in Escherichia coli. *Nucleic Acids Research*, 47(4):1847–1860, 12 2018.

[17] Brittany A. Niccum, Heewook Lee, Wazim MohammedIsmail, Haixu Tang, Patricia L. Foster, Julian E. Davies, Susan Lovett, and Vaughn Cooper. The symmetrical wave pattern of base-pair substitution rates across the escherichia coli chromosome has multiple causes. *mBio*, 10(4):e01226–19, 2019.

[18] Marcus M. Dillon, Way Sung, Michael Lynch, Vaughn S. Cooper, and Michael T. Laub. Periodic variation of mutation rates in bacterial genomes associated with replication timing. *mBio*, 9(4):e01371–18, 2018.

[19] J T Hynes. Statistical mechanics of molecular motion in dense fluids. *Annual Review of Physical Chemistry*, 28(1):301–321, 1977.

[20] Anjan Roy, Dotan Goberman, and Rami Pugatch. A unifying autocatalytic network-based framework for bacterial growth laws. *Proceedings of the National Academy of Sciences*, 118(33), 2021.

[21] Fangwei Si, Dongyang Li, Sarah E. Cox, John T. Sauls, Omid Azizi, Cindy Sou, Amy B. Schwartz, Michael J. Erickstad, Yonggun Jun, Xintian Li, and Suckjoon Jun. Invariance of initiation mass and predictability of cell size in escherichia coli. *Current Biology*, 27(9):1278–1287, 2017.

[22] M Schaechter, O Maaloe, and N O Kjeldgaard. Dependency on medium and temperature of cell size and chemical composition during balanced grown of salmonella typhimurium. *Journal of general microbiology*, 19(3):592–606, Dec 1958.

[23] T E Shehata and A G Marr. Effect of temperature on the size of escherichia coli cells. *Journal of Bacteriology*, 124(2):857–862, 1975.

[24] F J Trueba, E A van Spronsen, J Traas, and C L Woldringh. Effects of temperature on the size and shape of escherichia coli cells. *Archives of microbiology*, 131(3):235–40, May 1982.

[25] Pablo Sartori and Simone Pigolotti. Kinetic versus energetic discrimination in biological copying. *Physical review letters*, 110(18):188101, 2013.

[26] Kinshuk Banerjee, Anatoly B Kolomeisky, and Oleg A Igoshin. Elucidating interplay of speed and accuracy in biological error correction. *Proceedings of the National Academy of Sciences*, 114(20):5183–5188, 2017.

[27] William J Fitzsimmons, Robert J Woods, John T McCrone, Andrew Woodman, Jamie J Arnold, Madhumita Yennawar, Richard Evans, Craig E Cameron, and Adam S Lauring. A speed–fidelity trade-off determines the mutation rate and virulence of an rna virus. *PLoS biology*, 16(6):e2006459, 2018.

[28] Iain G. Duggin, R. Gerry Wake, Stephen D. Bell, and Thomas M. Hill. The replication fork trap and termination of chromosome replication. *Molecular Microbiology*, 70(6):1323–1333, 2008.

[29] B. Roecklein, A. Pelletier, and P. Kuempel. The tus gene of escherichia coli: autoregulation, analysis of flanking sequences and identification of a complementary system in salmonella typhimurium. *Research in Microbiology*, 142(2):169–175, 1991.

[30] Anurag Kumar Sinha, Christophe Possoz, Adeline Durand, Jean-Michel Desfontaines, François-Xavier Barre, David RF Leach, and Bénédicte Michel. Broken replication forks trigger heritable dna breaks in the terminus of a circular chromosome. *PLoS genetics*, 14(3):e1007256, 2018.

[31] Ben Langmead and Steven L Salzberg. Fast gapped-read alignment with bowtie 2. *Nature methods*, 9(4):357–359, 2012.

[32] Martin R. Evans and Satya N. Majumdar. Diffusion with stochastic resetting. *Phys. Rev. Lett.*, 106:160601, Apr 2011.