Journal Title Here, 2022, pp. 1–5

doi: DOI HERE
Advance Access Publication Date: 00.00.0000
Paper

PAPER

On the interaction of replication and transcriptional regulation

Pietro Gemo[1,2] and Matteo Brilli[1,*]

[1]Department of Biosciences, University of Milan, Via Celoria, 26, 20133, Milano, Italy and [2]Present affiliation: Infectious Disease Epidemiology Group, Max Planck Institute for Infection Biology, 10117, Berlin, Germany
*Corresponding author. email:matteo.brilli@unimi.it
FOR PUBLISHER ONLY Received on Date Month Year; revised on Date Month Year; accepted on Date Month Year

Abstract

DNA replication introduces a gradient of gene copy number, and in Bacteria it affects gene expression accordingly. In *E. coli* and other species, the average gradient in the population can be estimated on the basis of a relationship with the growth rate. In this work we integrated this growth- and position-dependent gradient into a classical transcriptional regulation model, to evaluate their interactions. The theoretical treatment of our model highlights that the sensitivity to transcription factor-mediated regulations acquires and extra-dimension related to the position of a locus on the oriter axis and division time. This reinforce the idea of replication as an additional layer of complexity in gene regulation. We highlight that the contributions of the replication- and transcription factor-mediated regulations to the final variance of the target gene can in theory work in concert or counteract each other, and we discuss why this is important from an evolutionary point of view. Finally, we highlight the fact that this treatment may improve the estimation of kinetic parameters for transcription factor activity using RNA-seq data, and the estimation of transcript variance in differential gene expression analysis.

Key words: DNA replication, global regulation, mathematical model,systems biology

Introduction

Most Bacteria in the environment experience a so-called *feast-famine* cycle [12]: short periods of exponential growth when nutrients happen to be abundant, followed by nutrient depletion and then famine times lasting much longer. Most cells die during the latter, and the remaining significantly slow down their growth rate. Intuitively, species that can promptly shift to fast growth rates in rich conditions can exploit a larger fraction of the nutrients before the situation changes thereby out-competing their peers. *E. coli* implements a successful strategy in this sense: it is mono-ploid when slow growing - e.g. at $\mu \approx 0.7h^{-1}$ ($\tau \approx 1h$) but in particularly rich conditions, it can shift to a mero-oligoploid growth *mode*, whereby cells are able to fire multiple replication forks during the same cell cycle round (e.g. at $\tau \approx 25min$ or $\mu \approx 1.7h^{-1}$, and Figure 1) [8]. This growth *mode* enables certain species to divide faster than the time required to copy the chromosome, which in *E. coli* is relatively constant ($C \approx 40min$) [5]. Since the activity of each replication fork has the effect of increasing by one the copy number of loci on the chromosome at its passage a mero-oligoploid growth mode confers each single cell with a (potentially different) gradient of gene copy number having a maximum near the Origin and a minimum near the Terminus (Figure 1).

This situation was formalized since the late 60s [3, 5, 9]; in the classical framework, the copy number n_a of locus a located at a relative position p_a on the Ori/Ter axis, can be calculated as:

$$n_a = 2^{(p_a C + D)/\tau}, \tag{1}$$

where C, D have the meaning as in Figure 1b and τ is the division time (B is usually so short, it does not take part in the equations); p_a is the relative distance of the locus from the terminus; in this coordinate system $p_{Ori} = 1$ and

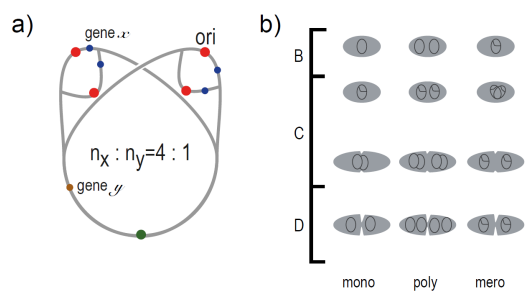

Fig. 1. a) Representation of a circular chromosome with three active replication forks, which translates in $n_{Ori} = 4$. Genes x and y, are in a 4 : 1 ratio but it changes with the number of replication forks; an *E. coli* cell can for instance be born with a single chromosome ($n_{Ori} = n_{Ter} = 1$) and after some time of exponential growth in a rich environment its progeny can have several genome equivalents and/or active replication forks. b) Different phases of cell life cycle for mono (exact), poly and mero-oligo ploids (idealized) bacterial cells. C is the time needed for the replication of an entire chromosome. D is the time needed for segregation and cell division. In *E. coli* B is a *virtual* cell cycle phase, as daughter cells immediately start replicating the genome in exponential phase. For this reason, B has no part in the equations for multiplicity prediction.

$p_{Ter} = 0$. By introducing these parameters into Eq. 1, we get:

$$n_{Ori} = 2^{(C+D)/\tau} \tag{2}$$

$$n_{Ter} = 2^{D/\tau}. \tag{3}$$

Although the model was developed on *E. coli*, where it has an extremely good fit over a wide range of growth rates, it is similarly valid in other species

© The Author 2022. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

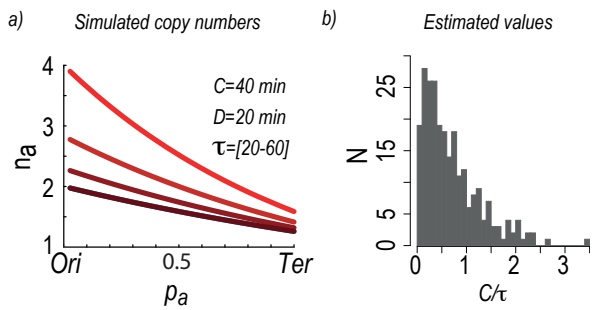

Fig. 2. (a) Multiplicity n_a as a function of genomic position p_a, computed in a range of 4 division times using Eq. 1; (b) Estimation of C/τ in many different species, using genome sequencing reads.

[14]. Indeed, since Eq. 2 and 3 imply:

$$\log_2 \frac{n_{Ori}}{n_{Ter}} = \frac{C}{\tau}, \tag{4}$$

we can extrapolate the growth rate *in vivo* by exploiting shotgun metagenomic sequencing data, i.e. the so-called peak-to-trough (PTR) analysis [15, 19, 10].

We previously estimated C/τ for many species grown in laboratory conditions, by using genome sequencing libraries mapped on selected $100kb$ regions centered on the Ori and Ter loci, previously localized by compositional analysis [4]. Since growth conditions for genome sequencing are generally designed to achieve massive growth in view of DNA extraction, we assume the values represent proxy of the maximum attainable by each species. Mero-oligoploid species can reach short division times by multi-fork replication, and they can be identified as the fraction of species with $C/\tau > 1$ (Figure 2b).

Replication and transcription occur concomitantly in Prokaryotes, therefore division time-dependent copy number variations affect transcript abundances [1, 16, 15], and replication becomes an additional player in determining the transcriptome configuration of an organism.

In this work we integrate the above ideas in a standard gene regulation model to disentangle the different contribution and we show that this approach highlights interesting evolutionary cues together with a deeper understanding of this global level of regulation.

Results

Mathematical formalization

A standard model describing the evolution in time of the abundance of a transcript a, under the regulation of a transcription factor T is:

$$\frac{da}{dt} = f(T, \boldsymbol{p}) - \gamma a \tag{5}$$

Where γ is the degradation rate, assumed to be constant across conditions and genes; $f(T, \boldsymbol{p})$ models the transcription rate of a promoter, as a function of the abundance of a transcriptional regulator T plus parameters. Often, this is a sigmoid (Eq. 6) or hyperbolic function that can be positive depending on the regulator's role [7]:

$$h^+(T, \boldsymbol{p} = [\kappa, \theta, n]) = \kappa \frac{T^n}{\theta^n + T^n} \tag{6}$$

$$h^-(T, \boldsymbol{p} = [\kappa, \theta, n]) = \kappa - h^+(T, \boldsymbol{p} = [\kappa, \theta, n]), \tag{7}$$

where κ is the maximum transcription rate achievable with this promoter at saturating regulator concentrations and θ (in units of concentration), is the regulator's abundance inducing an half-maximal rate. Implicitly, this function represents the output by a single copy of the promoter, and it has a hyperbolic functional response with $n = 1$, sigmoidal with $n \geq 2$. Typically, $n \leq 4$.

Therefore, since prokaryotic genes can be transcribed just after their replication, it is natural to use a locus' copy number as a multiplier of the above:

$$\frac{da}{dt} = n_a f(T, \boldsymbol{p}) - \gamma a, \tag{8}$$

where n_a is the multiplicity of locus a. Eq. 8 can be easily solved for the steady state (when $\frac{da}{dt} = 0$):

$$a^{ss} = n_a \frac{f(T, \boldsymbol{p})}{\gamma} = 2^{(p_a C + D)/\tau} \frac{f(T, \boldsymbol{p})}{\gamma} \tag{9}$$

A solution, that can be better discussed taking the $\log_2$:

$$\log_2 a^{ss} = \frac{p_a C + D}{\tau} + \log_2 \frac{f(T, \boldsymbol{p})}{\gamma}. \tag{10}$$

Eq. 10 highlights the different contributions to the steady state abundance of the transcript; as expected, the relative activity of the regulator with respect to the degradation rate is one, together with the dependence on division time and locus position. Therefore, changes in division time can affect the abundance of a transcript even with no change in the activity of its regulator. Since the two contributions can be considered as potentially independent, one pertinent question at this point is how they may interact in determining a certain steady state.

Sensitivity Analysis

To evaluate how the transcript steady state abundance depends on the two contributions in a more quantitative way, we derived analytical formulas for the sensitivity of the solution to the two parameters of interest. Sensitivities can be calculated by partial derivation of the steady state expression with respect to one model parameter at time, obtaining:

$$\frac{\partial a^{ss}}{\partial T_+} = n_i^a \frac{\kappa n}{\gamma T_+} \frac{\theta^n}{T_+^n + \theta^n} \frac{T_+^n}{T_+^n + \theta^n} \tag{11}$$

$$\frac{\partial a^{ss}}{\partial T_-} = -\frac{\partial a^{ss}}{\partial T_+}, \tag{12}$$

for the two possible regulatory effects, and:

$$\frac{\partial a^{ss}}{\partial \tau} = -\ln 2 \, \frac{(p_a C + D) 2^{(p_a C + D)/\tau}}{\tau^2} \, \frac{f_\star(T, \boldsymbol{p})}{\gamma}, \tag{13}$$

where $f_\star$ is the original regulatory function (positive or negative).

Sensitivities are represented in Figure 3 for biologically meaningful parameter ranges ($T \in [0.1, 250]\,\mu mol$, i.e. ≈ 11.3 log fold change around θ and $\tau \in [30, 70] min$). In Figure 3a we show the sensitivity of steady state transcript abundance to τ, when the transcription factor is fixed at a θ concentration, for all genomic positions. As expected, genes located closer to the origin are more sensitive to copy number changes; the negative sign comes out from the inverse relationship of division time and copy number i.e. when τ decreases (cells divide faster), a locus copy number grows. Additionally, copy number and sensitivity are positively correlated. In Figure 3b we show that the sensitivity to a positive regulator with a sigmoidal functional response acquires a positional pattern once replication is accounted for. As before, the sensitivity is higher nearby the origin and it increases at shorter division times (or equivalently, toward the origin), in agreement with what recently described by [20] for constitutive genes. Indeed our treatment is general in this sense, and T could also represent a virtual regulator of constitutive genes e.g. a σ factor. In Fig.3c the functional response is of the hyperbolic type; sensitivity in this case is maximum at very low regulator abundances and increases with copy number; the usual sigmoidal function instead translates into a log-normal shape, with a global maximum just before the regulator reaches a concentration θ (Fig. 3d). Increasing the steepness parameter (n) in the Hill function, makes the range

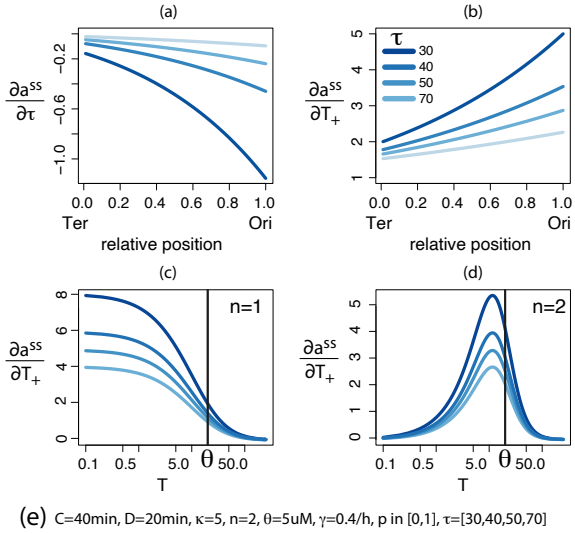

Fig. 3. Sensitivity analysis. (a) Sensitivity to τ around different τ values, relative to genome position. (b) Sensitivity to transcription factor abundance (T) at various τ values, at $T = \theta$; (c) Sensitivity to transcription factor abundance at various τ values, relative to$T$ itself and setting $n = 1$ in the Hill function (hyperbolic response); (d) as in (c) with $n = 2$ (sigmoidal response); (e) parameters used, if not otherwise indicated.

of non-zero sensitivity narrower (not shown). The above clearly illustrates that copy number affects the sensitivity to the regulator, therefore replication can introduce a positional and growth rate-dependent pattern in input/output relationships of gene regulatory networks.

Variance decomposition

Since the steady state transcript level depends on both replication and transcriptional control, its variance across conditions also is a function of both. Using Eq. 10, we apply the rule for calculating the variance of a sum of functions, obtaining:

$$Var(\log_2 a_{ss}) = Var(log_2 f(\tau, \boldsymbol{p})) + Var(log_2 f(T, \boldsymbol{p})) + \\ +2Cov(log_2 f(\tau, \boldsymbol{p}), log_2 f(T, \boldsymbol{p})), \quad (14)$$

where $f(\tau, \boldsymbol{p}) = n_a$ as from Eq. 1. The magnitude of the variance and covariance terms depend on the family distribution of functions and the exact values of parameters, but the implications are general. Replication introduces a position and division time-dependent pattern to the variance, because genes at different positions will have unequal variances when division time changes, even if expressed by identical promoters. More specifically, if changes in copy numbers driven by division time (i.e. indicated above as $f(\tau, \boldsymbol{p})$) and in the activity of the regulator ($f(T, \boldsymbol{p})$) are uncorrelated, the covariance term (in red in Eq. 14) is zero by definition; in this case, the variance of the log-transformed transcript abundance is simply additive in those of the two functions. On the other hand, if the covariance term is not zero, it may confer interesting effects on the variance of the transcript: when it is positive, the final variance of the transcript increases faster than in the purely additive case. This can be interpreted as a replication-induced amplification of the activity of the regulator and could therefore be particularly advantageous for genes involved in growth-rate dependent processes (e.g. σ70-dependent promoters). Incidentally, this reasoning provides an explanation to the documented tendency of *E. coli* and other fast-growing species to keep genes benefiting from high expression levels toward the origin [18]. On the converse, when the covariance term is negative, it will somehow buffer part of the total variance and the transcript abundance will be stabilized (Figure 4). We can therefore say that replication and transcription can cooperate or counteract each other, and the two may have important repercussions in the Input/Output relationships of the system.

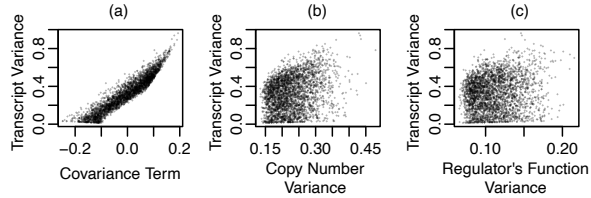

Fig. 4. Variance - covariance relationships from the correlation analysis. a) The variance of the transcript at steady state is positively related to the covariance term. In particular, a negative covariance in the inputs strongly reduces the variance in the output (stabilizing effect), while a positive covariance term causes the opposite. b) and c) show that the variance of the output is uncorrelated to either inputs' variance.

Input/Output Analysis

To study this aspect into more detail, we produce inputs (multiplicities, $f(\tau, \boldsymbol{p})$ and transcription rates, $f(T, \boldsymbol{p})$) over the entire covariation spectrum, we solve the system and then we check how the output (a^{ss}) relates to the inputs. Each simulation comprises: (i) drawing copy numbers for several random values of τ ($\in [20, 120]$min), (ii) generating transcription factor abundances to have a known covariance with copy numbers, (iii) calculating transcript abundances at steady state (Eq. 9), (iv) calculating input-outputs correlations ($corr(a^{ss}, f(\tau, \boldsymbol{p}))$ and $corr(a^{ss}, f(T, \boldsymbol{p}))$), (v) comparison of the latter with the original correlation imparted to the inputs.

Results of this analysis are summarized in Tables 1-4. We consider positive and negative regulation, and promoters with functional response of type II (hyperbolic) and III (sigmoidal). The emerging pattern can be summarized as: (i) positive correlation in the inputs **always** results in significant correlations of the output with both inputs. This should be a relatively common case for regulators not depending on post-translational activation, as their gene expression will be also affected by division time changes; (ii) Type III promoters tend to reflect the expected correlation with the regulator's function but often not with multiplicity, basically disentangling the transcript's abundance from the latter; (iii) Type II promoters tend to produce outputs that are instead more correlated with multiplicity than with the regulator's function; when T is a repressor it is more effective in defining patterns (ii) and (iii). (iv) However, the interference of copy numbers and regulator's activity often mask the expected input/output relationship of a^{ss} and transcription rate. While being aware that the exact values of correlations and their significance clearly depends on the parameters, by testing biologically meaningful regulator concentrations and division times, we here highlight that the interaction of copy number variations induced by changes in division time and transcription factor mediated regulation can result in an extended range of input/output relationships: positively correlated inputs provide a way to amplify the signal mediated by the regulator; on the converse, negative regulation on a type III promoter may provide a way to remove the effect of copy number variations and highlight the pattern imparted by the regulator.

Conclusions

The relationship between position and transcription rate due to replication-induced copy number variations introduces an additional layer of regulation for modulating gene expression that has a positional dimension, together with a dependency on growth rate. This hypothesis can find support in papers demonstrating that gene positioning is a good predictor of gene expression levels [1, 2, 6, 13, 16, 17]. In this work, we studied the interaction of replication-induced copy number variations and transcriptional regulation from a theoretical perspective. We derived analytical formula for the sensitivity of the steady state transcript's abundance to both division time and regulator's activity and we show that replication induces a positional and division-time dependent pattern to the sensitivity of a transcript's abundance to a regulator's activity. In this way, the cells have an additional mean to modulate the sensitivities of promoters on the basis of their position on the ori/ter axis as

	$f(\tau,\boldsymbol{p})$			$f(T,\boldsymbol{p})$		
output / input	-	0	+	-	0	+
-	0.22	0.00	0.00	0.00	0.00	0.00
0	0.78	0.82	0.00	0.59	0.29	0.00
+	0.00	0.18	1.00	0.41	0.71	1.00

Table 1. Input/output correlation analysis when the regulator is an activator. ($\kappa = 10$, $n = 2$, other parameters are indicated in the main text). Rows corresponds to correlations of the output (a^{ss}) with the function indicated in the header of each column block. In all cases, $+$ ($-$) means a significant ($p \leq 0.0001$) positive (negative) correlation and 0 means the correlation is not significant. Columns concern the correlation in the inputs ($corr(f(\tau,\boldsymbol{p}), f(T,\boldsymbol{p}))$). The "$-$" columns for instance mean that in the samples with inputs negatively correlated, a^{ss} is negatively correlated with multiplicity in 22% of the cases while the correlation is non significant in the remaining 78%; additionally, in only 41% of those cases, a significant and positive correlation with the transcription rate was detected; therefore, negatively correlated inputs often result in loss of correlation of the transcript abundance with respect to the regulator's function. If the inputs are uncorrelated, the outputs also can become uncorrelated from the inputs. Differently, positively correlated inputs always result in high correlations in the outputs.

	$f(\tau,\boldsymbol{p})$			$f(T,\boldsymbol{p})$		
output / input	-	0	+	-	0	+
-	0.00	0.00	0.00	0.22	0.00	0.00
0	0.39	0.25	0.00	0.78	0.72	0.00
+	0.61	0.75	1.00	0.00	0.28	1.00

Table 2. Same as Table 1 but with $n = 1$.

	$f(\tau,\boldsymbol{p})$			$f(T,\boldsymbol{p})$		
output / input	-	0	+	-	0	+
-	0.00	0.00	0.00	0.00	0.00	0.00
0	1.00	0.83	0.00	0.21	0.25	0.00
+	0.00	0.17	1.00	0.79	0.75	1.00

Table 3. Same as Table 1, but considering the case of T as a negative regulator. In this case, a negative correlation in the inputs translates into uncorrelated output with multiplicity, and correlation with transcription rate is significant in almost 80% of the cases, but the correlation with multiplicity is erased.

	$f(\tau,\boldsymbol{p})$			$f(T,\boldsymbol{p})$		
output / input	-	0	+	-	0	+
-	0.00	0.00	0.00	0.00	0.00	0.00
0	0.07	0.30	0.00	1.00	0.78	0.00
+	0.93	0.70	1.00	0.00	0.22	1.00

Table 4. Same as Table 3 but with $n = 1$.

a function of division time. We highlight that the contribution of copy numbers and transcriptional regulation can interfere *constructively* or *destructively* in determining the variance of transcripts, and that this produces an enriched set of regulatory outputs that can be manipulated by controlling how the inputs are related to each other. A recent work [20] showed that constitutive promoters with the highest sensitivity to growth rate changes after the copy number effect is removed, tend to be located toward the origin, and that their position is moreover conserved across many species. This could be explained by the evolutionary exploitation of the constructive effect of copy number and regulation variations highlighted by our treatment.

The decoupling of the variance of transcript abundance into its components, shows that detecting the dependencies of transcript abundance to transcription factor activity using RNA-seq compendia could be hampered by division time variations across conditions. Indeed, predicting the multiplicity of the loci from RNA-seq data could enable to get rid of the noise introduced by variable growth rates in gene expression data, when the target is for instance promoter / regulator kinetic parameter estimation. Indeed, by referring to Eq. 10, we notice that even with no change in regulator activity and degradation rate, growth in fluctuating conditions affecting division times, would instead be taken as consequent to genuine variations in regulator's activity and incorporated into the parameters. Similarly, as the relationship of replication with other transcriptional regulations can mask the dependency of the output to the regulator concentration, another possible application would be in devising a correction aimed at removing the effect of copy numbers and consequently at highlighting the relationship with the regulator's abundance, in gene regulatory network reconstruction. We conclude adding that this work also suggests that multiplicity may help in the estimation of gene-specific dispersion factors in differential gene expression analysis (e.g. [11]) when replicates are in small number, and the conditions differ in terms of division times.

Methods

Scripts in R for reproducing all the analysis and figures in the paper are available at the team's repository on github: Comparative Systems Biology Lab.

Section title of first appendix

Subsection title of first appendix

Competing interests

No competing interest is declared.

Author contributions statement

MB conceived the idea. MB and PG developed the models and performed the analysis.

Acknowledgments

MB wish to thank the students of his 2022 course at the Quantitative Biology Master Degree of the University of Milan(Mario Deligios, Daniele di Bella, Greta Carola Giannini, Elahe Hassanbeiki, Frederik Espersen Knudsen and Alberto Peruzzi) for participating to the development of this idea.

References

1. David H. Ardell and Leif A. Kirsebom. The genomic pattern of tDNA operon expression in e. coli. *PLoS Computational Biology*, 1(1):0086–0099, 2005.
2. Dena H.S. Block, Razika Hussein, Lusha W. Liang, and Han N. Lim. Regulatory consequences of gene translocation in bacteria. *Nucleic Acids Research*, 40(18):8979–8992, 2012.
3. Hans Bremer and Gordon Churchward. An examination of the Cooper-Helmstetter theory of DNA replication in bacteria and its underlying assumptions. *Journal of Theoretical Biology*, 69(4):645–654, 1977.
4. Delphine Charif and Jean R. Lobry. SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In U Bastolla, M Porto, H E Roman, and M Vendruscolo, editors, *Structural approaches to sequence evolution: Molecules, networks, populations*, pages 207–232. 2007.
5. Stephen Cooper and Charles E. Helmstetter. Chromosome replication and the division cycle of Escherichia coli Br. *Journal of Molecular Biology*, 31(3):519–540, 1968.
6. Etienne Couturier and Eduardo P.C. Rocha. Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. *Molecular Microbiology*, 59(5):1506–1518, 2006.
7. Hidde de Jong. Modeling and simulation of genetic regulatory systems: a literature review. *Journal of Computational Biology*, 9(1):67–103, 2002.

8. Charles E. Helmstetter. Timing of Synthetic Activities in the Cell Cycle. *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2:1627–1639, 1996.
9. Charles E. Helmstetter and Stephen Cooper. DNA synthesis during the division cylce of rapidly growing E. coli B/r. *Journal of molecular biology*, pages 507–518, 1968.
10. Tyler A. Joseph, Philippe Chlenski, Aviya Litman, Tal Korem, and Itsik Pe'er. Accurate and robust inference of microbial growth dynamics from metagenomic sequencing reveals personalized growth rates. *Genome Research*, 32(3):558–568, 2022.
11. Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15(12):550, 2014.
12. Manon Morin, Brice Enjalbert, Delphine Ropers, Laurence Girbal, and Muriel Cocaign-Bousquet. Genomewide Stabilization of mRNA during a Feast-to-Famine Growth Transition in Escherichia coli. *mSphere*, 5(3), 2020.
13. Jatin Narula, Anna Kuchina, Dong Yeon D. Lee, Masaya Fujita, Gürol M. Süel, and Oleg A. Igoshin. Chromosomal Arrangement of Phosphorelay Genes Couples Sporulation and DNA Replication. *Cell*, 162(2):328–337, 2015.
14. Vito Pecoraro, Karolin Zerulla, Christian Lange, and Jörg Soppa. Quantification of ploidy in proteobacteria revealed the existence of monoploid, (mero-)oligoploid and polyploid species. *PloS one*, 6(1):e16392, 2011.
15. Andrew Pountain, Peien Jiang, Magdalena Podkowik, Bo Shopsin, Victor J. Torres, and Itai Yanai. A quantitative model for the transcriptional landscape of the bacterial cell cycle Andrew. *bioRxiv*, 2022.
16. Jelle Slager, Morten Kjos, Laetitia Attaiech, and Jan Willem Veening. Antibiotic-induced replication stress triggers bacterial competence by increasing gene dosage near the origin. *Cell*, 157(2):395–406, 2014.
17. Carolina Sousa, Víctor De Lorenzo, and Angel Cebolla. Modulation of gene expression through chromosomal positioning in Escherichia coli. *Microbiology*, 143(6):2071–2078, 1997.
18. Marie Touchon and Eduardo P.C. Rocha. Coevolution of the Organization and Structure of Prokaryotic Genomes. *Cold Spring Harbor Perspectives in Biology*, 8(1):a018168, 2016.
19. Jake L. Weissman, Shengwei Hou, and Jed A. Fuhrman. Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. *Proceedings of the National Academy of Sciences of the United States of America*, 118(12):1–10, 2021.
20. Pablo Yubero and Juan F. Poyatos. The Impact of Global Transcriptional Regulation on Bacterial Gene Order. *iScience*, 23(4):101029, 2020.

Matteo Brilli, Ph.D. is an Associate Professor in Microbiology at the University of Milan (Italy). He is interested in basic research and general questions that he approaches by integrating his variegated expertise in bioinformatics and systems biology.

Pietro Gemo. At the time of writing he was an undergraduate of the Master in Quantitative Biology of the University of Milan and now a Ph.D. candidate at the Max Planck Institute for Infection Biology (Berlin).