

Supplemental Information:

The *in vivo* measurement of replication fork velocity and pausing by lag-time analysis

Dean Huang,¹ Anna E. Johnson,^{2,3} Brandon S. Sim,¹ Teresa W. Lo,¹ Houra Merrikh,^{2,3,*} and Paul A. Wiggins^{1,4,5,†}

¹Department of Physics, University of Washington, Seattle, Washington 98195, USA

²Department of Biochemistry, Vanderbilt University, Nashville, Tennessee 37205, USA

³Department of Pathology, Microbiology, and Immunology,

Vanderbilt University Medical Center, Nashville, Tennessee 37232, USA

⁴Department of Bioengineering, University of Washington, Seattle, Washington 98195, USA

⁵Department of Microbiology, University of Washington, Seattle, Washington 98195, USA

SUPPLEMENTARY TABLES

1. Strains used in this study

Short name:	Strain name:	Description:	Source:
<i>V. cholerae</i> WT	O1 biovar El Tor N16961	<i>ChapR ΔlacZ</i> gm ^R	[1]
<i>V. cholerae</i> MCH1	MCH1	Integration of N16961 Chr2, without <i>oriC2</i> and partition machinery region, in place of the <i>dif1</i> site in N16961 Chr1.	[1]
<i>V. cholerae oriR4</i>	EGV111	N16961 <i>ChapR ΔlacZ oriC1</i> @ R4 (1,898Mb) gm ^R	[1]
<i>B. subtilis</i> WT	JH642	<i>trpC2 pheA1</i>	[2]
<i>B. subtilis</i> 257°:: <i>oriC</i>	MMB703	<i>trpC2 pheA1 argG(257)::(oriC/dnaAN kan) Δ(oriC-L)::spc</i>	[2]
<i>B. subtilis</i> 94°:: <i>oriC</i>	JDW258	<i>trpC2 pheA1 aprE(94)::(oriC/dnaAN kan) Δ(oriC-L)::spc dnaB134ts-zhb83::Tn917(cat)</i>	[2]
<i>B. subtilis oriN</i>	MMB208	<i>pheA1 (ypjG-hepT)122 spoIIJ(359)::(oriN kan tet) ΔoriC-S</i>	[2]
<i>B. subtilis</i> 257°:: <i>oriN</i>	MMB700	<i>pheA1 (ypjG-hepT)122 argG(257)::(oriN kan) ΔoriC-S</i>	[2]
<i>B. subtilis</i> YB886	YB886	<i>trpC2 metB5 sigB amyE spβ-ICEBs° xin-</i>	[3]
<i>B. subtilis rrnIHG(pre-inv)</i>	JDW858	YB886 <i>kbaA'::neo' cat::'ybaN rrnG-5S'::erm 'neo::'ybaR</i>	[3]
<i>B. subtilis rrnIHG(inv)</i>	JDW860	YB886 <i>kbaA'::neo erm inv(ybaN::rrnG-5S) cat::'ybaR</i>	[3]
<i>E. coli</i> WT	K-12 MG1655	F- <i>lambda- ilvG- rfb-50 rph-1</i>	[4]

* houra.merrikh@vanderbilt.edu

† pwiggins@uw.edu; <http://mtshasta.phys.washington.edu/>

2. Datasets used in this study

All datasets were obtained from cells grown at 37 °C. Next-generation-sequencing FASTQ files and marker frequencies that were generated for this study are available from the NCBI Sequence Read Archive with the BioProject accession code [PRJNA919081](#). The data from Galli *et al.* [1] and Midgley-Smith *et al.* [4] are both obtained from the European Nucleotide Archive (ENA), with the accession codes [PRJEB28538](#) and [PRJEB25595](#), respectively. The individual run accessions are also included. The digitized data from Wang *et al.* [2] and Srivatsan *et al.* [3] are available in the Source Data file, with the corresponding sheets listed in the table below. In the following table, Exp is shorthand for exponential phase growth and Stat is shorthand for stationary phase. Growth media are described in more detail in Supplementary Method 1.

Short name:	Growth media:	Doubling time:	Source:	Project accession:	Run accession:	Sample size (read count):
<i>V. cholerae</i> WT	LB	22 ± 1 min	This study	SRA: PRJNA919081	Exp: SRR23003324 Stat: SRR23003328	5.4 × 10 ⁷ 6.0 × 10 ⁷
	M9 fructose	50 ± 4 min	From study [1]	ENA: PRJEB28538	Exp: ERX2796386 Stat: ERX2796387	3.2 × 10 ⁷ 2.0 × 10 ⁷
<i>V. cholerae</i> MCH1	LB	90 ± 6 min	This study	SRA: PRJNA919081	Exp: SRR23003321 Stat: SRR23003311	5.4 × 10 ⁷ 6.1 × 10 ⁷
	M9 fructose	50 ± 4 min	From study [1]	ENA: PRJEB28538	Exp: ERX2796384 Stat: ERX2796385	1.1 × 10 ⁷ 1.5 × 10 ⁷
<i>V. cholerae</i> <i>oriR4</i>	M9 fructose	55 ± 5 min	From study [1]	ENA: PRJEB28538	Exp: ERX2796379 Stat: ERX2796380	1.5 × 10 ⁷ 1.3 × 10 ⁷
<i>B. subtilis</i> WT	S7 fumarate	61.0 ± 1.2 min	From study [2]	N/A	N/A	N/A (Microarray)
<i>B. subtilis</i> 257°:: <i>oriC</i>	S7 fumarate	Not reported	From study [2]	Digitized	Source Data: Bs_257	N/A (Microarray)
<i>B. subtilis</i> 94°:: <i>oriC</i>	S7 fumarate	Not reported	From study [2]	Digitized	Source Data: Bs_94	N/A (Microarray)
<i>B. subtilis</i> <i>oriN</i>	S7 fumarate	64.3 ± 2.9 min	From study [2]	Digitized	Source Data: Bs_oriN	N/A (Microarray)
<i>B. subtilis</i> 257°:: <i>oriN</i>	S7 fumarate	Not reported	From study [2]	Digitized	Source Data: Bs_oriN257	N/A (Microarray)
<i>B. subtilis</i> <i>rrnIHG</i> (pre-inv)	LB	20 ± 1 min	From study [3]	N/A	N/A	N/A (Microarray)
	MOPS glucose CA	28 ± 1 min	From study [3]	N/A	N/A	N/A (Microarray)
	MOPS glucose MM	42 ± 1 min	From study [3]	N/A	N/A	N/A (Microarray)
<i>B. subtilis</i> <i>rrnIHG</i> (inv)	LB	> 160 min	From study [3]	N/A	N/A	N/A (Microarray)
	MOPS glucose CA	44 ± 5 min	From study [3]	Digitized	Source Data: Bs_IHG_MM	N/A (Microarray)
	MOPS glucose MM	44 ± 1 min	From study [3]	Digitized	Source Data: Bs_IHG_CA	N/A (Microarray)
<i>E. coli</i> WT	LB	19.3 ± 1.7 min	From study [4]	ENA: PRJEB25595	Exp: ERS2298483 Stat: ERS2298484	1.4 × 10 ⁷ 1.3 × 10 ⁷
	M9 glucose	68.8 ± 6.2 min	From study [4]	ENA: PRJEB25595	Exp: ERS2298504 Stat: ERS2298505	1.7 × 10 ⁷ 1.4 × 10 ⁷

SUPPLEMENTARY METHODS

1. Growth media and determination of growth phase

As we have used data from multiple sources, the minimal media and the determination of population growth phase are not consistent across all studies. All studies use the standard recipe for Luria-Bertani (LB) rich medium

(1% tryptone, 0.5% yeast extract, and 1% NaCl in H₂O), with the exception of Midgley-Smith et al. (2018), where 0.05% NaCl is used instead. All studies using M9 minimal media have the same base recipe (1X M9 salts, 2 mM MgSO₄, and 0.1 mM CaCl₂), with different supplements and carbon sources. In our study and in Galli et al. (2019) [1], M9 was supplemented with 10 $\mu\text{g mL}^{-1}$ thiamine HCl for the *V. cholerae* strains. The carbon source for M9 is 0.4% glucose for our study, 0.4% fructose for Galli et al. [1], and 0.2% glucose for Midgley-Smith et al. [4]. In Wang et al. (2007) [2], the relevant datasets exclusively use S7 minimal medium (50 mM MOPS, 10 mM (NH₄)₂SO₄, 5 mM potassium phosphate, 2 mM MgCl₂, 0.9 mM CaCl₂, 50 μM MnCl₂, 5 μM FeCl₃, 10 μM ZnCl₂, and 2 μM thiamine hydrochloride), supplemented with 1% sodium fumarate as the carbon source, 0.1% glutamate, 40 $\mu\text{g mL}^{-1}$ tryptophan, and 40 $\mu\text{g mL}^{-1}$ phenylalanine. For Srivatsan et al. (2010) [3], the minimal medium consists of 50 mM MOPS with 1% glucose, along with different supplements for the CA (0.5% casamino acids) and MM (40 $\mu\text{g mL}^{-1}$ tryptophan, 40 $\mu\text{g mL}^{-1}$ methionine, 40 $\mu\text{g mL}^{-1}$ phenylalanine, and 100 $\mu\text{g mL}^{-1}$ arginine) growth media.

All populations were grown at 37 °C. In our study, exponential phase cultures were grown to OD₆₀₀ of 0.60-0.80. Stationary phase samples were grown to OD₆₀₀ of 1.50 or greater. In Galli et al. [1], exponential phase cultures were grown to an OD₆₅₀ of 0.05 and 0.2 in M9 and LB, respectively. In Midgley-Smith et al. [4], exponential phase cultures were grown to an OD₆₀₀ of 0.48. In Srivatsan et al. [3], exponential phase cultures were grown to an OD₆₀₀ of 0.2-0.6.

2. Generation of marker frequency data for this study

For detailed protocols of marker-frequency generation for each dataset, see the corresponding references in Sec. . In this study, strains were struck on solid agar and grown overnight at 37 °C. Three individual colonies were selected from each strain and used to inoculate 10 mL of LB media, which was grown overnight at 37 °C with 260 rpm shaking. The following morning, these overnight pre-cultures were back-diluted to OD₆₀₀ of 0.05 in 5 mL of either LB or M9 media. The same biological replicate was used to inoculate both growth conditions. Exponential phase cultures were grown to OD₆₀₀ of 0.60-0.80. Stationary phase samples were grown to OD₆₀₀ of 1.50 or greater. To harvest, cultures were centrifuged at $8,000 \times g$ for 5 mins at 25 °C, and the supernatant was removed. gDNA was prepared immediately following harvest using GeneJet Genomic DNA Purification Kit (Thermo Scientific, K0721). Purified gDNA was quantified using Qubit dsDNA HS Assay Kit (Invitrogen, Q32851) and quality was assayed using a Nanodrop 2000 Spectrophotometer (Thermo Scientific, ND-2000). Whole-genome libraries were prepared from purified gDNA using Twist 96-Plex Library Prep Kit (Twist Bioscience, 104950) with dual adapters. Libraries were subsequently pooled and sequenced on Illumina NovaSeq6000 (S4) to a depth of 6×10^7 reads per sample.

3. Marker frequency analysis

A. Sequence alignment

To align deep sequencing read outputs to the bacterial reference genomes for MFA, we used the read alignment tool Bowtie 2 (v2.4.5) via the MATLAB (R2022b) function `bowtie2` [5], which requires the Bowtie 2 Support Package from the MATLAB Bioinformatics Toolbox (v4.16.1). We then extracted the position information from the resulting SAM file using an in-lab written MATLAB script and the command-line tool SAMtools [6] (v1.16.1).

B. Data selection

Not all marker-frequency experiments are of the same quality and many datasets we have analyzed have clear signatures of systematic error. We use two key criteria to determine the quality of the data. The first is a qualitative measure involving the shape of the marker-frequency profile and the second is a quantitative measure based on resolution analysis.

Marker profile near origin. Our analysis uses exponential growth as the stop-watch for resolving dynamics, so it is essential that the population is harvested at the right time for sequencing. We have found that the marker-frequency profile has a signature flattening near the origin of replication if the population is harvested too late in exponential phase, as shown in Supplementary Figure 15. The population is entering stationary phase, where nutrient depletion begins to cause slower growth. Entering stationary phase also causes a decrease in the population-wide rate of replication initiation, which leads to the flattening of the profile near the origin. For the rest of the chromosome, replication continues unchanged for the majority of the population. Since exponential growth is crucial

to our analysis, we have chosen to only use datasets with cusp-like behavior near the origin(s), as opposed to a rounded concave-down shape.

Resolution analysis. In the case where there is a single origin of replication for each chromosome, we expect only a single maximum in the marker frequency profile, corresponding to the origin. In an ideal noiseless situation, any local maxima distinct from the origin would correspond to other points of replication initiation and the fit would predict a retrograde fork velocity. However, due to the stochastic nature of the sequencing data, there can be spurious local maxima if the resolution is chosen to be too high. Fluctuations due to noise can cause the knots in the piecewise-linear fit along each arm of the chromosome to be non-monotonic, as shown in Panel C of Fig. 1 in the main paper. Thus, given two datasets for the same organism and growth condition, data quality was determined based on the best resolution that can be achieved without introducing spurious maxima.

C. To divide or not to divide by stationary phase data

To further remove any sequencing induced bias, we divide the copy number data obtained during exponential growth by the data obtained during stationary phase. This normalization step is done with the 1 kb binned values. Since the division of two noisy uncorrelated data sets typically results in a decreased signal-to-noise ratio (SNR), we used a comparison of the variance divided by the mean before and after division by stationary phase.

We expect the relative variance to decrease by normalization if the reduction in sequencing bias resulting from division has a larger effect than the increase in noise from dividing two noisy data sets. We first divide the noisy stationary phase data by its expected mean value. This ensures that it is of unit order and will not introduce a global scale factor during division, which would affect only the normalized data and not the pre-normalization data. We find that after division, the variance is roughly halved, suggesting that the normalization by stationary phase is successful in reducing sequencing bias without resulting in excessive noise. We thus chose to divide all exponential phase data with the corresponding stationary phase data. For some datasets, the stationary phase results were not provided for some of the ectopic origin mutants. The only difference in these mutants from their original strains is the addition of a short segment of DNA for the ectopic origin, which does not significantly change the locations of sequencing bias. Therefore, we chose to divide the exponential phase data for the ectopic origin strains by the stationary phase data from the original strains.

D. Filtering the data

To remove outliers from the marker-frequency data, we used a centered 200 kb median filter, such that at each position ℓ along the chromosome, we set the median marker-frequency from $\ell - 100$ kb to $\ell + 100$ kb as a baseline. The data was given periodic boundary conditions such that the centered medians still had 200 kb regions at the boundaries of the data set. We then discarded the 10% of points that most deviated from this centered median baseline, sorted by the absolute value of the difference. We also removed data points from regions known to be associated with mobile elements of low GC content, such as the Super Integron on Chr2 of *V. cholerae*, which is subject to anomalous sequencing bias.

E. Nearest-neighbor variance estimator

We used a nearest-neighbor estimate of variance:

$$\hat{\sigma}^2 \approx \frac{1}{N} \sum_{i=1}^N \frac{1}{2} (x_{i+1} - x_i)^2, \quad (1)$$

where $x_{N+1} = x_1$, in agreement with periodic boundary conditions. This approach allows us to obtain a variance measurement without assuming the underlying distribution of the data and the corresponding expected value of each bin.

F. Fitting slopes to the log of the copy number

After taking the natural logarithm of the copy number, we expect the data to follow a roughly linear trend along each arm of the chromosome. Since our goal is to obtain higher resolution measurements of fork velocity along the

chromosome, it is necessary to subdivide each arm into smaller segments with variable slopes. The fit to the data should be continuous, since discontinuities would create copy number ambiguity at segment junctions. Therefore, we use a continuous piecewise linear least squares fit to obtain slope measurements from the data. We call the MATLAB function `lsqnonlin` as part of an in-lab written fitter function. The `lsqnonlin` function is part of the MATLAB Optimization Toolbox (v4.16.1). The fitter obtains vertical control point values at the junctions between segments. Each control point is used in the measurement of the slope both to the left and to the right of it.

G. Estimating errors for the control points

To obtain error estimates for the control points of the least squares fit, we use the Jacobian that `lsqnonlin` returns, which is the Jacobian of the difference between the fit data and the experimental data. If we have k fit parameters θ_α , where $\alpha = 1, \dots, k$, and N data points y_i , where $i = 1, \dots, N$, then the Jacobian takes the form:

$$J_{i\alpha} = \frac{\partial}{\partial \theta_\alpha} (y_i - \mu_i(\theta)), \quad (2)$$

where $\mu_i(\theta)$ is the fitter estimate of y_i , based on the model with the set of parameters θ . We use the matrix form of the Fisher information:

$$[I(\theta)]_{\alpha,\beta} = \mathbb{E} \left[\left(\frac{\partial}{\partial \theta_\alpha} \log f(Y; \theta) \right) \left(\frac{\partial}{\partial \theta_\beta} \log f(Y; \theta) \right) \middle| \theta \right], \quad (3)$$

where $f(Y; \theta)$ is the probability density function (PDF) of the random variable Y , given parameters θ . In our case, we expect the noise to be Gaussian distributed around the mean, so we have the PDF:

$$f(Y; \theta) = \frac{1}{(\sigma\sqrt{2\pi})^N} \prod_{i=1}^N \exp \left(-\frac{(y_i - \mu_i(\theta))^2}{2\sigma^2} \right), \quad (4)$$

where σ is the standard deviation. Now we take the derivative of the logarithm:

$$\frac{\partial}{\partial \theta_\alpha} \log f(Y; \theta) = \frac{\partial}{\partial \theta_\alpha} \left(-\frac{\sum_i (y_i - \mu_i(\theta))^2}{2\sigma^2} - N \log(\sigma\sqrt{2\pi}) \right), \quad (5)$$

$$= -\frac{1}{\sigma^2} \sum_{i=1}^N (y_i - \mu_i(\theta)) \frac{\partial}{\partial \theta_\alpha} (y_i - \mu_i(\theta)), \quad (6)$$

$$= -\frac{1}{\sigma^2} \sum_{i=1}^N (y_i - \mu_i(\theta)) J_{i\alpha}. \quad (7)$$

Plugging this into Supplementary Equation (3) and noting that the matrix multiplication contracts over the i index, we are left with:

$$I(\theta) = \frac{J^T J}{\sigma^2}. \quad (8)$$

The Cramér-Rao bound states that the inverse of the Fisher information provides a lower bound for the covariance matrix of unbiased estimators. More explicitly, the relation is as follows:

$$\text{cov}_\theta(\hat{\theta}) \geq I(\theta)^{-1}. \quad (9)$$

We take the equality for our error estimates. The variances of the parameter estimates are given by the diagonal elements of the covariance matrix.

H. Transformation matrices for obtaining the fork velocity

To convert the control points of the piecewise linear fit into slopes and fork velocities, we use coordinate transformation matrices, which have the added benefit of also transforming the variance estimates from the Fisher information. The coordinate transformation matrices can either be obtained through direct reasoning or as Jacobians of the

final coordinates relative to the initial coordinates. The following examples will be square matrices of dimension 3, acting on sets of 3 parameters. These examples illustrate the procedure of obtaining the transformation matrices and are easily generalizable. To convert from a vector of control point values to a vector where the first element is the leftmost control point value and the other elements are the slopes, we use the following matrix:

$$\text{CtrlPts2Slopes} = \begin{pmatrix} 1 & 0 & 0 \\ -\Delta_{12}^{-1} & \Delta_{12}^{-1} & 0 \\ 0 & -\Delta_{23}^{-1} & \Delta_{23}^{-1} \end{pmatrix}, \quad (10)$$

where Δ_{ij} is the chromosomal position (horizontal value) of the j -th control point minus the position of the i -th control point. We can thus relate the vector of slopes $\vec{\alpha}$ and the vector of (vertical) control point values $\vec{\theta}$:

$$\vec{\alpha} = \text{CtrlPts2Slopes} \cdot \vec{\theta} = \begin{pmatrix} \theta_1 \\ (\theta_2 - \theta_1)/\Delta_{12} \\ (\theta_3 - \theta_2)/\Delta_{23} \end{pmatrix} \quad (11)$$

From the slopes, we can obtain the fork velocities with the following matrix:

$$\text{Slopes2Vels} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -k_G/\alpha_2^2 & 0 \\ 0 & 0 & -k_G/\alpha_3^2 \end{pmatrix}. \quad (12)$$

The velocities \vec{v} are related to $\vec{\alpha}$ like so:

$$\vec{v} = \text{Slopes2Vels} \cdot \vec{\alpha} = \begin{pmatrix} \alpha_1 \\ -k_G/\alpha_2 \\ -k_G/\alpha_3 \end{pmatrix}. \quad (13)$$

To properly transform the covariance matrix obtained in Sec. , we must use the transformation matrix on both sides, like so:

$$\text{cov}_\alpha = \text{CtrlPts2Slopes}^T \cdot \text{cov}_\theta \cdot \text{CtrlPts2Slopes}, \quad (14)$$

$$\text{cov}_v = \text{Slopes2Vels}^T \cdot \text{cov}_\alpha \cdot \text{Slopes2Vels}. \quad (15)$$

The error estimates for each parameter are the square roots of the corresponding diagonal elements.

I. Behavior near the terminus

Due to the stochastic nature of replisome progression, replication forks do not always meet at the terminus, which corresponds to the *dif* site [1]. If one fork arrives first, it can continue past the *dif* site, converting from antegrade to retrograde motion along the other arm. While there are Tus-Ter traps in *E. coli* and *B. subtilis* to limit the amount of retrograde motion, they are not 100% efficient and they are not present in *V. cholerae* [1]. Thus, in a population, the true termination points along the chromosome would form a distribution around the *dif* site. This means that in regions near the terminus, there can be antegrade and retrograde replication forks both contributing to the marker frequency, which contradicts one of the assumptions required for our analysis: replication proceeding only in one direction. Such retrograde motion would cause a flattening out around any fork convergence points observed in the marker-frequency data. Therefore, when doing a multi-segment analysis, we remove the two fork velocity measurements on either side of the marker-frequency minimum.

J. Selection of the number of fitted knots

First we bin the data into 1 kb regions, following the convention of previous marker frequency analysis studies. To determine the number of segments that would best fit the data, we use the Akaike Information Criterion (AIC). We compare a series of nested models with 2^k continuous piecewise-linear least squares fits, where $k = 1, \dots, 10$. The AIC-optimal model for fast growth (in LB) had 39 knots, spaced by 100 kb, generating 38 measurements of locus velocity across the two chromosomes of WT *V. cholerae*. Other strains either have similar or smaller region sizes that minimized the AIC value, so for ease of comparison with other mutant strains, we take 100 kb to be the step size for determining fork velocity.

4. Bilateral symmetry analysis

To analyze whether an observed velocity profile was consistent with the predictions of the time-dependent mechanism, we test for bilateral symmetry between the left and right arms:

$$v(\ell) = v(-\ell), \quad (16)$$

where ℓ is the locus position relative to the origin.

Assume the fork velocity has been computed over $2m$ equal-length genome segments, arranged symmetrically about the origin ($\ell = 0$). The segments $i = -1 \dots -m$ correspond to sequentially labeled segments along the left arm and the segments $i = 1 \dots m$ correspond to sequentially labeled segments along the right arm such that $\ell = \Delta\ell i$ corresponds to the end point of each segment where $\Delta\ell$ is the segment length. Let the mean fork velocity be defined:

$$\bar{v} \equiv \frac{1}{2m} \sum_{i=1}^m v_i + v_{-i}, \quad (17)$$

where v_i is the velocity over segment i .

To divide the variance into symmetric and antisymmetric contributions, we define symmetrized, (i) , and antisymmetrized velocities, $[i]$, for index pairs $\pm i$:

$$\delta v_{(i)} \equiv \frac{1}{2}(v_i + v_{-i} - 2\bar{v}), \quad (18)$$

$$\delta v_{[i]} \equiv \frac{1}{2}(v_i - v_{-i}). \quad (19)$$

We define the symmetric and antisymmetric variances

$$\sigma_S^2 \equiv \frac{1}{m} \sum_{i=1}^m \delta v_{(i)}^2 \quad (20)$$

$$\sigma_A^2 \equiv \frac{1}{m} \sum_{i=1}^m \delta v_{[i]}^2, \quad (21)$$

and the total variance is the sum of the symmetric and antisymmetric variances:

$$\sigma^2 \equiv \frac{1}{2m} \sum_{i=1}^m (v_i - \bar{v})^2 + (v_{-i} - \bar{v})^2 \quad (22)$$

$$= \sigma_S^2 + \sigma_A^2. \quad (23)$$

Finally, we define the symmetric fraction of the variance:

$$f_S \equiv \sigma_S^2 / \sigma^2. \quad (24)$$

If the variation in the velocity obeys the bilateral symmetry (Supplementary Equation (16)), the variance is all symmetric (*i.e.* $f_S = 1$). If on the other hand, the variation is randomly distributed along the genome, we expect equal symmetric and antisymmetric combinations (*i.e.* $f_S = 1/2$).

5. Estimation of average fork number per cell cycle

The following analysis is only an estimation. There are a few key assumptions that are inconsistent with cell phenomenology, but are kept to make the analysis tractable. In particular, we take the assumption that neither arm of the chromosome has fork arrest, which is inconsistent with the results for *B. subtilis*. This assumption allows us to obtain an estimate without prior knowledge of where fork arrest may occur. We also assume that the D period (time between end of chromosomal replication and cell division) is negligible, which is inconsistent when growth is extremely slow. This assumption allows us to use the copy number of the terminus as an approximation of the number of cells, which cannot be determined from MFA.

The population number density of forks $n_f(\ell)$ at any position ℓ is given by the change in copy number between two consecutive positions along the chromosome:

$$n_f(\ell) = -\frac{d}{d\ell} N(\ell). \quad (25)$$

This equation is a result of the following reasoning: If there are two copies at $\ell = x$ and one copy at $\ell = x + 1$, then there must be a fork that has just replicated $\ell = x$. We need to integrate the number density over the entire chromosome to get the total number of forks for the population:

$$N_{fork, population} = 2 \int_{ori}^{ter} n_f(\ell) d\ell = 2 \int_{ori}^{ter} -\frac{d}{d\ell} N(\ell) d\ell = 2(N_{ori} - N_{ter}), \quad (26)$$

where the factor of 2 came from having two arms. To get the average number of forks per cell \bar{N}_f , we divide by the total number of cells, which is roughly equal to the number of termini N_{ter} :

$$\bar{N}_f = 2 \frac{N_{ori} - N_{ter}}{N_{ter}}. \quad (27)$$

This can be related to lag time:

$$\tau_\ell = -k_G^{-1} \ln \frac{N_\ell}{N_{ori}} \implies \frac{N_{ori}}{N_{ter}} = \exp(k_G \tau_{ter}) = e^{\tau_{ter} \ln 2 / T} = (e^{\ln 2})^{\tau_{ter} / T} = 2^{\tau_{ter} / T}, \quad (28)$$

which can be substituted into the equation above to get:

$$\bar{N}_f = 2 * (2^{\tau_{ter} / T} - 1). \quad (29)$$

This is closely related to the result derived in [7]. For two chromosomes, we add the number of forks for both, but divide everything just by the minimum $N_{ter,i}$, which corresponds to the terminus of Chr1 in this case:

$$\bar{N}_{f,2chr} = 2 \frac{(N_{ori,1} - N_{ter,1})}{N_{ter,1}} + 2 \frac{(N_{ori,2} - N_{ter,2})}{N_{ter,1}}. \quad (30)$$

In general, when there are multiple chromosomes:

$$\bar{N}_{fork, multichr} = \frac{2}{\min(N_{ter,i})} \sum_i (N_{ori,i} - N_{ter,i}), \quad (31)$$

where i denotes which chromosome. To get a relative lag time, we use:

$$\tau_{\ell \rightarrow m} = \tau_m - \tau_\ell = -k_G^{-1} \left(\ln \frac{N_m}{N_{ori}} - \ln \frac{N_\ell}{N_{ori}} \right) = k_G^{-1} \ln \frac{N_\ell}{N_m}. \quad (32)$$

Letting the subscript *end* denote the lag time corresponding to $\min(N_{ter,i})$. We thus have:

$$\bar{N}_{fork, multichr} = 2 \sum_i \left(2^{(\tau_{end} - \tau_{ori,i}) / T} - 2^{(\tau_{end} - \tau_{ter,i}) / T} \right), \quad (33)$$

where τ_{end} is the maximum lag time measured. This can be simplified to:

$$\bar{N}_{fork, multichr} = 2 \times 2^{\tau_{end} / T} \times \sum_i \left(2^{-\tau_{ori,i} / T} - 2^{-\tau_{ter,i} / T} \right), \quad (34)$$

which enables a calculation of average fork number per cell cycle using lag times.

6. Statistically significant deviations of local fork velocity from the global mean

To test the statistical significance of the measured velocities, we use a null hypothesis test. The null hypothesis corresponds to a model with uniform fork velocity on the right and left arms. The alternative hypothesis corresponds to a model with a model-selected number of knots (m), corresponding to $m - 1$ genomic region-specific velocities.

To calculate the p-value, we begin by obtaining the χ^2 -test statistic:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{\sigma_i^2} \quad (35)$$

where O_i denotes the observed values, E_i denotes the expected values, and σ_i^2 is the variance of the fork velocities measured using the fitter function (as described in Supplementary Method 3H). In this case, E_i is the global mean fork velocity for all i , since that is the null hypothesis. Since there is only one degree of freedom for the model, we use the one-dimensional χ^2 cumulative distribution function to determine the probability of measuring a χ^2 statistic as extreme as the one we've measured, assuming the null hypothesis is true. This p-value is found to be $\ll 10^{-30}$ for all datasets except one with a p-value of 6×10^{-12} . All of the p-values are much smaller than the standard threshold of 0.05, so we reject the null hypothesis. We thus conclude that there are statistically significant variations of the fork velocity relative to the mean.

SUPPLEMENTARY NOTES

1. Stochastic simulations method

The purpose of the stochastic simulation was *to investigate the role of stochasticity in determining the log-phase growth demography of the population*. It was *not* to generate a mechanistically realistic and detailed model of the cell cycle.

To simulate the cell cycle, we performed an exact stochastic simulation using the Gillespie Algorithm [8]. We idealized the cell cycle as follows: **Genome representation:** The genome was divided into two equal length arms each consisting to 100 course-grained bases. An explicit realization of all 5 Mb would be too slow to rapidly explore different cell models. Finer coarse-grained basepairs were explored but made no difference to the results. **Replication initiation:** We experimented with a number of different models for initiation since we initially believed that the stochasticity of initiation would limit the ability to quantitate the fork velocity. (i) We initially investigate a *Terminus Model* in which there was a constant rate of initiation k_{init} after the termination of the on-going replication. (ii) In order to simulate a more realistic model, we then considered an *Origin Model* where there was constant rate k_{init} of initiation at each origin. (iii) We explored a *Precise Model* with precise timed B period length T_B . (iv) We explored a *Uniform Model* with a uniform distribution of B periods. **Fork velocity:** We used the replication of the course-grained bases had exponentially distributed wait times $k_{\text{rep}} = v/\Delta\ell$ (i.e. constant rate per unit time) where v is the replication velocity and $\Delta\ell$ is the length of the course-grained bases. We note that this assumption makes replication *more stochastic* than a more realistic model in which wait times are exponentially distributed at the single-base level. We experimented with a number of different types of locus-dependent fork velocities. In addition, we added exponentially-distributed pauses of various durations at a specific locus. **Termination:** We only allowed a single direction of fork propagation. When the fork reached the end of the arm, replication terminated. **Cell division:** We assumed cells divided immediately after termination of the slowest arm of the chromosome. **Initiation of the simulation:** All simulations initiated from a single new-born cell with an un-replicated chromosome. **Stop conditions:** Simulation were run until there were 10^5 cells. **Determination of growth rate.** The number of cells was fit to an exponential to determine the growth rate k_G between $N_{\text{cell}}(t) = 10^4$ and $N_{\text{cell}}(t) = 10^5$ cells. **Generation of simulated marker frequency.** The marker frequency was defined as the number of each genetic locus in the population at the termination of the simulation. In summary, the model was described by the parameters k_{init} and the fork velocity $v(\ell)$ (or pause time distribution) at each locus. The model output was the marker frequency at each locus $N(\ell)$ and the growth rate k_G . See Supplementary Figure 16.

2. Stochastic simulations match the predictions of the log-slope

To test whether the log-slope law applies locally, irrespective of stochasticity in cell-cycle timing, we used a stochastic simulation to generate simulated marker-frequency data. The basic strategy was to simulate a stochastically timed cell cycle, including stochastically timed B periods as well as stochastic fork dynamics, and compare the observed population demographics to the prediction of non-stochastic models.

Log slope. We define the log slope:

$$\alpha(\ell) \equiv \frac{d}{d\ell} \log N(\ell), \quad (36)$$

where $N(\ell)$ is the population copy number of the locus at position ℓ . Theoretically, the slope is predicted to be:

$$\alpha(\ell) = \frac{k_G}{v(\ell)}, \quad (37)$$

where k_G is the growth rate and $v(\ell)$ is the locus-dependent fork velocity.

Log difference. For long pauses, the dynamics are characterized by a step rather than a slope. The step is defined:

$$\Delta(\ell) \equiv \lim_{\delta \rightarrow 0^+} \log N(\ell + \delta) - \log N(\ell), \quad (38)$$

where $N(\ell)$ is the population copy number of the locus at position ℓ . Theoretically, the log difference is predicted to be:

$$\Delta(\ell) \equiv k_G \delta \tau(\ell) = \log\left(1 + \frac{k_G}{k}\right), \quad (39)$$

where k_G is the growth rate, $\delta \tau(\ell)$ is the exponential mean of the step wait time and the last equality applies in the special case of an exponentially-distributed wait time with rate k .

3. Simulation models

The model descriptions and motivations are summarized below:

Terminus Model 1: In the terminus model, the replication initiation rate is proportional to the terminus number (per cell). In this model, there is a locus-independent fork velocity.

Terminus Model 2: Same as above, but with two fork velocities. On the last half of the right arm (region 1), the fork velocity is higher than the rest of the chromosome (region 0).

Terminus Model 3: Same as above, but with an exponentially-distributed pause close to the origin on the right arm.

Terminus Model 4: Same as Terminus 1, but with a lower initiation rate.

Origin Model: In the origin model, the replication initiation rate is proportional to the origin number.

Uniform Model: The timing between initiations is uniformly distributed. This model was included to explore whether the observations were a special case of exponentially-distributed wait times.

Precise Model: The timing between initiations is uniformly distributed. This model was included to explore whether the observations were a special case of exponentially-distributed wait times.

4. Simulation results

The results from the simulations are summarized in the table below. All numbers have at least precision of 1% and are in simulation units: simulation bases (sb) and simulation time (st).

Model	Parameter Initiation	Parameter Fork dynamics	Observed Growth rate: k_G	Observed Log slope/difference	Predicted Log slope/difference
Terminus 1	$k_{\text{init}} = 1 \text{ st}^{-1}$	$v_0 = 20 \text{ sb st}^{-1}$	$2.6 \times 10^{-1} \text{ st}^{-1}$	$\alpha = 9.0 \times 10^{-3} \text{ sb}^{-1}$	$\alpha = 9.0 \times 10^{-3} \text{ sb}^{-1}$
Terminus 2	$k_{\text{init}} = 1 \text{ st}^{-1}$	$v_0 = 20 \text{ sb st}^{-1}$ $v_1 = 30 \text{ sb st}^{-1}$	$2.7 \times 10^{-1} \text{ st}^{-1}$	$\alpha_0 = 1.3 \times 10^{-2} \text{ sb}^{-1}$ $\alpha_1 = 8.8 \times 10^{-3} \text{ sb}^{-1}$	$\alpha_0 = 1.3 \times 10^{-2} \text{ sb}^{-1}$ $\alpha_1 = 8.8 \times 10^{-3} \text{ sb}^{-1}$
Terminus 3	$k_{\text{init}} = 1 \text{ st}^{-1}$	$v_0 = 20 \text{ sb st}^{-1}$ $v_1 = 30 \text{ sb st}^{-1}$ $k_3 = 2 \text{ st}^{-1}$	$2.6 \times 10^{-1} \text{ st}^{-1}$	$\alpha_0 = 1.3 \times 10^{-2} \text{ sb}^{-1}$ $\alpha_1 = 8.8 \times 10^{-3} \text{ sb}^{-1}$ $\Delta = 1.3 \times 10^{-1}$	$\alpha_0 = 1.3 \times 10^{-2} \text{ sb}^{-1}$ $\alpha_1 = 8.8 \times 10^{-3} \text{ sb}^{-1}$ $\Delta = 1.3 \times 10^{-1}$
Terminus 4	$k_{\text{init}} = 0.3 \text{ st}^{-1}$	$v_0 = 20 \text{ sb st}^{-1}$	$1.4 \times 10^{-1} \text{ st}^{-1}$	$\alpha_0 = 7.3 \times 10^{-3} \text{ sb}^{-1}$	$\alpha_0 = 7.3 \times 10^{-3} \text{ sb}^{-1}$
Origin	$k_{\text{init}} = 0.3 \text{ st}^{-1}$	$v_0 = 20 \text{ sb st}^{-1}$	$3.0 \times 10^{-1} \text{ st}^{-1}$	$\alpha = 1.5 \times 10^{-2} \text{ sb}^{-1}$	$\alpha = 1.5 \times 10^{-2} \text{ sb}^{-1}$
Uniform	$T \sim U([0, 6]) \text{ st}$	$v_0 = 20 \text{ sb st}^{-1}$	$2.7 \times 10^{-1} \text{ st}^{-1}$	$\alpha = 1.3 \times 10^{-2} \text{ sb}^{-1}$	$\alpha = 1.3 \times 10^{-2} \text{ sb}^{-1}$
Precise	$T = 0.2 \text{ st}$	$v_0 = 20 \text{ sb st}^{-1}$	$1.4 \times 10^{-1} \text{ st}^{-1}$	$\alpha = 6.8 \times 10^{-3} \text{ sb}^{-1}$	$\alpha = 6.8 \times 10^{-3} \text{ sb}^{-1}$

5. Re-scaling simulation units to compare to measured data

The simulations were performed in simulation units. We call the coarse-grained bases simulation bases (sb) and time units simulation time (st). To compare these simulations to experimental data, the prediction need to be rescaled

to place the observations in biological units. To model a 5 Mb bacterial genome, with typical fork velocity of 1 kb s^{-1} , the conversion factors are:

$$\frac{5 \times 10^6 \text{ bp}}{200 \text{ sb}} = 2.5 \times 10^4 \text{ bp sb}^{-1}, \quad (40)$$

$$\left(\frac{1 \text{ s}}{10^3 \text{ bp}} \right) \left(\frac{20 \text{ sb}}{1 \text{ st}} \right) \left(\frac{5 \times 10^6 \text{ bp}}{200 \text{ sb}} \right) = 5 \times 10^2 \text{ s st}^{-1}, \quad (41)$$

since we simulated typical fork velocities of 20 sb st^{-1} which we rescale the units to be equivalent to 1 kb s^{-1} .

Aside: Note that due to the smaller number of sb relative to bp, the stochasticity of fork dynamics should be exaggerated in the simulations, strengthening our argument that the stochasticity does not effect the model predictions.

6. Movies of marker frequency dynamics approaching steady state growth

We include two examples of movies tracking the marker frequency dynamics from a single cell to steady state growth. Supplementary Movies 1 and 2 show the dynamics of the Terminus 1 and Terminus 3 models respectively. Movie time is shown in simulation units. The dynamics emphasizes the importance of performing the calculations of the marker frequency in the steady-state limit. Not taking this limit correctly leads to anomalous results (*e.g.* [9]).

SUPPLEMENTARY FIGURES

1. Comparison with GC content

One potential rate-limiting factor for replication is the distribution of GC content across the genome. However, at the current genomic resolution of 100 kb, we find no significant evidence of GC content determining the rate of replication. See Supplementary Figure 1 for the results of wild-type *E. coli* strain MG1655 in LB and M9.

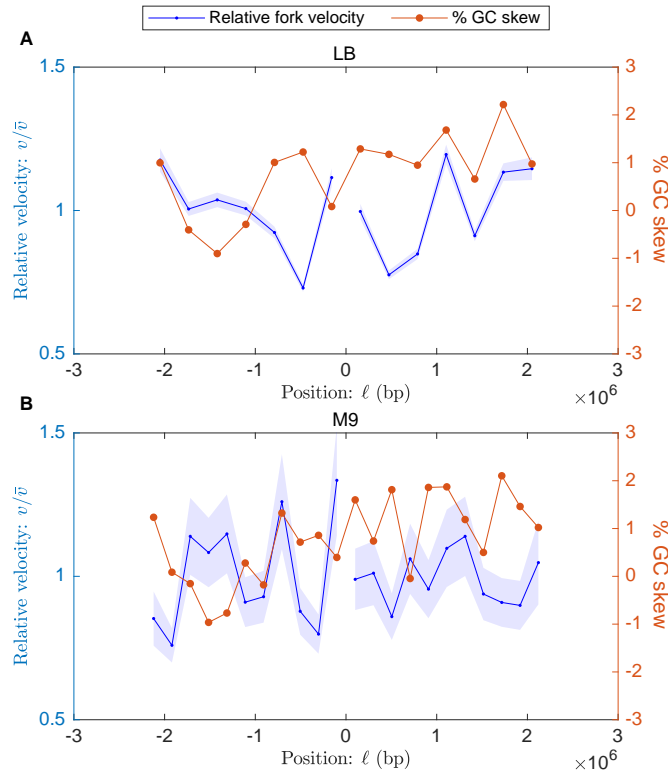
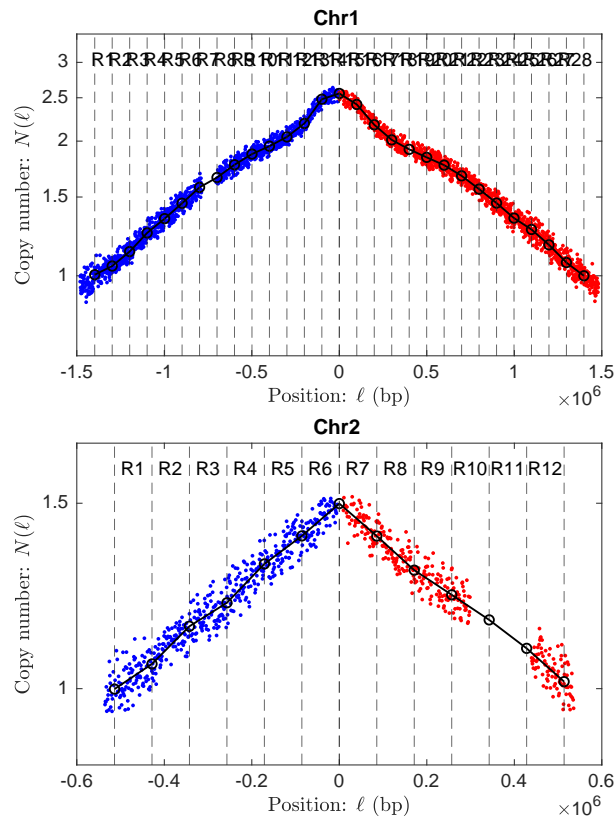


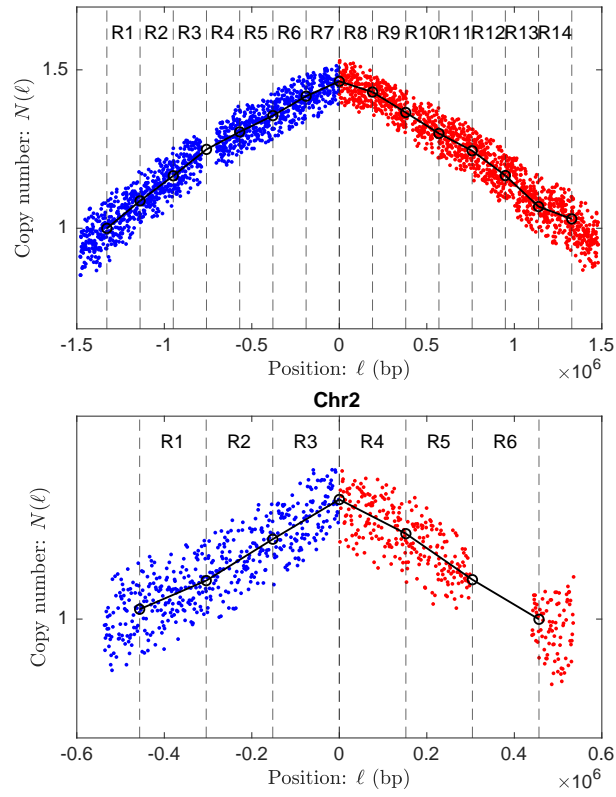
FIG. 1. **Relative fork velocity and % GC skew as a function of position.** Results for wild-type *E. coli* strain MG1655. The blue curve represents the fork velocity divided by the mean fork velocity, with shaded error bars. The orange curve represents the % GC skew. Data are presented as mean values \pm standard error of the mean (SEM). **Panel A:** Growth in LB medium. **Panel B:** Growth in M9 glucose minimal medium.

2. *V. cholerae* WT on LB



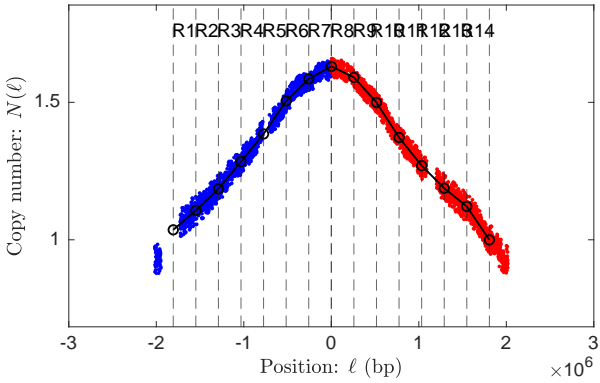
Detailed results tabulated in Supplementary Data.xlsx.

3. *V. cholerae* WT on M9 fructose



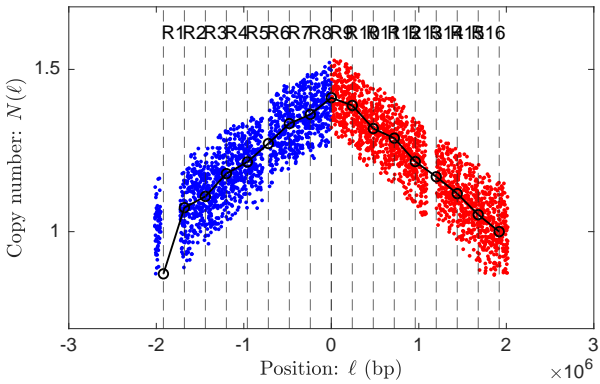
Detailed results tabulated in Supplementary Data.xlsx.

4. *V. cholerae* MCH1 on LB

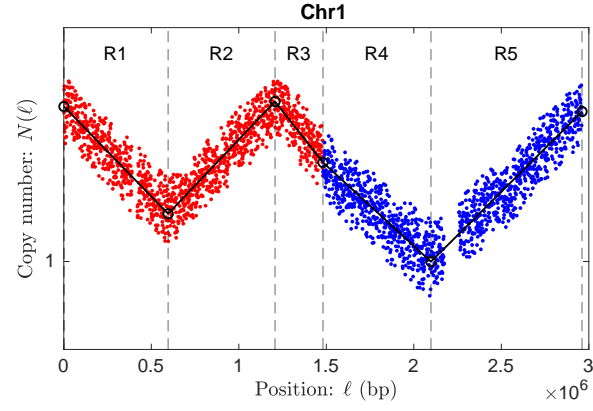


Detailed results tabulated in Supplementary Data.xlsx.

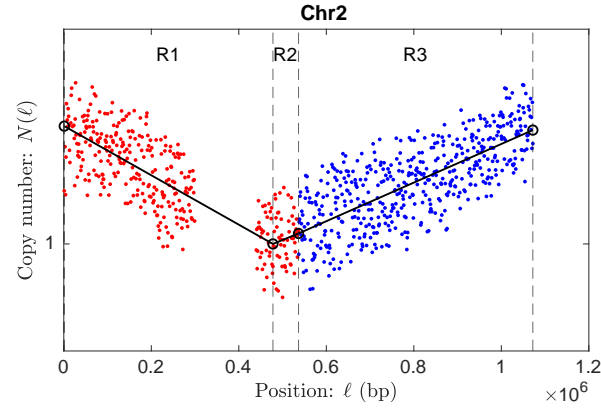
5. *V. cholerae* MCH1 on M9 fructose



Detailed results tabulated in Supplementary Data.xlsx.

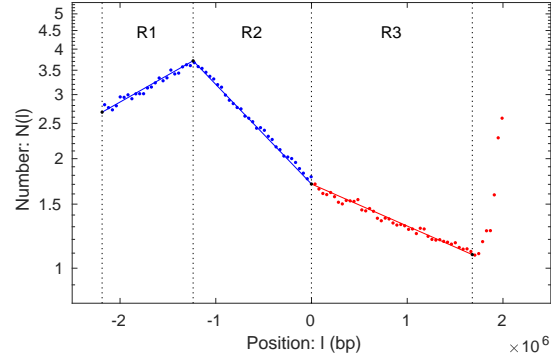
6. *V. cholerae* oriR4 on M9 fructose

Region name	Region start position: ℓ_- (Mb)	Region end position: ℓ_+ (Mb)	Fork direction	Replication orientation	Slope: α (Mb $^{-1}$)	Velocity: v (kb s $^{-1}$)	Log difference: Δ	Duration: $\Delta\tau$ (min)	Percent error
R1	0	0.596	+	A	0.308	0.750	0.183	13.2	$\pm 2.0\%$
R2	0.596	1.21	-	R	0.314	0.736	0.192	13.8	$\pm 1.7\%$
R3	1.21	1.48	+	A	0.380	0.608	0.104	7.51	$\pm 3.3\%$
R4	1.48	2.10	+	R	0.275	0.840	0.170	12.2	$\pm 1.9\%$
R5	2.10	2.96	-	A	0.297	0.777	0.256	18.5	$\pm 1.3\%$



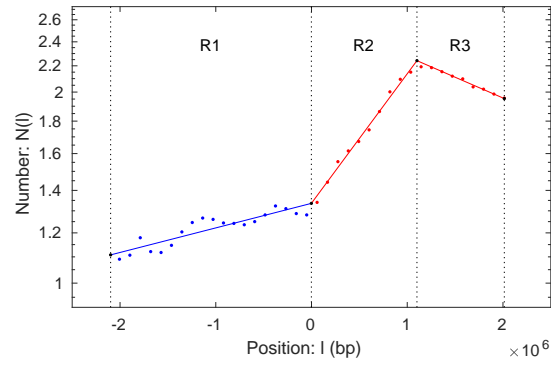
Region name	Region start position: ℓ_- (Mb)	Region end position: ℓ_+ (Mb)	Fork direction	Replication orientation	Slope: α (Mb $^{-1}$)	Velocity: v (kb s $^{-1}$)	Log difference: Δ	Duration: $\Delta\tau$ (min)	Percent error
R1	0	0.478	+	A	0.257	0.898	0.123	8.85	$\pm 5.0\%$
R2	0.478	0.537	-	R	0.186	1.24	0.0109	0.787	$\pm 52\%$
R3	0.537	1.07	-	A	0.201	1.15	0.108	7.76	$\pm 4.4\%$

7. *B. subtilis* oriC-257° on S7 fumarate



Region name	Region start position: ℓ_- (Mb)	Region end position: ℓ_+ (Mb)	Fork direction	Replication orientation	Slope: α (Mb $^{-1}$)	Velocity: v (kb s $^{-1}$)	Log difference: Δ	Duration: $\Delta\tau$ (min)
R1	-2.19	-1.24	-	A	$0.342 \pm 4.3\%$	NA	$0.325 \pm 4.3\%$	NA
R2	-1.24	-8.72×10^{-4}	+	R	$0.631 \pm 1.3\%$	NA	$0.779 \pm 1.3\%$	NA
R3	-8.72×10^{-4}	1.68	+	R	$0.266 \pm 2.6\%$	NA	$0.447 \pm 2.6\%$	NA

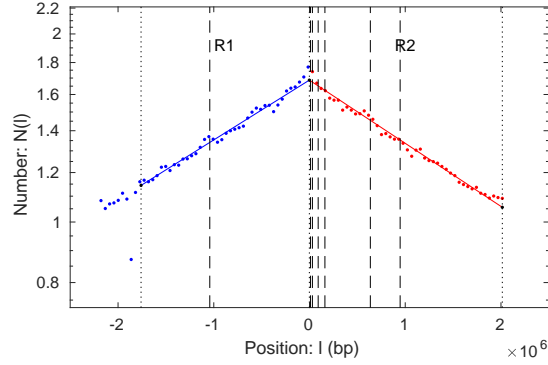
8. *B. subtilis* oriC-94° on S7 fumarate



Region name	Region start position: ℓ_- (Mb)	Region end position: ℓ_+ (Mb)	Fork direction	Replication orientation	Slope: α (Mb $^{-1}$)	Velocity: v (kb s $^{-1}$)	Log difference: Δ	Duration: $\Delta\tau$ (min)
R1	-2.10	-1.44×10^{-3}	-	A	$8.91 \times 10^{-2} \pm 10\%$	NA	$0.187 \pm 10\%$	NA
R2	-1.44×10^{-3}	1.10	-	R	$0.469 \pm 3.5\%$	NA	$0.518 \pm 3.5\%$	NA
R3	1.10	2.01	+	A	$0.151 \pm 18\%$	NA	$0.138 \pm 18\%$	NA

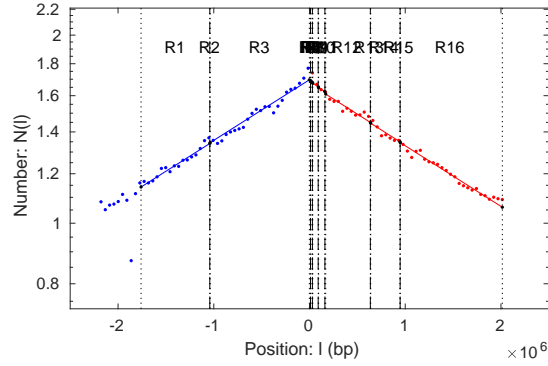
9. *B. subtilis* *oriN* on S7 fumarate

A. Two-slopes model



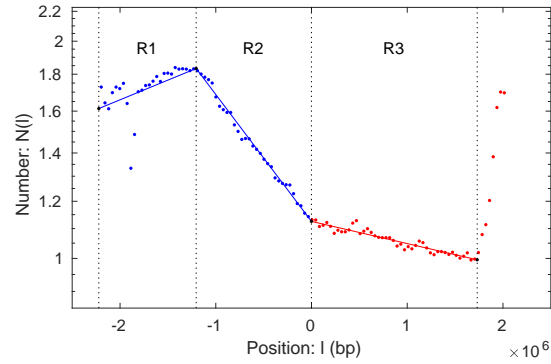
Region name	Region start position: ℓ_- (Mb)	Region end position: ℓ_+ (Mb)	Fork direction	Replication orientation	Slope: α (Mb $^{-1}$)	Velocity: v (kb s $^{-1}$)	Log difference: Δ	Duration: $\Delta\tau$ (min)
R1	-1.76	-1.59×10^{-3}	-	A	$0.221 \pm 2.3\%$	$0.948 \pm 2.3\%$	$0.388 \pm 2.3\%$	$30.9 \pm 2.3\%$
R2	-1.59×10^{-3}	2.01	+	A	$0.232 \pm 1.8\%$	$0.900 \pm 1.8\%$	$0.469 \pm 1.8\%$	$37.3 \pm 1.8\%$

B. Pause model



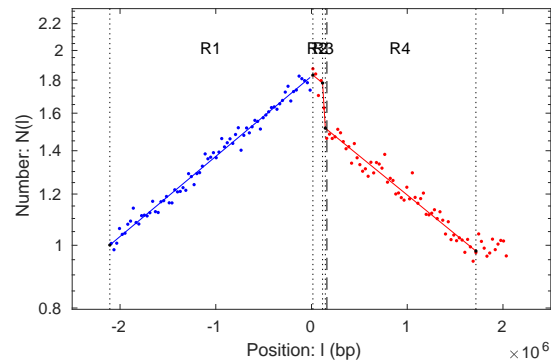
Region name	Region start position: ℓ_- (Mb)	Region end position: ℓ_+ (Mb)	Fork direction	Replication orientation	Slope: α (Mb $^{-1}$)	Velocity: v (kb s $^{-1}$)	Log difference: Δ	Duration: $\Delta\tau$ (min)
R1	-1.76	-1.04	-	A	$0.223 \pm 1.9\%$	$0.937 \pm 1.9\%$	$0.160 \pm 1.9\%$	$12.8 \pm 1.9\%$
R2	-1.04	-1.04	-	A	$0.973 \pm 21\%$	$0.215 \pm 21\%$	$3.13 \times 10^{-3} \pm 21\%$	$0.250 \pm 21\%$
R3	-1.04	-1.59×10^{-3}	-	A	$0.223 \pm 1.9\%$	$0.937 \pm 1.9\%$	$0.232 \pm 1.9\%$	$18.4 \pm 1.9\%$
R4	-1.59×10^{-3}	1.13×10^{-2}	+	R	$0.223 \pm 1.9\%$	$0.937 \pm 1.9\%$	$2.88 \times 10^{-3} \pm 1.9\%$	$0.229 \pm 1.9\%$
R5	1.13×10^{-2}	1.45×10^{-2}	+	A	$0.973 \pm 21\%$	$0.215 \pm 21\%$	$3.13 \times 10^{-3} \pm 21\%$	$0.250 \pm 21\%$
R6	1.45×10^{-2}	3.06×10^{-2}	+	A	$0.223 \pm 1.9\%$	$0.937 \pm 1.9\%$	$3.59 \times 10^{-3} \pm 1.9\%$	$0.286 \pm 1.9\%$
R7	3.06×10^{-2}	3.38×10^{-2}	+	A	$0.973 \pm 21\%$	$0.215 \pm 21\%$	$3.13 \times 10^{-3} \pm 21\%$	$0.250 \pm 21\%$
R8	3.38×10^{-2}	9.18×10^{-2}	+	A	$0.223 \pm 1.9\%$	$0.937 \pm 1.9\%$	$1.29 \times 10^{-2} \pm 1.9\%$	$1.03 \pm 1.9\%$
R9	9.18×10^{-2}	9.50×10^{-2}	+	A	$0.973 \pm 21\%$	$0.215 \pm 21\%$	$3.13 \times 10^{-3} \pm 21\%$	$0.250 \pm 21\%$
R10	9.50×10^{-2}	0.159	+	A	$0.223 \pm 1.9\%$	$0.937 \pm 1.9\%$	$1.44 \times 10^{-2} \pm 1.9\%$	$1.15 \pm 1.9\%$
R11	0.159	0.169	+	A	$0.973 \pm 21\%$	$0.215 \pm 21\%$	$9.40 \times 10^{-3} \pm 21\%$	$0.749 \pm 21\%$
R12	0.169	0.636	+	A	$0.223 \pm 1.9\%$	$0.937 \pm 1.9\%$	$0.104 \pm 1.9\%$	$8.30 \pm 1.9\%$
R13	0.636	0.639	+	A	$0.973 \pm 21\%$	$0.215 \pm 21\%$	$3.13 \times 10^{-3} \pm 21\%$	$0.250 \pm 21\%$
R14	0.639	0.945	+	A	$0.223 \pm 1.9\%$	$0.937 \pm 1.9\%$	$6.83 \times 10^{-2} \pm 1.9\%$	$5.44 \pm 1.9\%$
R15	0.945	0.948	+	A	$0.973 \pm 21\%$	$0.215 \pm 21\%$	$3.13 \times 10^{-3} \pm 21\%$	$0.250 \pm 21\%$
R16	0.948	2.01	+	A	$0.223 \pm 1.9\%$	$0.937 \pm 1.9\%$	$0.238 \pm 1.9\%$	$19.0 \pm 1.9\%$

10. *B. subtilis* oriN-257° on S7 fumarate



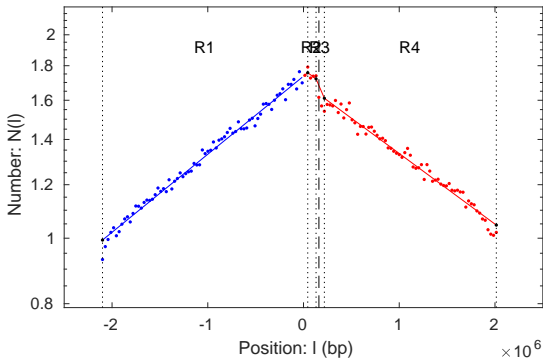
Region name	Region start position: ℓ_- (Mb)	Region end position: ℓ_+ (Mb)	Fork direction	Replication orientation	Slope: α (Mb $^{-1}$)	Velocity: v (kb s $^{-1}$)	Log difference: Δ	Duration: $\Delta\tau$ (min)
R1	-2.22	-1.20	-	A	$0.126 \pm 18\%$	NA	$0.128 \pm 18\%$	NA
R2	-1.20	-9.18×10^{-4}	+	R	$0.405 \pm 3.5\%$	NA	$0.488 \pm 3.5\%$	NA
R3	-9.18×10^{-4}	1.73	+	A	$7.08 \times 10^{-2} \pm 16\%$	NA	$0.123 \pm 16\%$	NA

11. *B. subtilis* rrnIHG inversion on MOPS glucose w/ Casamino Acids



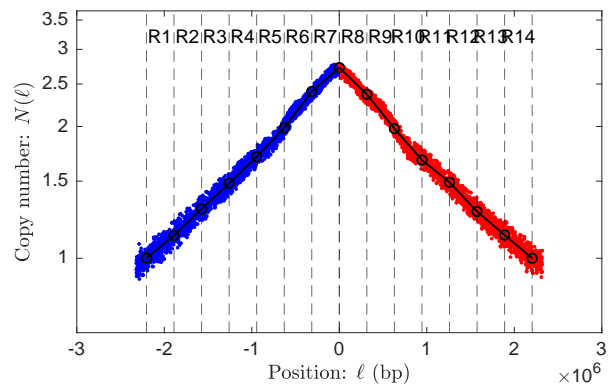
Region name	Region start position: ℓ_- (Mb)	Region end position: ℓ_+ (Mb)	Fork direction	Replication orientation	Slope: α (Mb $^{-1}$)	Velocity: v (kb s $^{-1}$)	Log difference: Δ	Duration: $\Delta\tau$ (min)
R1	-2.13	-6.09×10^{-3}	-	A	$0.285 \pm 2.3\%$	$0.921 \pm 2.3\%$	$0.605 \pm 2.3\%$	$38.4 \pm 2.3\%$
R2	-6.09×10^{-3}	9.49×10^{-2}	+	A	$0.278 \pm 4.0\%$	$0.945 \pm 4.0\%$	$2.81 \times 10^{-2} \pm 4.0\%$	$1.78 \pm 4.0\%$
R3	9.49×10^{-2}	0.124	+	A	$5.57 \pm 8.4\%$	$4.71 \times 10^{-2} \pm 8.4\%$	$0.161 \pm 8.4\%$	$10.2 \pm 8.4\%$
R4	0.124	1.70	+	A	$0.278 \pm 4.0\%$	$0.945 \pm 4.0\%$	$0.437 \pm 4.0\%$	$27.7 \pm 4.0\%$

12. *B. subtilis* *rrnIHG* inversion on MOPS glucose – Minimal



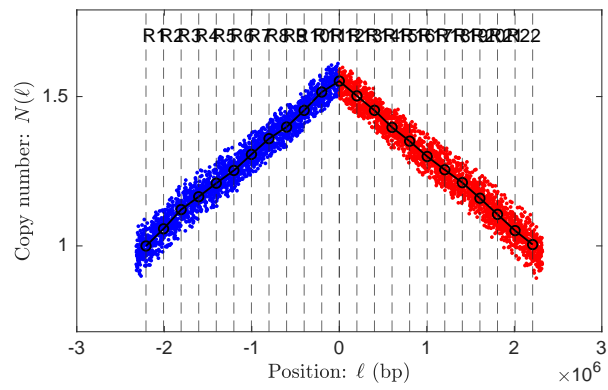
Region name	Region start position: ℓ_- (Mb)	Region end position: ℓ_+ (Mb)	Fork direction	Replication orientation	Slope: α (Mb $^{-1}$)	Velocity: v (kb s $^{-1}$)	Log difference: Δ	Duration: $\Delta\tau$ (min)
R1	-2.10	4.41×10^{-2}	-	A	$0.266 \pm 1.7\%$	$0.869 \pm 1.7\%$	$0.570 \pm 1.7\%$	$41.1 \pm 1.7\%$
R2	4.41×10^{-2}	0.131	+	A	$0.241 \pm 2.5\%$	$0.961 \pm 2.5\%$	$2.09 \times 10^{-2} \pm 2.5\%$	$1.51 \pm 2.5\%$
R3	0.131	0.218	+	A	$0.761 \pm 13\%$	$0.304 \pm 13\%$	$6.61 \times 10^{-2} \pm 13\%$	$4.77 \pm 13\%$
R4	0.218	2.01	+	A	$0.241 \pm 2.5\%$	$0.961 \pm 2.5\%$	$0.432 \pm 2.5\%$	$31.2 \pm 2.5\%$

13. *E. coli* WT on LB



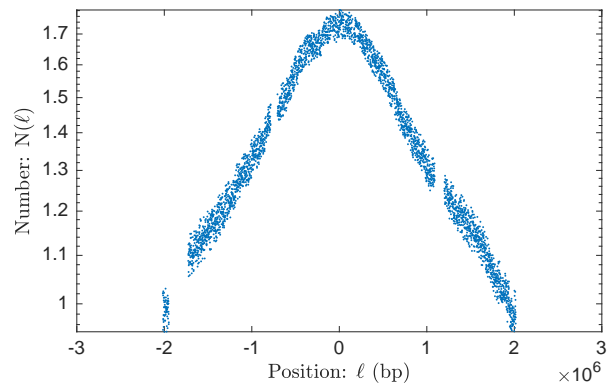
Detailed results tabulated in Supplementary Data.xlsx.

14. *E. coli* WT on M9 glucose



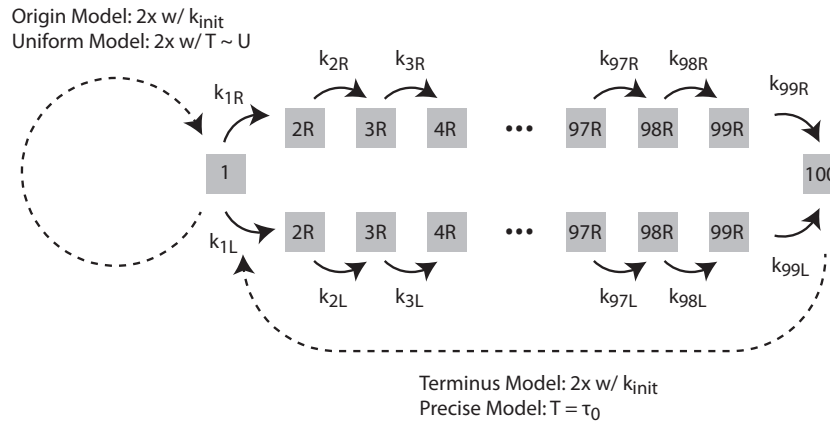
Detailed results tabulated in Supplementary Data.xlsx.

15. Marker frequency profile of *V. cholerae*



MCH1 in LB rich media. Flattening of the profile near the origin of replication suggests a slowdown in the rate of replication initiation, consistent with the population entering stationary phase.

16. Stochastic simulation model schematics.



Replication initiates at the origin (state 1) and proceeds bi-directionally along the left (L) and right arms (R). Each transition between simulation bases (*i.e.* state number) represents the replication of roughly 25 kb. States have exponentially-distributed lifetimes with a rate equal to the fork velocity at that position: $k_X = v(\ell_X)$. A number of different initiation schemes were simulated. In the *Origin Model*, the initiation rate was proportional to the number origins (per cell). In the *Terminus Model*, the initiation rate was proportional to the number termini (per cell). In the *Uniform Model*, there is a uniformly distributed wait time after initiation before the next initiation occurs. In the *Precise Model*, there is a fixed time interval after termination before the next initiation occurs.

SUPPLEMENTARY DISCUSSION

In this work, we have formalized the lag-time analysis approach. Although the approach has been understood at a conceptual level since the pioneering work of Cooper and Helmstetter [10], our recent exploration of stochastic models and the introduction of the exponential mean have clarified its interpretation [11] and, from an experimental perspective, the introduction of next-generation sequencing greatly expanded the potential of the approach for characterizing the dynamics of nucleic acids and in particular replication dynamics, where it has the potential to make precise measurements of replication timing. In multiple applications, we have used this approach to quantitatively measure time durations as short as seconds, a time resolution that is challenging, if not impossible, to achieve using other methods.

To appreciate the power of our approach, it is useful to compare our results to recent results of Nieduszynski and coworkers who have used experimental methods, cell synchronization (sync-seq), to directly resolve replication dynamics [12, 13]. In this case, the time resolution is limited by a combination of the precision of cell synchronization, which is an imperfect tool [14], and the frequency of fraction collection (every 5 minutes). Although it would be interesting to compare the relative precision of our approach to this competing method, the authors do not report fork velocities, pause durations, or provide an error analysis of their reported replication times, questioning to what extent the approach is truly quantitative. Since the fractions are collected on five-minute intervals, this time resolution is the floor of the direct time resolution achieved by this approach. In contrast, we report on a range of pause durations that are shorter than 5 minutes.

A significant experimental shortcoming of the sync-seq approach is the necessity of cell synchronization. In most systems, synchronization requires cell-cycle arrest, which introduces a significant potential for artifactual results [15]; whereas lag-time analysis probes dynamics in steady-state growth. Our own preliminary analysis suggests that the timing of initiation at a subset of loci in *Saccharomyces cerevisiae* is changed by the cell synchronization procedure relative to steady-state growth.

In addition to these quantitative and high-time-resolution applications, we have also demonstrated the approach in a more conceptual context: using lag-time analysis to argue that the observed oscillations in fork velocity were temporal rather than locus dependent.

It may seem perplexing that we have not pooled many existing datasets from multiple independent experiments. This would naïvely increase the statistical resolution and sensitivity from an analytical perspective. However, it is important to emphasize that not all marker-frequency experiments are of equal quality and that many datasets we have analyzed have clear signatures of systematic error. (See the Supplementary Methods 3B.) A signature of systematic error that appears in many datasets is significant flattening of the marker frequency in the vicinity of the

origin which appears intermittently. This feature is consistent with a culture that has begun to reduce the rate of initiation and suggests that early harvesting and rapid cell processing may be essential for the generation of optimal datasets. In our analysis, we have prioritized the selection of artifact-free datasets over the indiscriminate pooling of data.

We emphasize that to date, datasets have not been generated with quantitative replication dynamics analysis as a goal and we are confident that experimental protocols can be optimized to improve the data. Ref. [16] describes a promising approach, including harvesting populations earlier in exponential phase. We too are developing new protocols to increase data quality.

-
- [1] Galli, E. *et al.* Replication termination without a replication fork trap. *Scientific Reports* **9** (2019).
 - [2] 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* **104**, 5608–13 (2007).
 - [3] Srivatsan, A., Tehranchi, A., MacAlpine, D. M. & Wang, J. D. Co-orientation of replication and transcription preserves genome integrity. *PLoS Genet* **6**, e1000810 (2010).
 - [4] Midgley-Smith, S. L. *et al.* Chromosomal over-replication in *Escherichia coli* recG cells is triggered by replication fork fusion and amplified if replicore symmetry is disturbed. *Nucleic Acids Research* **46**, 7701–7715 (2018).
 - [5] Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359 (2012).
 - [6] Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10** (2021). URL <https://doi.org/10.1093/gigascience/giab008>. Giab008, <https://academic.oup.com/gigascience/article-pdf/10/2/giab008/36332246/giab008.pdf>.
 - [7] Bremer, H. & Churchward, G. An examination of the Cooper-Helmstetter theory of DNA replication in bacteria and its underlying assumptions. *J Theor Biol* **69**, 645–54 (1977).
 - [8] Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry* **81**, 2340–2361 (1977). URL <https://doi.org/10.1021/j100540a008>. <https://doi.org/10.1021/j100540a008>.
 - [9] Retkute, R., Hawkins, M., Rudolph, C. J. & Nieduszynski, C. A. Modeling of DNA replication in rapidly growing bacteria with one and two replication origins. *Preprint at bioRxiv* (2018). URL <https://www.biorxiv.org/content/early/2018/06/29/354654>. <https://www.biorxiv.org/content/early/2018/06/29/354654.full.pdf>.
 - [10] Cooper, S. & Helmstetter, C. E. Chromosome replication and the division cycle of *Escherichia coli* B/r. *J Mol Biol* **31**, 519–40 (1968).
 - [11] Huang, D., Lo, T., Merrikh, H. & Wiggins, P. A. Characterizing stochastic cell-cycle dynamics in exponential growth. *Phys. Rev. E* **105**, 014420 (2022). URL <https://link.aps.org/doi/10.1103/PhysRevE.105.014420>.
 - [12] Batrakou, D. G., Müller, C. A., Wilson, R. H. C. & Nieduszynski, C. A. DNA copy-number measurement of genome replication dynamics by high-throughput sequencing: the sort-seq, sync-seq and mfa-seq family. *Nat Protoc* **15**, 1255–1284 (2020).
 - [13] Müller, C. A. *et al.* The dynamics of genome replication using deep sequencing. *Nucleic Acids Res* **42**, e3 (2014).
 - [14] Withers, H. L. & Bernander, R. Characterization of *dnac2* and *dnac28* mutants by flow cytometry. *J Bacteriol* **180**, 1624–31 (1998).
 - [15] Bates, D. *et al.* The *Escherichia coli* baby cell column: a novel cell synchronization method provides new insight into the bacterial cell cycle. *Mol Microbiol* **57**, 380–91 (2005).
 - [16] Knöppel, A., Broström, O., Gras, K., Fange, D. & Elf, J. The coordination of replication initiation with growth rate in *Escherichia coli*. *Preprint at bioRxiv* (2022).