

Evolution of Complex Regulation for Cell-Cycle Control

Samuel H. A. von der Dunk *, Berend Snel , and Paulien Hogeweg

Theoretical Biology and Bioinformatics, Department of Biology, Science Faculty, Utrecht University, Padualaan 8, 3584 CH, Utrecht, The Netherlands

*Corresponding author: E-mail: s.h.a.vonderdunk@uu.nl.

Accepted: 20 April 2022

Abstract

Many questions remain about the interplay between adaptive and neutral processes leading to genome expansion and the evolution of cellular complexity. Genome size appears to be tightly linked to the size of the regulatory repertoire of cells (van Nimwegen E. 2003. Scaling laws in the functional content of genomes. *Trends Gen.* 19(9):479–484). In the context of gene regulation, we here study the interplay between adaptive and nonadaptive forces on genome and regulatory network in a computational model of cell-cycle adaptation to different environments. Starting from the well-known *Caulobacter crescentus* network, we report on ten replicate *in silico* evolution experiments where cells evolve cell-cycle control by adapting to increasingly harsh spatial habitats.

We find adaptive expansion of the regulatory repertoire of cells. Having a large genome is inherently costly, but also allows for improved cell-cycle behavior. Replicates traverse different evolutionary trajectories leading to distinct eco-evolutionary strategies. In four replicates, cells evolve a generalist strategy to cope with a variety of nutrient levels; in two replicates, different specialist cells evolve for specific nutrient levels; in the remaining four replicates, an intermediate strategy evolves. These diverse evolutionary outcomes reveal the role of contingency in a system under strong selective forces.

This study shows that functionality of cells depends on the combination of regulatory network topology and genome organization. For example, the positions of dosage-sensitive genes are exploited to signal to the regulatory network when replication is completed, forming a *de novo* evolved cell cycle checkpoint. Our results underline the importance of the integration of multiple organizational levels to understand complex gene regulation and the evolution thereof.

Key words: genome expansion, complexity, cell cycle, computational model, generalism, *in silico* evolution.

Significance

Throughout evolution, we see genome expansion and increase in cellular complexity, but many questions remain about the interplay between adaptive and neutral processes leading to these two trends. Using a computational evolutionary model, we find that genome expansion is driven by and required for adaptation to occur; large genomes are costly, minimizing neutral forces. This study highlights how adaptations to new habitats are driven by various emergent mechanisms enabled by the multilevel genotype–phenotype map, which apart from genome expansion include genome organization, regulatory network architecture, and cell-cycle checkpoints.

Introduction

The evolution of cellular complexity is a major question in evolutionary biology. Gene regulation plays a key role in many complex cellular features, from metabolic plasticity in generalist bacteria to cell differentiation in multicellular organisms.

With increasing genome size, a larger fraction of the coding genome is devoted to regulatory function compared to information processing and metabolism (van Nimwegen, 2003). Thus, for understanding the evolution of complexity, it is essential to understand how gene regulatory networks evolve.

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution- Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use journals.permissions@oup.com

Changes in gene expression have been associated with adaptation and speciation, suggesting that selection is the main architect of gene regulatory networks (Gilad et al., 2006; Whitehead and Crawford, 2006; Fay and Wittkopp, 2008; Zheng et al., 2011). Yet, models have shown that non-adaptive forces also shape the network (Cordero and Hogeweg, 2006; Lynch, 2007). Large networks are selected to promote evolvability (Cuypers and Hogeweg, 2012, 2014) and mutational robustness (Soyer and Bonhoeffer, 2006). Conversely, small networks are selected to limit off-target interactions in the face of gene expression noise (Jenkins and Stekel, 2008, 2010). The increase in complexity of regulatory networks likely results from a combination of adaptive, mutational, and stochastic processes. The interplay between these processes and further mechanistic details remain elusive.

Modeling approaches facilitate mechanistic understanding of gene regulatory network evolution, complementing comparative studies on real data which are limited to extant organisms (Romero et al., 2012). So far, models have focused on simple cases where regulatory networks needed to solve fixed functional challenges, thereby potentially overlooking important evolutionary dynamics, such as neutral variation during long periods of stasis (Quayle and Bullock, 2006). To overcome these issues, we turn to modeling of the cell cycle, which provides an intrinsic and dynamic fitness criterion for a cell's regulatory network. Even in prokaryotes, cell-cycle control consists in integration of internal and external cues for correct timing of major cellular events such as genome replication, growth, and cell division. Furthermore, due to the lack of strictly separated stages as in eukaryotes, replication impacts ongoing gene expression (Pelve et al., 2011; Paijmans et al., 2016; Walker et al., 2016; Jaruszewicz-Błóńska and Lipniacki, 2017).

Cell-cycle regulation in prokaryotes is best understood in the model species *Caulobacter crescentus*, an alpha-proteobacterium that lives in nutrient-poor freshwater streams (Sánchez-Osorio et al., 2017). The five main transcription factors that orchestrate the *C. crescentus* cell cycle have been identified: CtrA, GcrA, DnaA, CcrM, and SciP (Tan et al., 2010; Quiñones-Valles et al., 2014). These five genes produce a cyclic expression pattern that drives all downstream cell-cycle functions (Zhou et al., 2015) and which can be reproduced in a simple Boolean model (Quiñones-Valles et al., 2014; Sánchez-Osorio et al., 2017). We use this model as starting point to study the evolution of complex cell-cycle behavior in cells exposed to more challenging, nutrient-limited conditions.

Results

Model

We define a cell cycle by the five established core genes from *C. crescentus* (dubbed g1–5) whose combined

expression states represent the four cell-cycle stages observed in Quiñones-Valles et al. (2014) (dubbed G1, S, G2, and M, fig. 1; see Methods for a more detailed description of the model). To allow evolution of this well-studied system, we embed the Boolean cell-cycle network in cells that die, replicate their genome, and divide. The gene

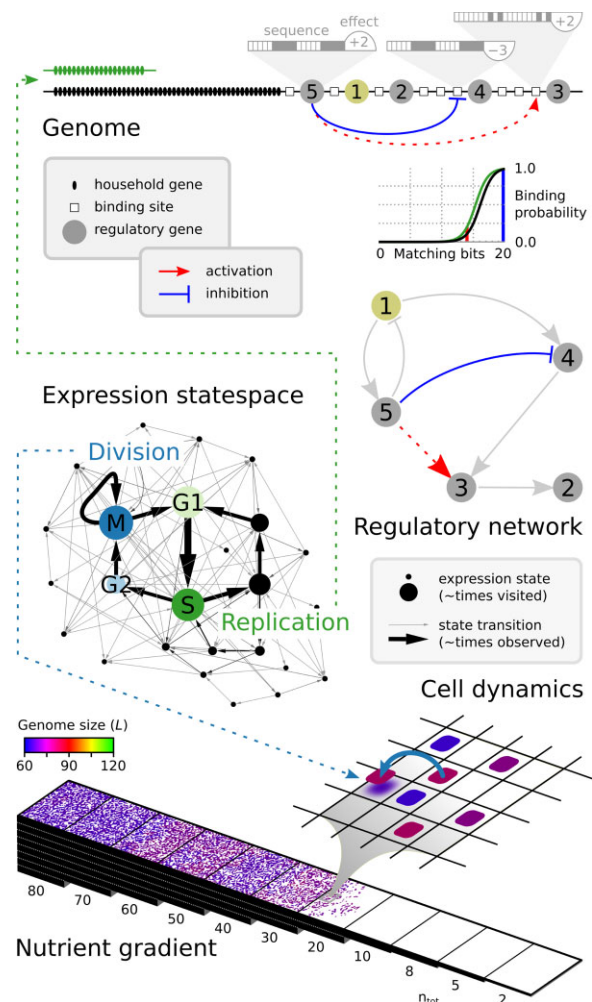


FIG. 1.—Overview of the multilevel evolutionary model. Regulatory interactions are defined at the genome level: binding probability is a function of bitstring similarity (see eq. 3 in Methods). The gene regulatory network is a useful simplification of these dynamics. Regulatory interactions give rise to gene expression dynamics shown in the statespace, where four states (G1, S, G2, and M) define the cell cycle. For instance, state S is defined by expression of only g3 and g5. In S, cells replicate a discrete segment of their genome (green dotted arrow). Replicated gene copies also participate in regulatory dynamics, increasing effective binding probability (i.e. from black to green curve). In state M, cells divide—given that genome replication is finished (blue dotted arrow). As shown, division may lead to overgrowing of a neighboring cell. Evolution of cell dynamics takes place on a nutrient gradient; in addition to this gradient, local cell density determines the amount of available nutrients (which corresponds to the size of the genome segment that is replicated if the cell is in S). Grid size is 50 × 550, giving a maximum population size of 27,500 cells.

regulatory network is coded in genomes consisting of discrete *beads*, that is, genes and binding sites (this representation is known as a beads-on-a-string genome; e.g. Crombach and Hogeweg, 2007). Noise is introduced through sequence-specific binding affinities between gene products and binding sites (where affinity is a function of bitstring similarity); out of all expressed genes, only one may bind a given binding site each timestep.

Rather than imposing an explicit fitness criterion on cells, cellular dynamics are directly linked to regulatory dynamics. A cell divides once it completes the four-stage gene expression cycle upon reaching the division stage (M). A crucial part of the cell cycle is genome replication, which happens at a rate of n beads per timestep when a cell is in the replication stage (S). How long a cell takes to replicate its entire genome depends on the genome size L and the replication rate which in turn depends on habitat quality, hereafter referred to as nutrient abundance n . The need to execute multiple replication steps in poor nutrient conditions provides a drive for the evolution of more regulation. Yet replication takes time, so there is inherent selection against genome expansion as larger genomes take longer to replicate. Finally, due to explicit genome replication, the cellular growth dynamics also feed back onto the regulatory dynamics through changes in gene dosages.

Initialization

Cells with the *C. crescentus* cell-cycle network (Quiñones-Valles et al., 2014) encoded as described above were preevolved for 10^5 timesteps (which here amounts to 2.5×10^4 generations) in rich nutrient conditions to adapt to the model formalism and specifically the newly introduced stochasticity in regulatory dynamics. The resulting most common genotype is streamlined and regulates a cell cycle that is robust to the intrinsic stochasticity (cf. Jenkins and Stekel, 2010, see also “Evolution of *C. crescentus* Cell cycle under Stochastic Gene Expression” in Appendix 1). With this genotype, we set up 10 replicate evolution experiments (R1–R10) where cells are confronted with nutrient limitation (fig. 2, Supplementary fig. S6, Supplementary Material online). To enable step-wise adaptation to more challenging habitats, nutrients are increasingly limited in sectors along a gradient, analogous to well-known *in vivo* evolution experiments with *Escherichia coli* (Baym et al., 2016). Locally, nutrients are further depleted by the cell count (see Methods). To expand their range to poorer conditions, cells need to evolve to adjust the duration of the replication stage (S) such that the genome can be fully replicated before reaching the division stage (M).

Range Expansion Correlates with Genome Size

In all replicates, cells succeed in evolving a cell cycle to cope with poor nutrient conditions (fig. 3). Four of the ten replicate experiments (R10, R3, R2, and R8) when ordered by

final population size illustrate the following general pattern (fig. 2; see Supplementary fig. S6, Supplementary Material online for all 10 replicates). On a comparatively short time-scale ($t \sim 10^5$), rapid range expansion to sector 5–7 is observed, and cell density increases in rich environments. Whenever cells adapt to poorer nutrient conditions, higher cell density (causing further nutrient depletion) can be sustained in previously colonized sectors. Subsequent range expansion occurs in infrequent bursts and is less consistent across replicates than the initial expansion.

By the end of the experiment ($t = 2 \times 10^6$), cells have invaded far into the gradient in all replicates and even established themselves at the end of the gradient in some replicates. Persisting at these extremely nutrient-poor conditions requires extraordinary change in cell-cycle regulation. While initial cells obtained S expression once per cycle, the most commonly evolved cell in sector 11 of R8 requires S expression at least 97 times per cycle to complete replication of its genome ($L = 82$, $n \approx 0.84$). A slow cell cycle translates to long generation time, which means that evolution operates up to an order of magnitude slower in poor habitats relative to rich habitats.

During range expansion, genome size increases, even though larger genomes require longer to replicate. Coinciding with the early range expansion, average genome size rises from $L = 64$ to $L \approx 70$. In half of the replicates (e.g. R10 and R3), no further substantial genome expansion occurs. In the other half of the replicates however (e.g. R2 and R8), genomes grow up to $L \approx 80$ and larger. In R8, there is also transient genome expansion to $L > 100$ along the line of descent. Strikingly, the five replicates with the largest genomes reach the greatest population size in the end (figs. 2 and 3). Moreover, genome size is generally larger in nutrient-poor sectors which is remarkable because in those nutrient-poor conditions, selection for shorter genomes would seem to be more severe. Several times when a strain appears with an expanded genome, it propagates not only to poorer environments, but also back into the rich environments, outcompeting resident cells with smaller genomes. It is from this propagation to richer environments that a strain in R8 finally invades the poorest environment of sector 11 (which is connected to sector 1), showing that it evolved a very effective strategy to deal with all nutrient conditions. Altogether, these observations show that genome expansion is associated with and likely enables adaptation to nutrient-poor conditions.

Cell-Cycle Adaptation to Nutrient Condition

To quantify how individual cells have adapted to distinct and potentially highly fluctuating nutrient conditions along the gradient, we use phenotypic profiling (fig. 4). Single cells are tracked under different fixed nutrient levels, without a gradient or local cell density. During 10^4 timesteps,

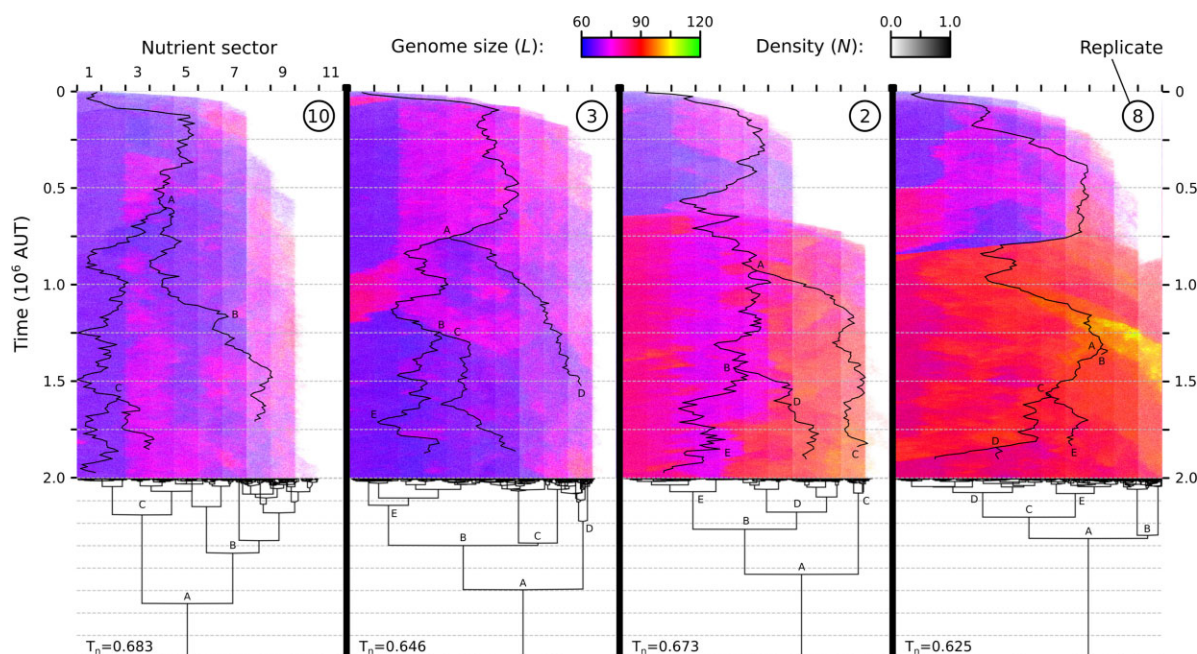


Fig. 2.—Replicate populations (R1–R10) invade and adapt to a nutrient gradient. Four replicates are shown, ordered from smallest to largest final population (see [Supplementary fig. S6, Supplementary Material](#) online for all ten replicates). Top: spacetime plots with population density and average genome size, bottom: cladogram of the final population at $t = 2 \times 10^6$, including the treeness statistic (T_n ; see “Evolutionary Signatures of Generalist and Specialist Strategies”). For the main ancestral lineages, their location on the gradient is shown on the spacetime plots with letters identifying the corresponding branching events in the cladogram. Each pixel row (1×550) in the spacetime plot is an aggregate across columns of the actual grid at a specific timepoint (50×550 , see [fig. 1](#)).

we measure in each condition the mean cell-cycle duration $\tau(n)$ and the fraction of cell cycles that results in proper division $\rho(n)$, and calculate from these a fitness as $R_0(n) = \rho(n)/[\delta \cdot \tau(n) + 1 - \rho(n)]$ (see [Appendix 2](#)).

Evolved cells with larger genomes (R2 and R8) execute faster cell cycles than cells with smaller genomes (R10 and R3) despite being larger. Provided that the cell does not reach M prematurely (which would decrease ρ), faster execution of the cell cycle is advantageous because it increases the division rate (since R_0 increases with decreasing τ). Phenotypic profiling thus confirms that genome expansion was adaptive and likely responsible for the previously noted range expansion.

Individual- versus Population-Level Adaptive Strategies

Phenotypic profiling revealed an additional unexpected difference among replicate experiments: different levels of control over the cell cycle evolved ([fig. 4](#)). In R3 and R8, the duration of the cell cycle is modulated by nutrient availability: a longer cell cycle is executed when conditions are poor, requiring more replication steps. Consequently in replicates where this phenotypic plasticity arises, individual cells demonstrate enlarged viable ranges (e.g. compare cells from R8 with those from R2 in [fig. 4](#)). In R10 and R2, cells did not evolve phenotypic plasticity and even show a slightly negative response to nutrient levels (see [fig. 3](#)).

Phenotypic profiling thus reveals two ecological strategies by which populations from different replicate experiments are adapted to the gradient: a generalist strategy in which cells mostly individually tune their cell cycle to the local conditions, and a specialist strategy in which the population has diversified into cells that are specialized to the different sectors of the gradient. Strikingly, these strategies seem to be independent from the level of genomic complexity in the population ([fig. 3](#)). Evolution has independently produced specialists with small genomes (R10), specialists with large genomes (R2), generalists with small genomes (R3), and generalists with large genomes (R8). The remaining six replicates are also found along these two axes (see [fig. 3](#)).

Evolutionary Signatures of Generalist and Specialist Strategies

Further analysis of the replicates showed that the specialist and generalist strategies which were identified in cell-cycle behavior can also be identified at the phylogenetic and genomic level. The specialist strategy is reflected by a deeper-branching phylogeny than the generalist strategy, as quantified by the *treeness* statistic T_n ([fig. 2](#); Spearman correlation of treeness versus mean plasticity, $T_n \times \sigma$: $r = -0.92$, $p = 0.0002$, $N = 10$). The treeness shows that specialists form distinct subspecies, whereas generalists resemble a single quasi-species.

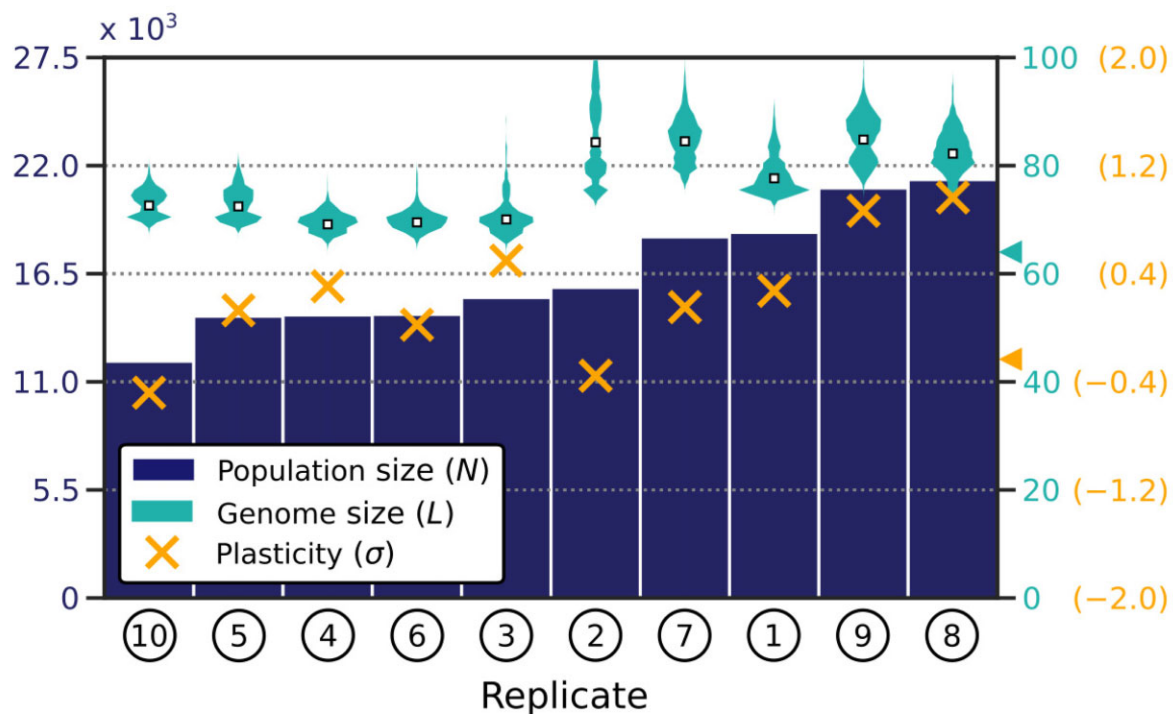


Fig. 3.—Populations that evolved independently from the same initial conditions show diverse outcomes in population size and average genome size. Replicates are ordered by final population size (dark blue bars), which reflects adaptation to poor nutrient conditions. Metrics are averaged across the entire gradient. Plasticity is quantified per cell as $\sigma = \log_{10} [\tau(0.1)/\tau(100)]$, that is, the measured cell-cycle duration at $n = 0.1$ relative to that at $n = 100$ (see fig. 4). Plasticity of the most common genotype of each sector was measured, and these were then averaged per replicate. Cyan and orange triangles on the right indicate genome size (64) and plasticity (-0.23) for the ancestor.

At the genomic level, the specialist strategy is reflected by stronger correlation between gene copy number variations and the spatial distribution of cells along the gradient compared to the generalist strategy (Supplementary fig. S2c, Supplementary Material online, $p = 0.034$). In R10 for instance, most cells encode a single copy of gene g6, but cells in sectors 1 and 2 mostly encode two copies (Supplementary fig. S3, Supplementary Material online). The same picture emerges when the combined variations in genetic properties such as the activation threshold or regulatory weight are analysed (Supplementary fig. S2d, Supplementary Material online, $p = 0.00044$). In sum, generalist and specialist strategies which ultimately lie encoded in evolved genomes and regulatory mechanisms, have an impact on eco-evolutionary dynamics of a population that reaches beyond the cellular phenotype.

Mechanisms of Cell-Cycle Adaptation to Nutrient Condition

In our replicate experiments, cells that experienced more genome expansion execute faster cell cycles, and generalist as well as specialist strategies emerge to deal with varying nutrient conditions. Our modeling framework offers a unique opportunity to investigate how evolved cells accomplish these behaviors mechanistically. To do this, we study

cell-cycle regulation at three different levels (fig. 5; see Supplementary fig. S7, Supplementary Material online for all ten replicates). First, dynamic insight into the cell cycle is provided by the state-space (top panels), which graphs expression states visited during the cell cycle and the likelihood of transitions between them. These states and transitions are determined empirically by tracking single cells for 10^4 timesteps (see “Cell-cycle Adaptation to Nutrient Condition”). Second, a concise picture of the regulatory architecture is provided by the network representation (middle panels), which depicts summed interactions between genes. Third, finer regulatory details are provided by the genome representation (bottom panels), which depicts individual interactions between genes and binding sites embedded in the genome. In the next sections, we discuss in terms of these three levels—cell-cycle dynamics, network topology and genome organization—how cells accomplish adaptation to nutrient conditions in different replicates.

Cell-Cycle Dynamics: Contiguous Versus Intermittent S-phase

For each replicate, state-space reveals how the ancestral cell cycle was modified by evolution to complete genome replication under limited nutrients, requiring more steps in S

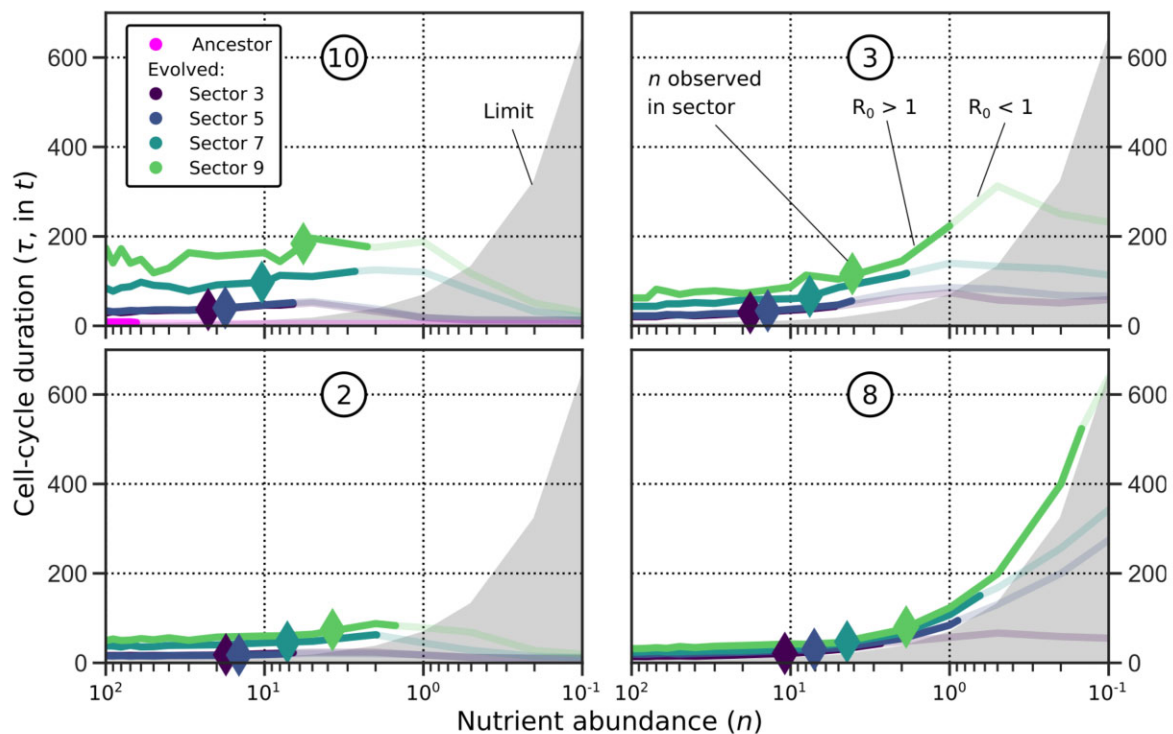


Fig. 4.—Phenotypic profiling of cells from four populations (R10, R3, R2, and R8). Cells with large genomes (bottom; R2 and R8) regulate faster cell cycles than those with smaller genomes (top; R10 and R3) in a given sector of the gradient. Because of this, populations in R2 and R8 grow denser, depleting nutrients further (diamonds have moved further to the right). Additionally, cells from R3 and R8 (right panels) time their cell cycle according to nutrient abundance. This is reflected in the convex nutrient curve and in the larger viable range of individual cells compared to R10 and R2. The grey area marks the minimal duration (τ_{\min}) of a complete cell cycle for a cell with the same genome size as the ancestor ($L = 64$). For all cells, R_0 drops below 1 before $\tau < \tau_{\min}$ because cells have larger genomes ($L > 64$) and because cell-cycle regulation is noisy ($\rho < 1$).

before reaching M (fig. 5, top panels). Besides the original cell cycle, cells exhibit an alternative cyclic trajectory which runs from S to S and by-passes M. By iterating through this new loop, which we term *S-loop*, cells stall their main cell cycle and execute additional replication steps.

Differences between the *S-loops* in state-space reveal how the fitter cells from R2 and R8 manage to execute faster cell cycles than cells from R10 and R3 (i.e. fig. 4). In R10, passage through the *S-loop* takes four timesteps, but genome replication only happens in one of these, that is, in S itself. Thus, cells spend most of their time in nonfunctional states during stalling of cell division. In R8 in contrast, the *S-loop* goes directly from S to S, so the cell executes replication contiguously. Under the same conditions, cells from R8 can suffice with an overall much shorter cell cycle than cells from R10. Thus, expansion of the regulatory repertoire allows cells to optimize their cell-cycle behavior supporting higher cell densities (i.e. as shown in fig. 3).

Evolution of the S-phase: Rewiring the Core Oscillator

To understand why a contiguous S-phase evolved in some replicates (e.g. R8) but not in others (e.g. R10), we turn to

the underlying regulatory networks (fig. 5, middle panels). Genome expansion and the resulting complex networks give rise to the more efficient *S-loops* of R2 and R8. Relative to the ancestor, multiple new genes and interactions arose in R2 and R8, whereas only one new gene arose in R10 and R3. Such genome expansion has resulted in adaptive rewiring of the regulatory core of the ancestral network.

In the ancestor, an oscillator composed of g1 and g5 drives both the cell cycle and the *S-loop*—the difference between these trajectories is in the activation of g3 by g5 which occurs downstream of the oscillator and hence does not interfere with the cyclic behavior. The g1/g5 oscillator is still intact in R10, yielding an *S-loop* with the same period as the main cell cycle. In R3 and R2, additions to the g1–g5 motif have accomplished shorter *S-loops*. In R8, the ancestral oscillator has even been entirely replaced by a new module of four genes. Thus, genome expansion was clearly adaptive: innovation of regulatory logic was necessary to organize an *S-loop* decoupled from the main cell cycle. Strikingly, the course of evolution is largely predetermined by early adaptations. Similar outcomes are found when the evolution experiments are repeated from

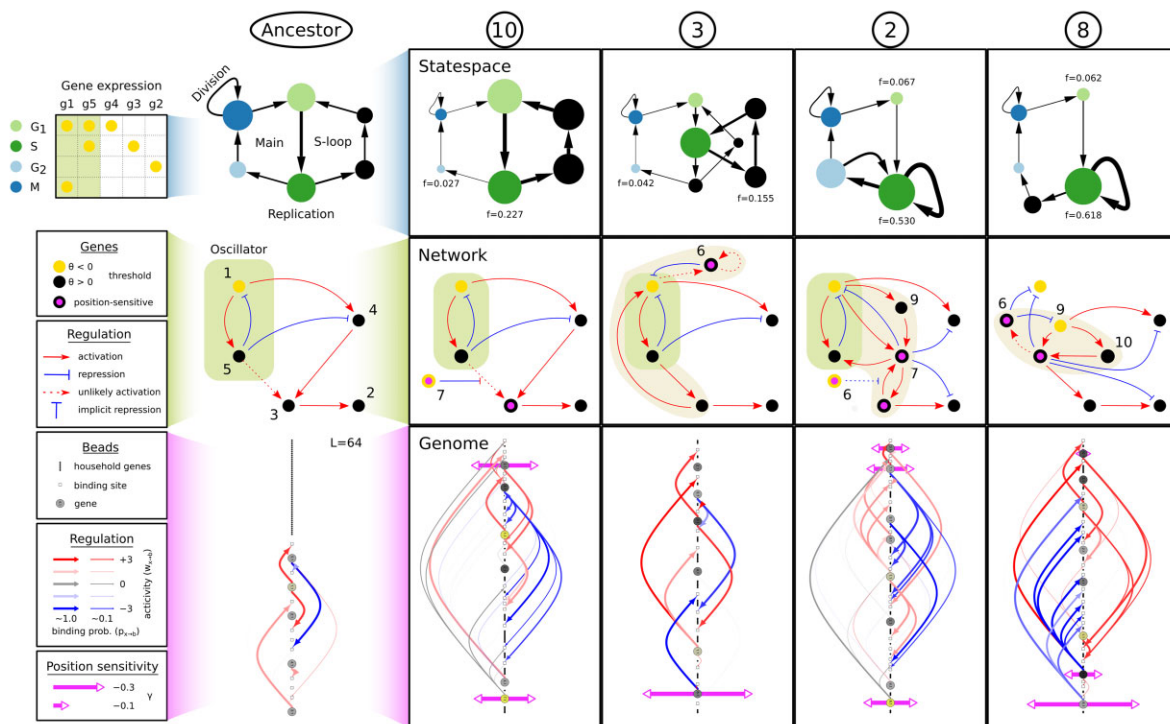


Fig. 5.—Mechanisms underlying cell-cycle adaptation at three different levels. For a representative from each replicate, the state-space (top panels), network (middle panels), and genome (bottom panels) are shown. The state-space only shows the most frequently observed states and transitions ($f > 0.02$). On the left, the ancestral cell for all replicates is shown, with the central oscillator highlighted in the networks. The additional paler shaded highlight in evolved networks of R3, R2, and R8 identifies other genes in the top level of the regulatory hierarchy. For each replicate, from the 11 most abundant cells per sector a network was chosen that represents the core network (see “Gene Family Analysis” in Appendix 1). The same views were made for the other six replicates (see [supplementary fig. S7, Supplementary Material](#) online). In the network panels, arrows with dashed lines represent unlikely interactions that result from low binding affinity; other interactions (arrows with solid lines) occur with a probability close to 1. Implicit repression, as seen in the networks of R10 and R2, refers to genes that bind with zero effect ($w = 0$) but which thereby prevent another gene with a positive effect from binding to the same binding site. Position-sensitive genes (indicated with pink horizontal arrows in the genome panel and a pink dot in the network panel), are derived from the analysis described in “Functional Genome Organization” and shown in [fig. 6](#).

2.5×10^5 timesteps onwards, which is prior to the final range expansions ([Supplementary fig. S8, Supplementary Material](#) online). While the range expansions occur at different time points, the final population sizes are very similar to the original replicate.

Beyond Regulatory Logic: Cell-Cycle Timing

The evolved regulatory architecture in each replicate ([fig. 5](#), middle panels) is conserved throughout the gradient (see “Gene Family Analysis” in Appendix 1). Specialists and generalists employ this single regulatory blueprint to obtain a wide range of cell-cycle timings across the gradient using genomic speciation and phenotypic plasticity, respectively. Through extensive investigation, we could show that network topology and genome organization operate in concert to achieve fine-tuning of the cell cycle to nutrient conditions ([fig. 5](#), middle and bottom panels).

Emergence of Functional Genome Organization

Given the coherence at the network level, variation in cell-cycle timings suggested that other levels such as the genome might be involved. To test this, we performed a gene swap experiment ([fig. 6](#); Methods). For a given genome, we created mutants which have the location of two genes swapped, for all pairs of genes. Most mutants have reduced fitness compared to the wildtype which can be attributed to 1–4 genes that are located close to the origin or close to the terminus in the wildtype ([fig. 5](#)). In line with this pattern, the genomic location of these genes is more conserved than that of other core genes ([Supplementary fig. S2b, Supplementary Material](#) online, $p = 2.6 \times 10^{-7}$, see Appendix 1).

Genes that display high sensitivity to genomic location have an important role in the regulatory network: control of cell-cycle progression. They accomplish this control through a weak regulatory interaction that is sensitive to gene dosage. Consequently, such genes link genome

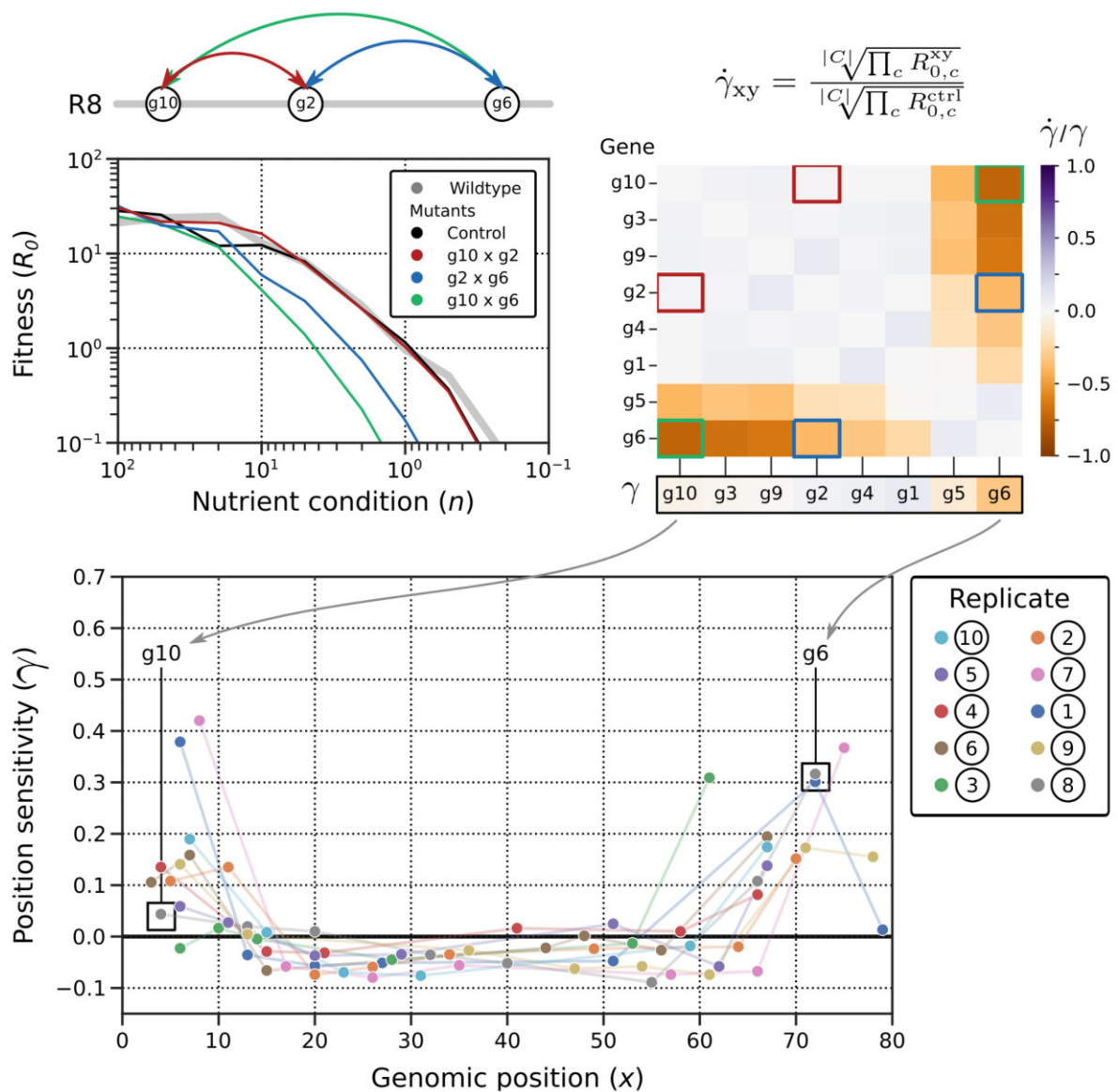


Fig. 6.—Gene swap experiment reveals that genome organization is functional, in particular for specific genes at the genomic extremities (see text). Top left panel shows the measured R_0 -curves for the representative cell from R8 (cf. fig. 5), the swap control and three swap mutants. The top right shows the matrix with relative fitness of swap mutants compared to the swap control; the three mutants shown on the left are highlighted. From this matrix, relative “position-sensitivity” of each gene is calculated (added row below matrix). In the bottom panel, position-sensitivity is shown for each gene from the 10 representative genomes for R1–10. Each genome harbors at least one gene whose position is crucial for correct cell-cycle behavior.

organization, which determines when each gene is replicated, to regulatory dynamics, that is, the likelihood of their interactions. In R8, g6 activation by g5 provides the cue for cell-cycle progression. Since g5 affinity for g6 binding sites is weak, the cell cycle is stalled in the S-loop long enough, on average, to allow replication to complete. Moreover, the position of g5 and g6 on the genome near the terminus favors g6 activation and progression late in the cell cycle, when replication is nearly finished. In general, dosage-sensitive genes that need to be active early in the cell cycle are located near the origin of replication (e.g. g3 in R10),

whereas dosage-sensitive genes that need to be active late are located near the terminus (e.g. g5 and g6 in R8). Thus, a simple physical feature like replication-induced change in gene dosage is exploited by an emergent coupling between genome organization and regulatory architecture.

Generalists: from Noisy Control to Cell-Cycle Checkpoint

All replicates evolved stochastic control over cell-cycle progression as well as a functional genome organization to

enhance this control. Generalists accomplish phenotypic plasticity through an even tighter coupling between genome and network. The strong response to replication status is achieved by two mechanisms: first, through nonlinearity between the cue for progression and progression itself; second, by using gene activation (g6 in R3 and R8) rather than inactivation (g3 in R10, g7 in R2) as the cue. These two features give rise to behavior that resembles a checkpoint: in the representative cell from R8, cell-cycle progression is nearly four times more likely at the end of replication than it is at the start.

Nonlinearity between the cue for progression and progression is achieved in at least two ways. In R3 and R8, it presents itself in a timewise manner. The cue for progression, g6 activation, has to be obtained two or three consecutive timesteps in order to steer the cell towards G2. If g6 is only activated once, the cell will revert to the S-loop. As a consequence, when replication increases the likelihood of g6 activation, the likelihood of progression depending on two or three such activations increases quadratically or cubically. Alternatively, in R4, nonlinearity arises from the promoter architecture (fig. 7): activation of g6—which immediately triggers progression—requires binding of g1 and g4 at two different binding sites constituting an AND-gate. When replication increases the probability of each binding event, the likelihood of progression depending on two such events increases quadratically.

As mentioned, the particular regulatory motif that functions as progression cue (i.e. gene activation or inactivation) also affects how much cell-cycle control can be acquired through genome organization. These observations show that, while genome organization is always relevant whenever low binding affinities play a role, the complex generalist behavior additionally requires specific network topology to achieve the finer integration between replication and regulation.

Evolution of a Cell-Cycle Checkpoint

To investigate how the generalist strategy evolved, we focus on one of the replicate experiments, R4, whose checkpoint we mechanistically fully understand as it arose in three separate stages (fig. 7). The invention of g6, constituting the first stage, improved cell cycle timing by competitively inhibiting binding of g5 to g3. Plasticity increases because the newly evolved g6 is located at the end of the genome which enhances its regulatory role—promotion of cell-cycle progression. Even more plasticity, that is, generalism, appears in the second stage, when cells switch to activation as cue for progression. Implicit repression of g3 by g6 becomes explicit through subfunctionalization of two binding sites of g3. In addition, g6 activation by g4 becomes less frequent, replacing binding competition for g3 as timing mechanism. In the third stage, the nonlinear response of cell-cycle

progression to replication status emerges as g6 activation becomes dependent on weak binding of both g1 and g4. During each stage, reorganization of the genome coincides with (stage I) or rapidly follows (stages II and III) the novel regulatory architecture, underscoring their interplay.

Specialists: Functional Genomic Speciation

Specialists rely on genomic speciation for adaptation to the gradient. To understand how genomic speciation allows for a wide range of cell-cycle timings, we investigated the previously identified genotypic variation within specialist populations in the context of regulatory topology (see “[Gene Family Analysis](#)” in Appendix 1). Cell-cycle timing is set by the competition between two regulatory genes for a critical target gene (g3 in R10, g7 in R2). Activation of the target gene by an activator (g5 in R10, g3 in R2) leads to stalling of the cell cycle in the S-loop; inactivation through hindrance of the activator by an implicit inhibitor (g6 in R10 and R2) triggers progression. This competitive balance is tuned to the local nutrient abundance through variation in various genetic properties. For instance, cells in nutrient-poor sectors of R2 encode an additional g7 binding site where competition between the activator and inhibitor takes place (Supplementary fig. S3, Supplementary Material online). This promotes long cell cycles, since cell-cycle progression by hindrance of the g7-activator (g3) at all binding sites is less likely. Similarly, short cell cycles are favored in nutrient-rich sectors of R10 by means of a second copy of the implicit g3-inhibitor (g6), shifting the competitive balance towards g3 inactivation and cell-cycle progression. Because cell-cycle timing is achieved with a critical noisy interaction, the same regulatory logic can easily be exploited in different conditions by tuning of that interaction. Instead of evolving phenotypic plasticity to cope with different nutrient conditions, specialists increased their evolvability to allow rapid adaptation to those different conditions.

Discussion

We reformulated a minimal model of the *C. crescentus* cell cycle to study the evolution of complex gene regulation. Our modeling framework is simple, yet allows for interesting emergent phenomena such as the evolution of genome organization, generalist, and specialist strategies, phenotypic plasticity, and cell-cycle checkpoints.

Genome Expansion by Adaptive Forces

We have demonstrated how key cell-cycle adaptations can be linked to the expansion of genomes and their encoded regulatory networks, which was exploited to organize replication more efficiently in an S-phase. In previous evolutionary models (Soyer and Bonhoeffer, 2006; Cuypers

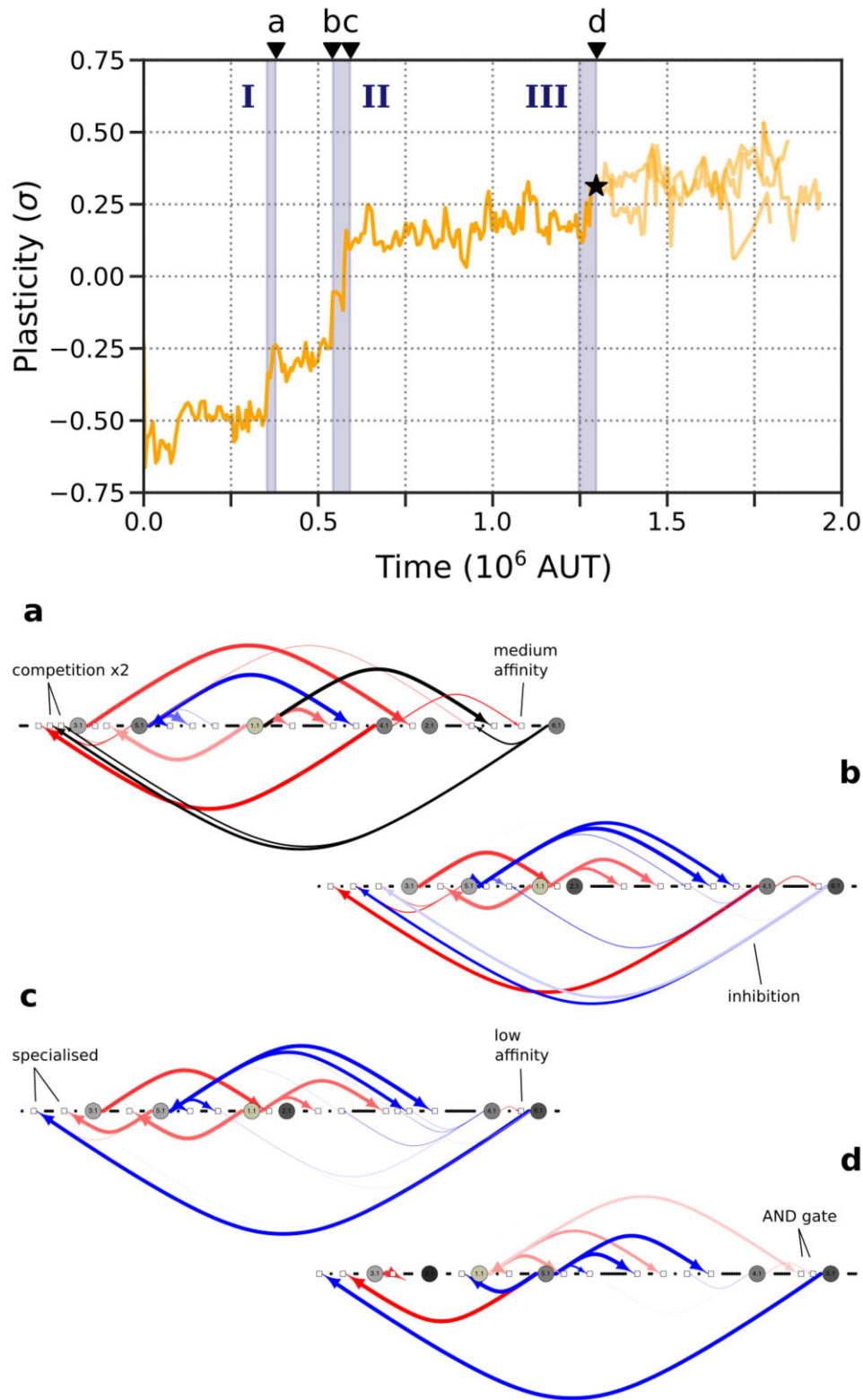


Fig. 7.—Emergence of plasticity in the generalist replicate R4. Along the ancestral lineage, three phases can be distinguished: (I) g6 appears and starts competing with (hindering) g3 activation by g5. (II) g6 starts inhibiting g3 directly and the two binding sites of g3 become specialized for g5 and g6. (III) g1 starts activating g3 and the AND-gate is formed. Below the time panel, four indicative genomes along the ancestral lineage are shown (birth times indicated by the letters at time points a, b, c, d).

and Hogeweg, 2012, 2014), network or genome expansion was largely driven by nonadaptive forces—although genome expansions in Cuypers and Hogeweg (2012) were eventually adaptive. In these models, the cost of genome expansion was low (Soyer and Bonhoeffer, 2006) or absent (Cuypers and Hogeweg, 2012, 2014), so neutral processes could easily overcome the weak purifying selection to increase cellular complexity. The explicit incorporation of replication into our model imposes on cells a much stronger selection against genome expansion. As a result, neutral processes play a much smaller role.

In literature, population size is often considered to be a crucial determinant for genome size evolution (Lynch and Conery, 2003). In contrast to Lynch and Conery (2003), we here see a positive correlation between population size and genome size because population size reflects adaptation which is achieved by genome expansion.

In Appendix 1, we describe additional experiments to investigate the sensitivity of our results to various model features (“Model Testing”, Appendix 1). Importantly, in our setup there is a slight mutational bias towards genome expansion due to the innovation rate. Without innovations, as expected from such lack of novel genetic material, populations have more trouble adapting to poor nutrient conditions. Yet, 3 out of 10 replicate populations achieve extensive range expansion comparable to R8 and R9 of the main model, and this again coincides with genome expansion (Supplementary fig. S4, Supplementary Material online). Other experiments using modifications of the main model further support the dominant role of adaptive forces in the observed genome expansion.

Emergence of No-Cost Generalists

To cope with different nutrient conditions, generalist and specialist strategies emerged in different replicate experiments. In both strategies, a single noisy regulatory motif is responsible for tuning of cell-cycle timing. Specialists employed hard-coded genomic variation of this motif—such as extra gene copies—to adapt to different conditions. Generalists exploited the interplay between regulation and genome organization to upgrade the noisy motif into a *de novo* cell-cycle checkpoint. The checkpoint acts to ensure that cells do not finish their cell cycle until replication is finished. Thus, range expansion was achieved through genomic speciation or by phenotypically plastic individuals.

It is due to the multiple organizational levels in our model that we find such diverse behavior. Considered separately, neither genome nor regulatory network defines the complex adaptive behavior of generalists (R8, R9, R3, and R4). Rather, generalist behavior emerges from the interplay between these organizational levels. In contrast, the efficient S-phase behavior found in half of the replicates (R2, R7, R1, R9, and R8) can be determined from the regulatory

network alone, that is, by the size of the core cyclic module. These independent forms of adaptive behavior demonstrate the power of a multilevel view for understanding the evolution of complex patterns in biology.

Whether generalism is an evolutionary stable strategy has been questioned (Dennis et al., 2011; Loxdale et al., 2011; Sriswasdi et al., 2017), based on the ideas that generalism is costly (e.g. requiring a larger genome) and that selection eventually steers adaptation to one primary niche (e.g. the local environment). However, the phenotypic plasticity that emerges in our model is an example of “no-cost generalism” (Remold, 2012) as it does not require cells to have a larger genome. Generalists also turn out to be slightly more successful between replicates, exemplified by the two replicates with the largest final populations, R8 and R9. We have thus shown mechanistically how no-cost generalism can emerge, and that this is a stable strategy rather than just an evolutionary intermediate state.

Genome Organization is Integrated into the Cell Cycle

One of the striking results from our model is the consistent evolution of functional genome organization in all replicates. Specific genes evolved to be located such that their time of replication aligns with their time of activity, that is, their role in cell-cycle regulation. Similar patterns have been observed in prokaryotes. In *C. crescentus*, the genes coding for DnaA, GcrA, CtrA, and CcrM are positioned sequentially on the genome coinciding with their sequential activation during the cell cycle (Collier et al., 2007; Seong et al., 2021). In *E. coli*, where gene copy number distribution is dominated by the growth cycle, transcriptomic data has revealed that the position of many genes correlates with their time of expression during the growth cycle (Sobetzko et al., 2012). Yet even certain binding sites for DnaA in *E. coli* have been shown to play a position-dependent role in cell-cycle regulation (Frimodt-Møller et al., 2017).

The observation that genome organization can be important for cellular function is not new (Li et al., 2008; Shen et al., 2008; Walker et al., 2016; Jaruszewicz-Błońska and Lipniacki, 2017). However, we have here gained insight into the evolutionary context—genome organization is maintained despite frequent shuffling mutations—and into the mechanisms that allow cells to exploit the interplay between genome organization and cell-cycle behavior. For instance, it is not *a priori* clear that cells can use cell-cycle-induced changes in gene dosage to their advantage (Paijmans et al., 2016). In our model, only generalists manage to evolve a cell-cycle checkpoint that gives rise to phenotypic plasticity. Two particularly interesting mechanisms used by generalists to increase sensitivity of cell-cycle behavior to gene copy number were the integration of multiple signals (binding events) in an AND-gate configuration (cf. Hermsen et al., 2006), and the integration of a

single signal over multiple consecutive timepoints (cf. Liu et al., 2015; Malaguti and Ten Wolde, 2021).

Prokaryotes also have complex biochemistry available to exploit the physical genome organization. Cooperative binding of transcription factors to DNA (Hermsen et al., 2006) or regulation beyond the level of transcription (Bryant et al., 2014; Frimodt-Møller et al., 2017; Krogh et al., 2018) would only seem to provide more routes for evolution to couple replication to regulation. In *C. crescentus*, chromatin states are used to control gene expression during the cell cycle (Collier et al., 2007; Seong et al., 2021). DNA is normally fully methylated but becomes hemi-methylated after replication. Promoters of genes may be designed to activate only in the hemi-methylated state—for example, *ctrA*, whose expression is triggered upon replication—or in the fully methylated state—for example, *dnaA*, whose expression collapses upon replication. Aside from these powerful biochemical processes, it is remarkable that functional genome organization already evolves in our relatively simple model.

In conclusion, our results show that the evolution of gene regulation cannot be understood from the architecture of the gene regulatory network alone. The interplay between genome organization, regulatory network, and cell-cycle behavior is as important as any of these levels on its own.

Materials and Methods

We study the evolution of regulatory networks in an individual-based model embedded on a grid (fig. 1). Each cell in a population consists of a genome with regulatory genes (or genes, for short) and binding sites, that together form a regulatory network. Expression of five core gene types (g1–5) defines the cellular state (\bar{s}). Reproduction takes place when a cell completes a trajectory through four states that represent the *C. crescentus* cell cycle (taken from the model of Quiñones-Valles et al., 2014; Sánchez-Osorio et al., 2017).

The Cell Cycle

Quiñones-Valles et al. (2014) derived from literature a Boolean network for cell-cycle regulation in the model bacterium *C. crescentus*. With five core regulatory genes (CtrA, GcrA, DnaA, CcrM, and SciP), the network already yields a cyclic attractor of four states, resembling a cell cycle. We use these states—which we will denote G1, S, G2, and M—to define the cell cycle in our model.

When a cell acquires M expression, one of two things happens. If the cell has progressed through all previous stages, it divides. A new cell is formed at an adjacent site on the grid, killing if necessary the cell that occupied that location. If the cell reaches M prematurely however, it dies. For proper cell-cycle progression the order of stages is important, but they do not have to follow on one another directly. A cell in G1 can reach any number of alternative states or even perform the expression that defines G2.

This merely stalls its cell cycle until it acquires the expression corresponding to the next stage (S in this case).

One of the key and most time-consuming events in the prokaryotic cell cycle is replication of DNA. The explicit incorporation of genome replication in our model leads to a natural constraint on genome size: an information cost. Cells replicate a piece of their genome (n beads, where n is the nutrient abundance) every timestep they spend in stage S. The copied beads participate in the regulatory dynamics. With time spent in S, the copy number of each gene increases from one to two, impacting the regulatory dynamics (see next section).

Initially, we evolve individuals with the basic cell-cycle circuit under unlimited nutrients ($n = \infty$), such that the entire genome is replicated in a single step (see “Evolution of *C. crescentus* Cell cycle under Stochastic Gene Expression”). This allows cells with the Quiñones-Valles et al. (2014) regulatory network, which spend only one step in S, to successfully complete their cell cycle. Afterwards, we expose cells to nutrient limitation, which requires them to express multiple replication steps (S) in one cell cycle (see Results). Cells are inoculated on a nutrient gradient ($n_{\text{tot}} = [2, 80]$) and nutrients are locally depleted by other cells, so that a site with x neighbors has a nutrient level $n = n_{\text{tot}}(1 - x/8)$. For cells with the maximum of eight neighbors, no replication is possible ($n = 0$) imposing an upper limit to population size even without instantaneous death rate δ .

Due to real-time genome replication and nutrient limitation, there is a strong cost to genome size. Cells with a smaller genome require fewer replication steps, which allows them to complete their cell-cycle faster and produce more offspring. Note that regulating more replication steps than required (i.e. when the genome is already completely replicated) does not harm the cell except that it costs time, decreasing reproduction rate. Besides regulatory components, cells need to maintain 50 household genes in their genome. When a cell cycle is completed, the reproduction probability is reduced by $|n_h - 50|/10$ where n_h is the number of household genes.

The Genome

Genes, binding sites and household genes are organized on a linear beads-on-a-string genome with an origin of replication and a terminus at the two opposite ends (cf. Crombach and Hogeweg, 2007). A gene is expressed if the sum of regulatory effects at its upstream binding sites reaches its specific activation threshold θ :

$$\epsilon_i = \begin{cases} 1 & \text{if } \sum_b w_{x \rightarrow b} > \theta_i, \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

The regulatory effect $w_{x \rightarrow b}$ results from the binding of expressed gene x to binding site b . This effect depends on

the regulatory weight of the binding site and the bound gene:

$$w_{x \rightarrow b} = w_x \cdot w_b \quad (2)$$

Only one gene may bind per binding site, and this is (re-)determined stochastically every timestep. All genes that are expressed at that moment have a probability to bind, based on their affinity for the binding site. There is always a probability that no gene binds, the relative propensity of which is set to 1:

$$p_{x \rightarrow b} = \frac{\epsilon_x k_0 e^{S(x,b)}}{1 + \sum_y^{n_g} \epsilon_y k_0 e^{S(y,b)}} \quad (3)$$

Here, $S(x, b)$ is the similarity (number of matching bits) between the binding sequences of gene x and site b (bitstrings of length $l = 20$); n_g is the total number of genes in the genome, and ϵ_x is the expression status of gene x (0 or 1). Note that when a gene has been replicated, both copies may be expressed and get a chance to bind to the binding site (see binding probability, green curve in fig. 1). Thus, gene products are incorporated implicitly. Furthermore, if the binding site and its upstream gene have also been replicated, both copies of the locus have a chance to become regulated. The parameter k_0 was set to $k_0 = 1.0 \times 10^{-7}$ because this leads to low average affinity for two random sequences, strong binding for two very similar sequences ($p_{x \rightarrow b} \rightarrow 1.0$), and an intermediate region where affinity takes on values in between these two extremes.

Mutations

Mutations happen at the level of the genome. Upon division of a cell, genes, binding sites, and household-genes in the new daughter cell can undergo various mutations (table 1).

First, the genetic properties of beads can be mutated. With probability μ_B per position, a bit in the binding sequence \vec{B} is flipped. With probabilities μ_w and μ_θ per bead, the regulatory weight w and the activation threshold θ are mutated, respectively (μ_θ only applies to genes). The

Table 1. Parameter Values used in the Evolution Experiments

Parameter	Value	Description
k_0	1.0×10^{-7}	Basal binding probability
l_B	20	Length of binding sequence
μ_B	1.0×10^{-4}	Per-bit mutation rate of binding sequence
μ_w	5.0×10^{-4}	Regulatory weight mutation rate
μ_θ	5.0×10^{-4}	Activation threshold mutation rate
μ_{dup}	5.0×10^{-4}	Per-bead duplication rate
μ_{del}	5.0×10^{-4}	Per-bead deletion rate
μ_{rel}	1.0×10^{-3}	Per-bead relocation rate
$\mu_{b,in}$	5.0×10^{-3}	Per-genome innovation rate for binding sites
$\mu_{g,in}$	5.0×10^{-4}	Per-genome innovation rate for genes

value of θ or w is then incremented or decremented (probability 0.2) or randomly redrawn from the range $[-3, 3]$ (probability 0.8).

Second, each bead has a probability to be duplicated (μ_{dup}), deleted (μ_{del}), or relocated (μ_{rel}). The new location of a duplicated or relocated bead is random. When genes are duplicated, deleted or relocated, adjacent upstream (“proximal”) binding sites move with them.

Third, there is a per-genome innovation rate for binding sites ($\mu_{b,in}$) and for genes ($\mu_{g,in}$). These result in the creation of a new bead of the respective type with random properties and at a random location in the genome.

As mentioned, the cellular state \vec{s} is defined by the expression of the five genes originally modeled by Quiñones-Valles et al. (2014). We do not want to constrain the regulation of these transcription factors themselves during *in silico* evolution. Therefore, we assign gene types t to unique combinations of binding sequence \vec{B} and regulatory weight w (these parameters define the gene product; in contrast, θ is a property of the locus) where g1–5 denote CtrA, GcrA, DnaA, CcrM, and SciP. Genes keep their identity even if \vec{B} or w changes. New gene types are only assigned upon gene innovation or when a duplicated gene diverges from the other copy and establishes a new unique combination of \vec{B} and w .

Gene Swap Experiment

A gene swap experiment was used to assess the importance of genome organization and to pinpoint genes whose position is crucial to cellular fitness. First, a wildtype genome is constructed from the original by placing all binding sites directly adjacent to their downstream gene. Then, genes are spaced from one another such that the maximal number of binding sites of any gene can fit in between. In the resulting wildtype, genes together with their binding sites can be swapped without displacing/moving any of the other genes in the genome. For each pair of genes, a mutant is created with the genomic location of those two genes swapped. Then, in addition to the wildtype, the fitness of each of these mutants is estimated in a range of fixed nutrient conditions $C = \{100, 50, 20, 10, 5, 2, 1, 0.5, 0.2, 0.1\}$:

$$R_{0,c}^{gt} = \max\left(\frac{\rho_c}{\delta \cdot \tau_c + 1 - \rho_c}, 0.001\right) \quad (4)$$

Here the genotype gt stands for the wildtype or any mutant xy with genes x and y swapped. The fitness of each mutant across all conditions relative to the wildtype is calculated by comparing the harmonic mean:

$$\dot{\gamma}_{xy} = \frac{|\mathcal{C}| \sqrt{\prod_c R_{0,c}^{xy}}}{|\mathcal{C}| \sqrt{\prod_c R_{0,c}^{wt}}} \quad \text{with } c \in \mathcal{C} \quad (5)$$

Then, to find the relative sensitivity to relocation of a particular gene, its contribution to the fitness reductions in all its swaps is calculated. Basically, the off-diagonal xy -elements in the matrix γ are normalized by the average of the column (all contributions of gene y), and then averaged per row (all contributions of gene x):

$$\gamma_x = \left\langle \log_{10} \left(\frac{\dot{\gamma}_{xy}}{\langle \dot{\gamma}_{zy} \rangle_z} \right) \right\rangle_y \quad \text{where } \langle \dot{\gamma}_{xy} \rangle_y = \sum_y \dot{\gamma}_{xy} / Y \quad (6)$$

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgements

The authors gratefully acknowledge the help of Jan Kees van Amerongen for running the local computer cluster, Bram van Dijk for his advice during writing of the model's code, and Rutger Hermsen for his advice on a suitable physical function for the binding probability between a gene and a binding site.

Data Availability

The data underlying this article were generated from our computational model which is freely available at <https://github.com/samvonderdunk/Prokaryotes>.

Literature Cited

- Baym M, Lieberman TD, Kelsic ED, Chait R, Gross R, Yelin I, Kishony R. 2016. Spatiotemporal microbial evolution on antibiotic landscapes. *Science* 353(6304):1147–1151.
- Bryant JA, Sellars LE, Busby SJ, Lee DJ. 2014. Chromosome position effects on gene expression in *Escherichia coli* k-12. *Nucl Acids Res*. 42(18):11383–11392.
- Collier J, McAdams HH, Shapiro L. 2007. A dna methylation ratchet governs progression through a bacterial cell cycle. *Proc Natl Acad Sci*. 104(43):17111–17116.
- Cordero OX, Hogeweg P. 2006. Feed-forward loop circuits as a side effect of genome evolution. *Mol Biol Evol*. 23(10):1931–1936.
- Crombach A, Hogeweg P. 2007. Chromosome rearrangements and the evolution of genome structuring and adaptability. *Mol Biol Evol*. 24(5):1130–1139.
- Cuyppers TD, Hogeweg P. 2012. Virtual genomes in flux: an interplay of neutrality and adaptability explains genome expansion and streamlining. *Genome Biol Evol*. 4(3):212–229.
- Cuyppers TD, Hogeweg P. 2014. A synergism between adaptive effects and evolvability drives whole genome duplication to fixation. *PLoS Comput Biol*. 10(4):e1003547.
- Dennis RL, Dapporto L, Fattorini S, Cook LM. 2011. The generalism–specialism debate: the role of generalists in the life and death of species. *Biol J Linn Soc*. 104(4):725–737.
- Fay J, Wittkopp P. 2008. Evaluating the role of natural selection in the evolution of gene regulation. *Heredity* 100(2):191–199.
- Frimodt-Møller J, Charbon G, Løbner-Olesen A. 2017. Control of bacterial chromosome replication by non-coding regions outside the origin. *Curr Genet*. 63(4):607–611.
- Gilad Y, Oshlack A, Rifkin SA. 2006. Natural selection on gene expression. *Trends Genet*. 22(8):456–461.
- Hermsen R, Tans S, Ten Wolde PR. 2006. Transcriptional regulation by competing transcription factor modules. *PLoS Comput Biol*. 2(12):e164.
- Jaruszewicz-Błońska J, Lipniacki T. 2017. Genetic toggle switch controlled by bacterial growth rate. *BMC Syst Biol*. 11(1):1–11.
- Jenkins DJ, Stekel DJ. 2008. Effects of signalling on the evolution of gene regulatory networks. In: Bullock S, Noble J, Watson R, Bedau MA, editors. *Artificial life XI – proceedings of the eleventh international conference on the simulation and synthesis of living systems*. Cambridge, MA: The MIT Press. p. 289–296.
- Jenkins DJ, Stekel DJ. 2010. Stochasticity versus determinism: consequences for realistic gene regulatory network modelling and evolution. *J Mol Evol*. 70(2):215–231.
- Krogh TJ, Møller-Jensen J, Kaleta C. 2018. Impact of chromosomal architecture on the function and evolution of bacterial genomes. *Front Microbiol*. 9:2019.
- Li S, Brazhnik P, Sobral B, Tyson JJ. 2008. A quantitative study of the division cycle of *Caulobacter crescentus* stalked cells. *PLoS Comput Biol*. 4(1):e9.
- Liu X, Wang X, Yang X, Liu S, Jiang L, Qu Y, Hu L, Ouyang Q, Tang C. 2015. Reliable cell cycle commitment in budding yeast is ensured by signal integration. *Elife* 4:e03977.
- Loxdale HD, Lushai G, Harvey JA. 2011. The evolutionary improbability of ‘generalism’ in nature, with special reference to insects. *Biol J Linn Soc*. 103(1):1–18.
- Lynch M. 2007. The evolution of genetic networks by non-adaptive processes. *Nat Rev Genet*. 8(10):803–813.
- Lynch M, Conery JS. 2003. The origins of genome complexity. *Science* 302(5649):1401–1404.
- Malaguti G, Ten Wolde PR. 2021. Theory for the optimal detection of time-varying signals in cellular sensing systems. *Elife* 10:e62574.
- Pajmans J, Bosman M, Ten Wolde PR, Lubensky DK. 2016. Discrete gene replication events drive coupling between the cell cycle and circadian clocks. *Proc Natl Acad Sci*. 113(15):4063–4068.
- Pelve EA, Lindås A-C, Martens-Habbena W, Stahl DA, Bernander R. 2011. Cdv-based cell division and cell cycle organization in the thaumarchaeon *Nitrosopumilus maritimus*. *Mol Microbiol*. 82(3):555–566.
- Quayle AP, Bullock S. 2006. Modelling the evolution of genetic regulatory networks. *J Theor Biol*. 238(4):737–753.
- Quiñones-Valles C, Sánchez-Osorio I, Martínez-Antonio A. 2014. Dynamical modeling of the cell cycle and cell fate emergence in *Caulobacter crescentus*. *PLoS ONE* 9(11):e111116.
- Remold S. 2012. Understanding specialism when the jack of all trades can be the master of all. *Proc R Soc B: Biol Sci*. 279(1749):4861–4869.
- Romero IG, Ruvinsky I, Gilad Y. 2012. Comparative studies of gene expression and the evolution of gene regulation. *Nat Rev Genet*. 13(7):505–516.
- Sánchez-Osorio I, Hernández-Martínez CA, Martínez-Antonio A. 2017. Modeling asymmetric cell division in *Caulobacter crescentus* using a boolean logic approach. In: Tassan JP, Kubiak JZ, editors. *Asymmetric cell division in development, differentiation and cancer. Results and problems in cell differentiation*. Vol. 61. Cham: Springer. p. 1–21.
- Seong HJ, Han S-W, Sul WJ. 2021. Prokaryotic dna methylation and its functional roles. *J Microbiol*. 59(3):242–248.
- Shen X, Collier J, Dill D, Shapiro L, Horowitz M, McAdams HH. 2008. Architecture and inherent robustness of a bacterial cell-cycle control system. *Proc Natl Acad Sci*. 105(32):11340–11345.

- Sobetzko P, Travers A, Muskhelishvili G. 2012. Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle. *Proc Natl Acad Sci.* 109(2):E42–E50.
- Soyer OS, Bonhoeffer S. 2006. Evolution of complexity in signaling pathways. *Proc Natl Acad Sci.* 103(44):16337–16342.
- Sriswasdi S, Yang C-c., Iwasaki W. 2017. Generalist species drive microbial dispersion and evolution. *Nat Commun.* 8(1):1–8.
- Tan MH, Kozdon JB, Shen X, Shapiro L, McAdams HH. 2010. An essential transcription factor, SciP, enhances robustness of *Caulobacter* cell cycle regulation. *Proc Natl Acad Sci.* 107(44):18985–18990.
- van Nimwegen E. 2003. Scaling laws in the functional content of genomes. *Trends Gen.* 19(9):479–484.
- Walker N, Nghe P, Tans SJ. 2016. Generation and filtering of gene expression noise by the bacterial cell cycle. *BMC Biol.* 14(1):1–10.
- Whitehead A, Crawford DL. 2006. Neutral and adaptive variation in gene expression. *Proc Natl Acad Sci.* 103(14):5425–5430.
- Zheng W, Gianoulis TA, Karczewski KJ, Zhao H, Snyder M. 2011. Regulatory variation within and between species. *Annu Rev Genom Hum Genet.* 12:327–346.
- Zhou B, Schrader JM, Kalogeraki VS, Abeliuk E, Dinh CB, Pham JQ, Cui ZZ, Dill DL, McAdams HH, Shapiro L. 2015. The global regulatory architecture of transcription during the *Caulobacter* cell cycle. *PLoS Genet.* 11(1):e1004831.

Associate editor: Laurence Hurst