THESIS

SLOW AND NOISY: DEVELOPMENTAL TIME AND GENE EXPRESSION KINETICS IN BIG CELLS

Submitted by

Alexandra Taylor

Department of Biology

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2023

Master's Committee:

Advisor: Rachel Mueller

Co-Advisor: Ashok Prasad

Kim Hoke

Diego Krapf

ABSTRACT

SLOW AND NOISY: DEVELOPMENTAL TIME AND GENE EXPRESSION KINETICS IN BIG CELLS

Evolutionary increases in genome size, cell volume, and nuclear volume have been observed across the tree of life, with positive correlations documented between all three traits. It is well documented that developmental tempo slows as genomes, nuclei, and cells increase in size, yet the driving mechanisms are poorly understood. Meanwhile, the dramatic increases in cell volume seen across the tree of life pose interesting questions about a potential relationship between cell volume and stochastic noise at the single cell level, but this remains an underexplored area of research. To bridge these knowledge gaps, we use a mix of deterministic and stochastic, as well as species-specific and more general, models of the somitogenesis clock. In doing so, we explore the impact of changing intra-cellular gene expression kinetics induced by increasing genome size, nuclear volume, and cell volume on developmental tempo and gene expression noise. Results suggest that longer transcriptional and nuclear export times act to slow cell and developmental processes down as genome size and cell volume increase, and that "search processes" undergone by gene products within a cell become noisier with increasing volume. Analyses of stochastic model simulations and existing empirical data bring into question whether or not cell-autonomous oscillations can truly exist in the absence of cell-to-cell signaling.

TABLE OF CONTENTS

Introduction

Across the tree of life, genome size and cell volumes span a remarkable range, and a positive correlation has been observed between increases in genome size, cell volume, and nuclear volume (Gregory 2001, Malerba and Marshall 2021, Sessions 2008). However, the mechanisms underlying these relationships and the implications associated with such increases remain areas of ongoing research. For example, evolutionary increases in genome, cell, and nuclear size have been found to slow developmental processes (Jockusch 1997, Sessions and Larson 1987, Wyngaard et al. 2005), but the driving mechanisms are poorly understood. Development emerges from the progression and interaction of a wide range of processes taking place at the single-cell level, where increasing genome, cell, and nuclear size impact transcription dynamics, intra-cellular distances, surface area to volume ratios, and other fundamental characteristics (Cadart et al. 2023, Sessions and Wake 2021). We therefore consider how alterations in single-cell processes might translate to slowed developmental tempo as increasing genome, cell, and nuclear sizes change cell structure and functionality.

Additionally, relationships between cell volume and stochastic intracellular noise remain underexplored. Stochastic noise is present in all processes and interactions that shape developmental outcome, yet relative homogeneity in outcome tends to be maintained (Keskin et al. 2018, Munsky et al. 2012, Tsimring 2014). We are also interested in whether or not larger cells should be expected to endure higher levels of gene expression noise at the single-cell level, which would likely propagate out to broader biological scales. If larger cells do in fact deal with higher levels of noise relative to their smaller counterparts, this poses important questions about how the cells of genomic giants have evolved to cope with or damp high levels of intracellular stochasticity.

Vertebrates comprise a large portion of the overall range in genome and cell size across the tree of life, and despite variation in genome size, cell volume, and developmental tempo, many developmental processes remain conserved. Somitogenesis is one such process that is relatively well understood; it is the process through which bilateral pairs of somites, blocks of presomitic mesoderm (PSM) tissue, are patterned along the head-to-tail axis in vertebrate embryos. Somitogenesis is typically described as operating through a clock and wavefront mechanism in which cell-autonomous oscillations of a somitogenesis gene (i.e. the "segmentation clock") interact with Notch, Wnt, FGF, and retinoic acid pathways across the PSM tissue to coordinate proper timing of segmentation of groups of neighboring cells into bilateral pairs of somites (Aulehla and Pourquié 2010, Cooke and Zeeman 1976, Gibb et al. 2010, Klepstad and Marcon 2023). The segmentation clock operates via oscillatory gene expression at the single-cell level, while the wavefront takes place at the intercellular level across the PSM. Vertebrates exhibit species-specific segmentation clocks directly related to oscillatory expression of an autoregulated somitogenesis gene at the single cell level (Diaz-Cuadros et al. 2023, Lázaro et al. 2023, Matsuda et al. 2020). One period of oscillation, the time required for one cycle of expression, at the single cell level determines the time needed for a group of neighboring cells to segment off from the larger block of unsegmented PSM tissue, creating a bilateral pair of somites. As a conserved phenomenon that drives developmental tempo while operating at the single-cell level, the segmentation clock is an appropriate lens through which to examine single-cell processes as an underlying link between increasing genome and cell size and slowed development. While species-specific segmentation rates have been linked to biochemical differences at the intracellular level (Diaz-Cuadros et al. 2023, Lázaro et al. 2023, Matsuda et al. 2020, Rayon et al. 2020), the role of genome and cell size as potential mediators of species-specific gene expression kinetics and therefore developmental tempo have not been explicitly examined. This is what we explore in Chapter 1 using a deterministic model.

Then, in Chapter 2, we turn our focus away from species-specific rates of development, and we use a simple stochastic model of the gene regulatory network underlying the segmentation clock to probe more generally for a relationship between cell volume and stochastic noise in gene expression.

Chapter 1: Amphibian segmentation clocks suggest mechanisms of slowed development across increasing genome size and nuclear volume

Introduction to a deterministic segmentation clock model

In our first chapter, we use a pre-existing deterministic model of the somitogenesis segmentation clock to explore potential mechanisms that may be slowing developmental tempo across increasing genome size and nuclear volume, both of which go hand-in-hand with increasing cell volume.

The segmentation clock is well modeled by the following system of delayed differential equations, first proposed by Lewis [2003], that describe coupled oscillatory expression of mRNA and protein,

$$\frac{dp}{dt} = am(t - T_p) - bp(t) \tag{1}$$

$$\frac{dm}{dt} = f(p(t - T_m)) - cm(t) \tag{2}$$

where: p is protein expression (i.e. the number of protein molecules in a cell); m is mRNA expression (i.e. the number of mRNA molecules in a cell); a is a rate constant for protein synthesis (protein/mRNA/min); b is the degradation rate for protein (protein/minute) and is given by the following expression: $b = ln(2)/h_p$, where h_p is the half-life (minutes) of the protein molecule; c is the degradation rate for mRNA (mRNA/minute) given by the following expression: $c = ln(2)/h_m$, where h_m is the half-life of the mRNA molecule; and T_p and T_m are the delays associated with protein and mRNA production, respectively.

Table 1: Model parameters

Notation	Description
p_{crit}	Critical protein threshold, number of protein molecules needed to achieve 10^{-9} concentration in nucleus
T_p	Delay associated with protein production (min)
T_m	Delay associated with mRNA production (min), sum of:
T_{tx}	Transcription delay (min)
T_{in}	Intron splicing delay (min)
T_{exp}	mRNA export delay (min)
a	Rate of protein synthesis (protein/mRNA/min)
k	Rate of mRNA synthesis, no repression (mRNA/min/cell)
b	Rate of protein degradation (protein/min), equal to $ln(2)/h_m$
h_m	Protein half-life
c	Rate of mRNA degradation (mRNA/min), equal to $ln(2)/h_p$
h_p	mRNA half-life

Equation (1) models the change in protein expression over time, $\frac{dp}{dt}$, $am(t - T_p)$ gives the rate at which proteins emerge in the cytoplasm, and $bp(t)$ gives the rate at which they degrade. Similarly, equation (2) models the change in mRNA expression over time, $\frac{dm}{dt}$. $cm(t)$ describes the rate at which mRNA molecules degrade while $f(p)$, given below, is a Hill function that describes the rate at which mRNA molecules emerge from the nucleus into the cytoplasm,

$$f(p) = \frac{k}{1 + (\frac{p}{p_{crit}})^n} \tag{3}$$

where k is a rate constant for mRNA synthesis (mRNA/min/cell) in the absense of repression, p_{crit} is the number of protein molecules in the nucleus needed to yield an assumed critical concentration of $10^{-9}M$ within the nucleus (from Lewis [2003]) associated with transcriptional repression. Notation and definitions of the parameters described above are given in Table 1. Equation (3) also includes an additional parameter n, the so called "Hill coefficient". The rate of mRNA emergence is dependent upon and inversely related to protein quantity, indicating the

presence of an auto-repressive mechanism (i.e. a negative feedback loop). Like Lewis [2003], we assume that these repressive proteins act as dimers and let $n = 2$.

As touched on above, the parameters T_p and T_m account for delays associated with protein and mRNA production, respectively. Delay parameters reflect the reality that biological processes are often a non-instantaneous affair. Mathematically, they distinguish the DDE system above from an ordinary differential equation system and are necessary to generate the sustained oscillations that yield species-specific periods of gene expression (Lewis 2003), corresponding to species-specific rates of somite segmentation.

Here, we adapt Lewis' model to assess the impact of increasing genome and cell size on the segmentation clock periods associated with two amphibian species that exhibit a 10-fold difference in genome size: *Xenopus laevis* and *Ambystoma mexicanum*, the model frog and salamander, respectively. We break delay parameters down into specific transcription, post-transcription, and translation processes, and we consider the potential impact of increasing genome and/or cell size on each individual component. We also adjust critical protein threshold values to reflect species-specific nuclear volume estimates. Finally, we consider additional potential roles for mRNA and protein stability, which are not directly related to genome or cell size, in the mediation of developmental tempo. We simulate the Lewis model under all of these different scenarios to test whether we can reproduce the observed periodicity of the somitogenesis clock. We buttress our simulations with an analytical derivation of the minimal conditions for oscillations in the Lewis model. With our approach, we are able to establish direct links between increases in genome size and nuclear volume and their specific impacts on gene expression (i.e. increases in network delays and threshold value) that yield slower developmental tempo, and we are also able to assess potential indirect roles for gene product stability in the mediation of developmental tempo.

Methods

Taxon selection

We choose to adapt the Lewis model to *X. laevis* and *A. mexicanum*. Their recorded genome sizes are ∼3.1 Gb (*X. laevis*) and ∼32 Gb (*A. mexicanum*) (Hellsten et al. 2010, Smith et al. 2019), and, although volumes vary across type, *A. mexicanum* cell volumes are typically larger than their *X. laevis* counterparts. *A. mexicanum* nerve cells, for example, are ∼2-times larger than *X. laevis* nerve cells (Roth and Walkowiak 2015), and their red blood cells (RBCs) are ∼10-times than in *X. laevis* (Gregory 2023, using cell volume measurements for *Ambystmoma tigrinum* whose average reported genome size is also ∼32 Gb. Note also that amphibian RBCs are nucleated). Meanwhile, there is about a 3-fold difference in the rate of somite segmentation. In *X. laevis*, bilateral pairs of somites are segmented off from the PSM every 50 minutes (extrapolated by Curran et al. [2014] from Faber and Nieuwkoop [1994] and Hamilton [1969]); in *A. mexicanum*, somite segmentation occurs every ∼155 minutes (Armstrong and Graveson 1988).

Generating species-specific parameter values

Our first goal was to test if parameter changes directly related to increasing genome, cell, and nucleus size are sufficient to recapitulate the slowed rate of somite segmentation in *A. mexicanum* relative to *X. laevis*. To this end, we start by generating species-specific delay time and critical threshold parameters that capture genome size and nuclear volume, while holding all other parameters (rates of production and degradation) at constant values or ranges between species.

We derive protein and mRNA-production delays, T_p and T_m, by applying estimation methods from Lewis to somitogenesis gene candidates for *X. laevis* and *A. mexicanum*. Vertebrate clock genes are members of the Hairy and enhancer of Split (*Hes/Her*) family of basic helix-loop-helix

(bHLH) genes (Kageyama et al. 2007). *Hes/Her* gene family size varies across vertebrates, and the individual cycling members of the clock network also vary (Eckalbar et al. 2012, Krol et al. 2011). In zebrafish and mice, *hes7* has been shown to be the central component of the oscillator (Bessho et al. 2001, Hirata et al. 2004, Holley et al. 2002, Oates and Ho 2002), and its oscillatory pattern has also been shown in the PSM in the lizard *Anolis carolinensis* (Eckalbar et al. 2012). We therefore infer that *hes7* is the ancestral clock for vertebrates and that it retains clock function in *A. mexicanum*. In *Xenopus laevis*, in contrast, the *Hes/Her* gene family is greatly expanded to 37 copies, following both ancient allotetraploidization and tandem duplication (Kuretani et al. 2021, Watanabe et al. 2017). The *hes7* orthologs in *X. laevis* are not cyclically expressed in the PSM and, therefore, cannot act as the clock (Davis et al. 2001, Jen et al. 1999, Shinga et al. 2001). In contrast, in *Xenopus laevis*, three *Hes/Her* family genes are known to oscillate in the PSM: *hes5.3*, *hes5.5*, and *hes5.7* (Blewitt 2009, Li et al. 2003). Of these, the strongest candidate is *hes5.7L* based on experimental data showing that *de novo* protein synthesis is required to repress *hes5.7L* transcription during somite formation, and that *hes5.7L* RNA instability is part of the mechanism underlying its cyclic expression (Davis et al. 2001, Li et al. 2003). *Hes5.7* is absent from the closely related *X. tropicalis* and is thus inferred to be specific to *X. laevis* (Kuretani et al. 2021, Watanabe et al. 2017), suggesting a case of developmental system drift (Haag and True 2021). In *X. laevis*, *hes5.7L* has a primary sequence length of 1,604 nt; it is made up of 3 exons and 2 introns (lengths: 166 and 113 nt), and its coding sequence is 465 nt (Sayers et al. 2022). In *A. mexicanum*, *hes7* has a primary sequence length of 8272 nt; it is made up of 4 exons and 3 introns (lengths 3017, 1260, and 2030 nt), and its coding sequence is 783 nt (Smith et al. 2019 visualized on `http://genome.ucsc.edu/`). This information is used to derive species-specific parameters:

Critical protein threshold, p_{crit} We calculate the number of protein molecules needed to achieve a critical concentration of $10^{-9}M$ in the nucleus, based on species-specific nuclear vol-

umes. We assume a spherical nucleus ($V = \frac{4}{3}\pi r^3$) and estimate species-specific radii of 4 and 5.5 μm using Fiji analyses (Schindelin et al. 2012) of stained PSM nuclei in *X. laevis* (Hidalgo et al. 2009) and *A. mexicanum* (Banfi et al. 2012), respectively.

Protein production delay, T_p We assume that the delay associated with protein production is equal to translation delay. We estimate species-specific translation delays by applying a translation rate of 6 nucleotides per second (Lewis 2003) to the reported coding sequence lengths.

mRNA production delay, T_m We consider mRNA-production delay to be a cumulative sum of transcription T_{tx}, intron-splicing T_{in}, and mRNA export T_{exp} delays. We estimate species-specific T_{tx} values by applying a transcription rate of 20 nucleotides per second (Lewis 2003) to the reported primary sequence lengths. Hoyle and Ish-Horowicz [2013] find that in vivo intron-splicing delay constitutes a relatively constant proportion of ∼8.3 % of the overall segmentation clock period in mice, chick, and zebrafish embryos. We apply this proportion to the reported clock periods in *X. laevis* (∼50 min) and *A. mexicanum* (∼155 min) to get species-specific T_{in} values. We estimated species-specific mRNA export delays, T_{exp}, using simulations of particle diffusion within a sphere, described in detail below.

Simulation of mRNA export time parameter, T_{exp} Before they are released into the cytoplasm, newly transcribed mRNA molecules must journey from their chromatin address to the nuclear periphery where they locate an exit pore. Journeying to the periphery constitutes a relatively large part of this process, on the order of minutes, whereas locating and exporting through a pore once there is relatively rapid, typically on the order of fractions of seconds (Ben-Yishay and Shav-Tal 2019, Mor and Shav-Tal 2010, Mor et al. 2010). mRNA movement through the nucleoplasm takes place via passive diffusion. Both normal and obstructed (sub-) diffusion have been observed (Ben-Ari et al. 2010, Ishihama and Funatsu 2009, Mor and Shav-Tal 2010, Mor et al. 2010, Oeffinger and Zenklusen 2012, Shav-Tal et al. 2004). Normal and obstructed diffusion of RNA

can be described by simple and fractional Brownian Motion, respectively (Jeon et al. 2014, Lampo et al. 2017, Mor and Shav-Tal 2010, Oeffinger and Zenklusen 2012), with obstructed diffusion in the nucleus typically attributed to constrained pathways arising from chromatin organization (Ben-Ari et al. 2010, Ishihama and Funatsu 2009, Mor and Shav-Tal 2010, Oeffinger and Zenklusen 2012, Sheinberger and Shav-Tal 2013). Given that both normal and obstructed diffusion of transcripts in the nucleus have been observed, we run normal and fractional Brownian Motion simulations and compare results.

To generate species-specific estimates for nuclear export, we simulate the (3D) random walk of a diffusing particle within spheres of radii between 3 and 6 μm. This range of radii captures measurements for *X. laevis* and *A. mexicanum* PSM cell nuclei, 4 and 5.5 μm, respectively. We simulate mRNA transcript trajectories and record the number of steps needed for our simulated mRNA molecule to first cross the nuclear periphery from the nuclear center. 10,000 trajectories are simulated for each sphere in the specified range (radius of 3 to 6 with intervals of 0.5).

The transcript trajectory for normal diffusion is simulated by generating x, y, and z position vectors as cumulative sums of increments (i.e. step sizes) chosen from a normal random distribution. The transcript trajectory for obstructed diffusion is simulated similarly, with the x, y, and z position vectors generated directly by a fractional Brownian Motion function in MATLAB, `wfbm`, with a Hurst parameter of 0.25. For both simulations, the number of steps required by a particle trajectory to first exit the domain is retrieved and averaged. We assume that each step takes one second, and we scale the particle trajectory such that the estimated mean first exit times agree reasonably well with mRNA exit times observed for the somitogenesis gene in *Danio rerio*, or zebrafish. It has been reported that nuclear export of *her1*(*hes7*), the somitogenesis clock gene in zebrafish, takes 3.36 minutes (Hoyle and Ish-Horowicz 2013). We used Fiji (Schindelin et al. 2012) and images of stained nuclei across the PSM in zebrafish from Keskin et al. [2018] to

estimate a corresponding nuclear radius of 3μm. Then, we scaled our trajectory time scales in both simulations such that it takes, on average, about 202 seconds, i.e., 3.36 min (corresponding to 202 steps of the simulation) to first exit a sphere of radius 3μm.

We assume the nuclear center as the initial mRNA position based on known chromosomal territories associated with the somitogenesis gene in humans and mice coupled with patterns of synteny observed across vertebrates. That is, in both mice and humans, the somitogenesis gene, *hes7*, is found on gene-rich chromosomes 11 and 17 (which are homologs), respectively; both chromosomes localize in their respective nuclear centers (Boyle et al. 2001, Kile et al. 2003, Mayer et al. 2005, Zody et al. 2006). Broad patterns of synteny conservation and topologically associated domain conservation have been observed across vertebrates (Schloissnig et al. 2021, Smith et al. 2019), suggesting that this pattern extends beyond humans and mice.

All species-and diffusion specific parameter values (described above) are given in the first part of Table 2. In the second part of Table 2, we provide model parameters that are held at constant values across all models. Selection of these values are described below.

Parameters held constant For the first set of analyses, we hold the rates of mRNA production (in the absence of inhibition), protein production, and mRNA degradation constant, at $k = 33$ mRNA/min, $a = 4.5$ protein/mRNA/min, and $c = ln(2)/3$ mRNA/min corresponding to a half-life $h_m = 3$ minutes (Lewis 2003); and we consider a set range of protein stability/degradation $ln(2)/23 < b < ln(2)/3$ protein/min, corresponding to a half-life of $3 < h_p < 23$ minutes. This range is chosen based on typical reported and estimated protein stability of the somitogenesis gene across model vertebrates, namely zebrafish and mice; we test the model across a range of protein half-lives due to its important role in mediating the period of gene expression (Hirata et al. 2004, Lázaro et al. 2023, Lewis 2003, Matsuda et al. 2020, Takashima et al. 2011).

Table 2: Initial parameter values and what is captured by changes in species-species values

Parameter		Species-specific values		Genome size or nuclear volume captured?
		X. laevis	*A. mexicanum*	
p_{crit}		161	420	Increasing nuclear volume
T_p		1.29	2.18	Neither
T_{tx}		1.34	6.89	Increasing genome size
T_{in}		4.15	12.87	Neither
T_{exp}	*BM model*	6.39	11.97	Increasing nuclear volume
	fBM model	8.36	26.27	Increasing nuclear volume
T_m	*BM model*	11.88	31.73	Increasing genome size and nuclear volume
	fBM model	13.85	46.03	Increasing genome size and nuclear volume

Parameter	Values and ranges held constant across species and diffusion-type
a	4.5
k	33
c	$ln(2)/h_m$, where $h_m = 3$ minutes
b	$ln(2)/h_p$, where $3 < h_p < 23$ minutes

Extrapolation of periodicity to compare with known somite segmentation rates

For our first set of analyses, we plug the parameter values described above back into the Lewis model, and we use the DDE solver `ddesd` in Matlab to generate solutions across identical ranges of protein stability h_p and species- and diffusion-specific ranges of total delay time, $T_m + T_p$. We assess the period of gene expression that emerges for each set of solutions, and we compare it to the known somite segmentation rate of the corresponding species. In doing so, we aim to first verify that our parameter selection does in fact yield the correct period of oscillation for *X. laevis*, and to then determine if parameter changes directly driven by genome and cell size differences are sufficient to capture slowed developmental rate in *A. mexicanum* relative to *X. laevis*.

To assess periodicity, we create vectors to store the local extrema (i.e. local minimum and maximum values) and corresponding time stamps for each solution. Gene expression tends to spike in the first 4 to 5 cycles of oscillation before settling into a long-term pattern, so we remove the first 5 cycles of oscillation from our data to avoid skewing. The period of oscillatory

gene expression is calculated by taking the average difference between successive time stamps associated with local minima (using local maxima would yield the same results). We do this across a time span of 0 to 3,100 minutes. The time span, 3,100 minutes, is chosen based on the fact that (at least) 20 somites are observed in *A. mexicanum* embryos, each requiring ∼155 minutes to form (Armstrong and Graveson 1988). This time span also works well for *X. laevis*, with approximately 50 somites (Dali et al. 2002), each taking ∼50 minutes to form (Li et al. 2003). For both species, oscillations must remain robust throughout this time span. We define robust oscillations as having an amplitude (the height of an oscillation, or difference between local minimum and maximum) of no fewer than 10 molecules throughout the time span. If this requirement is not met, the oscillations are considered damped and we define the periodicity as `Inf` (infinite). We choose 10 molecules as a conservative finite cut-off based on observed average RNA transcript amplitudes of zebrafish segmentation clock genes *her1* and *her7* which are ∼41 and ∼49 molecules, respectively (Keskin et al. 2018).

The assessment of periodicity described above is done for both mRNA and protein counts, and the difference in periodicity is always within 0.2 minutes (see Appendix: Supplemental Material 1). In other words, results are relatively similar for both sets of oscillations, so we choose mRNA periodicity to represent overall system behavior.

Calculating average amplitude of expression

We also calculate the average amplitude of mRNA expression across the parameter combinations in each model. To do this, we use the extrema vectors described above to create a vector that stores the difference between local minima and maxima corresponding to every complete cycle of oscillation, excluding data from the first 5 oscillation cycles that were cut out. The resulting vector gives the amplitude associated with each complete oscillation cycle, and we take the average of

all vector entries to get an average amplitude associated with a particular parameter combination. For all combinations with periodicity defined as `Inf` (infinite), we set amplitude equal to 0, to keep results consistent with each other.

Sensitivity analysis

To assess the impact of individual changes on the period of gene expression, we increase and decrease each individual parameter (assessing total delay as a parameter as opposed to taking T_p and T_m individually) by 50% while holding all other parameters constant. We first extrapolate the resulting period and amplitude of gene expression for an original set of parameters (corresponding to an *A. mexicanum* Brownian Motion model with protein half-life arbitrarily set at $h_p = 15$ minutes) and then for each parameter change using the methods described above.

Testing whether scaling mRNA stability with export time yields the rate of somite segmentation in A. mexicanum

Our mRNA export simulations suggest that transcripts take much longer to leave larger nuclei. We therefore assess the impact of increasing mRNA stability with estimated nuclear export time on the *A. mexicanum* segmentation clock period. To this end, we re-considered both *A. mexicanum* models, normal and fractional Brownian Motion corresponding to normal and obstructed diffusion, under 3 different levels of mRNA stability: half-life equal to mRNA export time, half-life equal to 50% of mRNA export time, and half-life equal to 25% of mRNA export time. While holding all other *A. mexicanum* species- and diffusion-specific parameters at their initial values (Table 2), we re-generate solutions and assess periodicity for the Lewis model given $h_m = T_{exp}$, $h_m = \frac{1}{2}T_{exp}$, and $h_m = \frac{1}{4}T_{exp}$.

Generalizing mRNA export simulations

We generalize our mRNA export simulations by running the simulations described above across a range that encompasses what has been observed across the tree of life, that is nuclei of radius between ∼0.5 and 13 μm, based on the minimum and maximum nuclear volumes reported in the dataset used by Malerba and Marshall [2021] (`https://doi.org/10.5061/dryad.vq83bk3ss`) while assuming a spherical volume $V = \frac{4}{3}\pi r^3$. We also run simulations for which the initial position is selected from a uniform distribution as opposed to always being set at the origin (assumed to be at the nuclear center). The uniform distribution we draw our initial x, y, and z positions from encompasses a domain that is $\frac{3}{4}$ of each radius. In doing so, we allow for initial positions to be drawn from positions throughout our theoretical nuclei, excluding the periphery. We choose to exclude the nuclear periphery because this is where heterochromatin is spatially concentrated and reduced transcriptional activity has been observed (Bizhanova and Kaufman 2021).

Results

Period of gene expression is most sensitive to changes in delay and stability parameters

Conditions for the emergence of oscillations for simplified versions of the Lewis model, assuming a single degradation rate and a single delay, have been obtained analytically Verdugo and Rand [2008], but it was not clear whether the results would generalize. We therefore derived conditions for the emergence of oscillations for the full Lewis Model (see Appendix: Supplemental Material 2). We found that oscillating solutions require (i) the geometric mean of the degradation rates to be less than an upper bound,$\sqrt{K}$, and (ii) total delay $T_m + T_p$ to be equal to or greater than a critical value, T_{crit}. Both $\sqrt{K}$ and T_{crit} depend upon the kinetic constants of the model. In

Table 3: Sensitivity analysis

Original parameter set	**Period of mRNA expression** (minutes)		**Amplitude of mRNA expression** (molecules)	
$a = 4.5$, $k = 33$, $p_{crit} = 420$, $h_m = 3$, $h_p = 15$, Total delay = 33.91 (*A. mexicanum* BM model)	109.08		31.21	
Parameter changed	50% increase	50% decrease	50% increase	50% decrease
a	110.02	107.84	28.76	26.23
k	109.93	107.87	43.13	13.01
p_{crit}	108.41	110.48	30.58	26.19
h_m	114.55	102.88	41.84	11.77
h_p	`Inf`	98.41	0.99	112.07
Total delay ($T_m + T_p$)	152.79	`Inf`	78.34	0.05

particular, T_{crit} is positively correlated with h_p, the protein half-life. Thus sustained oscillations with longer lived proteins require longer delay times.

Since analytical solutions of the Lewis model for arbitrary parameters do not exist, we ran a sensitivity analysis simulating the model for different parameter values and calculated the period of oscillation. The results of our sensitivity analysis are given in Table 3. In agreement with results in Lewis [2003], we find that the period of gene expression (we look at mRNA specifically) is most sensitive to changes in total delay, $T_m + T_p$, and stability, h_m and h_p, parameters. Additionally, we find that when protein is too stable and total delay time is too low, robust oscillations do not emerge. In other words, total delay must be relatively large compared to protein half-life for oscillations to emerge, in agreement with our analytical results and previous mathematical and empirical results (Hirata et al. 2004, Lewis 2003, Takashima et al. 2011).

Parameter changes directly linked to increasing genome and cell size can mathematically recapitulate slowed developmental tempo

Using simulated export times, we generate species- and diffusion-specific general delay times associated with mRNA production, a sum of transcription, intron splicing, and nuclear export delays. All other parameters are held constant either across species (a, k, b, c) or at species-specific values across diffusion models (p_{crit}, T_p). All of these values are shown in Table 2.

The resulting periods of oscillation for each model are shown in Figure 1 (with non-oscillatory combinations/regions, assigned period `Inf`, shown in dark purple). We test across a range of parameter combinations: on the x-axis, we have values of total delay, $T_m + T_p$, that fall within ± 5 minutes of our species- and diffusion-specific values given in Table 2; on the y-axis, we have a range of protein stability corresponding to $3 < h_p < 23$ minutes. The observed rate of somite segmentation in *X. laevis* ($\sim$50 minutes) is captured by a subset of total delay and protein stability combinations under both normal and sub-diffusive (obstructed diffusion) conditions; this subset is outlined in the plot by a dashed-line in figures 1A and 1B. In confirming that we can achieve the correct period of oscillation for *X. laevis*, the results in figures 1A and 1B provide support for our methods of parameter estimation; these results also act as a plausible baseline against which the *A. mexicanum* models can be compared. Meanwhile, the observed rate of somite segmentation in *A. mexicanum* ($\sim$155 min) is only captured under sub-diffusive conditions. Genome and nucleus size-driven increases in delay time and concentration threshold are sufficient to fully recapitulate slowed development in *A. mexicanum* when nucleoplasmic movement of transcripts is assumed to be sub-diffusive (Figure 1D), but are insufficient when normal diffusion is assumed (Figure 1C). When normal diffusion is assumed, genome and cell size-driven parameter changes can slow the period of oscillatory gene expression down to $\sim$125 minutes, 30 minutes faster than the known

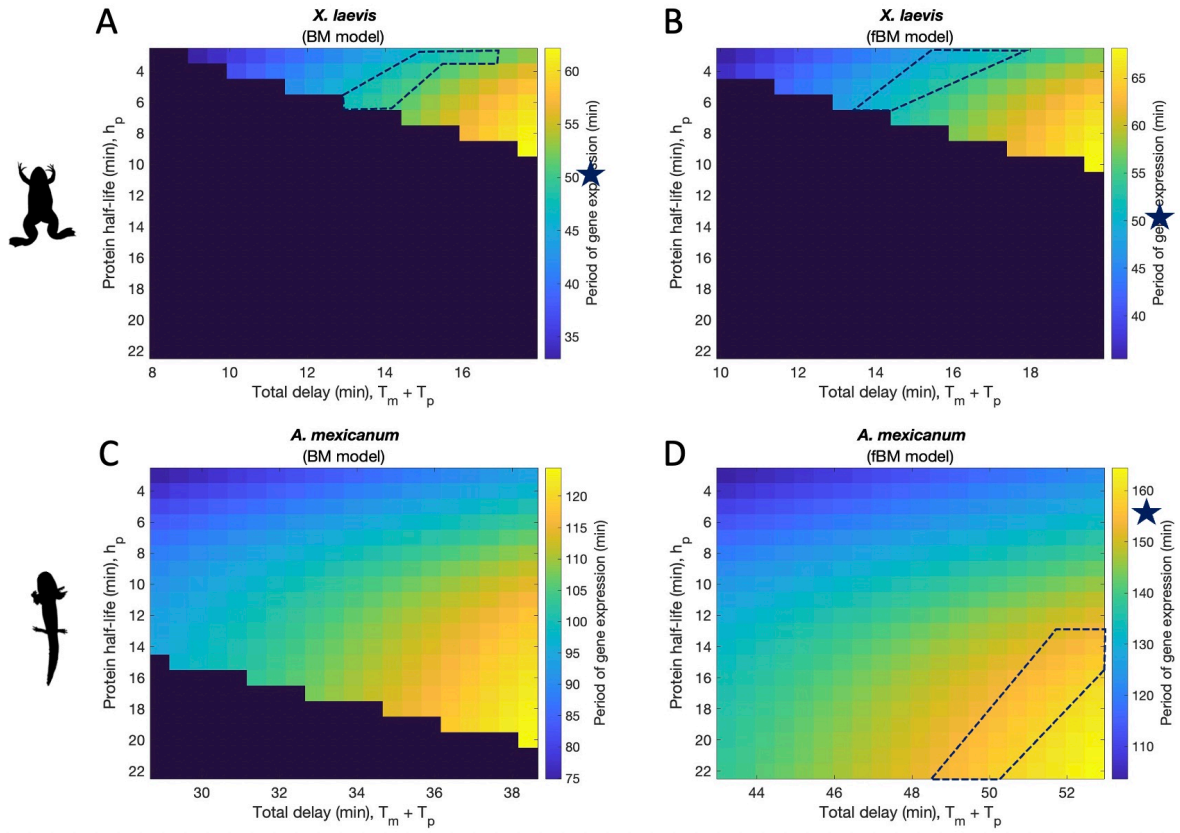

Figure 1: Resulting periods of gene expression given for each species- and diffusion-specific model. Each combination of protein half-life and total delay time corresponds to a period shown in the colorbar to the right of each model plot. Dark purple areas are non-oscillatory; stars show where the known species-specific rates of somite segmentation are found on the colorbar, and the corresponding regions are outlined in dashed lines. Results for: A *X. laevis* model when normal diffusion is assumed/Brownian Motion is modeled; B *X. laevis* model when obstructed diffusion is assumed/fractional Brownian Motion is modeled; C *A. mexicanum* model when normal diffusion is assumed/Brownian Motion is modeled; D *A. mexicanum* model when obstructed diffusion is assumed/fractional Brownian Motion is modeled.

rate of somite segmentation. The additional delay introduced by assuming sub-diffusion, about 14 minutes longer, is needed to produce a set of total delay and protein stability combinations that yield a ∼155 min period of oscillation.

mRNA export is 2- to 3-fold slower in A. mexicanum relative to X. laevis

mRNA export time roughly doubles in *A. mexicanum* relative to *X. laevis* when normal diffusion is assumed, and roughly triples when obstructed diffusion is assumed (see T_{exp} in Table 2). The

impact of obstructed diffusion on export time becomes more pronounced (i.e. the gap between export times for the two diffusion types becomes wider) as nuclear radius and therefore volume increase.

Scaling mRNA stability with nuclear export time yields biologically plausible recapitulation of slowed developmental tempo

Although the fBM model shown in Figure 1C yields the known rate of somite segmentation in *A. mexicanum* based solely on genome and cell size differences in parameter values, an additional increase in mRNA stability seems logically necessary given our simulated nuclear export times. The models shown in Figure 1 assume an mRNA degradation rate associated with a half-life of 3 minutes. However, we are working with simulated mean export times of ∼12 and ∼26 minutes in *A. mexicanum* PSM nuclei under normal and sub-diffusive conditions, respectively. Under these assumptions, a vast majority of mRNA molecules are expected to degrade long before ever leaving the nucleus.

We therefore test if the *A. mexicanum* segmentation clock can also be recovered when mRNA stability is increased relative to export time such that a greater proportion of transcripts are able to exit the nucleus before degrading. Although it is established that some fraction of RNA transcripts will degrade in the nucleus before exiting into the cytoplasm, the extent of degradation seems to differ across transcript types and remains relatively understudied (Smalec et al. 2022). Our parameter estimates for *X. laevis* yield the following patterns: under normal diffusion, mRNA half-life (3 minutes) is ∼47% of mRNA export delay (6.32 minutes), and under obstructed diffusion, mRNA half-life (3 minutes) is ∼36% of mRNA export delay (8.41 minutes). In terms of degradation, ∼76% of mRNA transcripts would be expected to degrade before leaving the nucleus (∼24% would leave the nucleus before degrading) under normal diffusion. Under ob-

structed diffusion, ∼82% of mRNA transcripts would be expected to degrade before leaving the nucleus (∼18% would leave the nucleus before degrading). Meanwhile, when we compare the typical estimate for mRNA half-life in *Danio rerio*, also 3 minutes, to the reported in vivo export time, 3.36 minutes (Hoyle and Ish-Horowicz 2013), we have that mRNA half-life is 83% of export time. This corresponds to an expectation that ∼58% of mRNA transcripts would degrade before leaving the nucleus (∼42% would leave before degrading).

To account for the wide range of empirical and theoretical estimates of mRNA degradation in the nucleus, we explore 3 potential scaling scenarios for mRNA stability and nuclear export time in *A. mexicanum*. We first set mRNA half-life equal to diffusion-specific export times, $h_m = T_{exp}$, for which 50% of mRNA transcripts degrade before leaving the nucleus (and 50% leave the nucleus before degrading); then we set mRNA half-life equal to 50% of the simulated export times, $h_m = \frac{1}{2}T_{exp}$, for which 75% of mRNA transcripts degrade before leaving the nucleus (and 25% leave before degrading); finally, we set mRNA half-life equal to 25% of the simulated export times, $h_m = \frac{1}{2}T_{exp}$, for which 87.5% of mRNA transcripts degrade before leaving the nucleus (and 12.5% leave before degrading).

In Figure 2, we show plots of *A. mexicanum* model results under both normal and fractional BM conditions, while decreasing the level of mRNA stability. When normal diffusion of transcripts through the nucleoplasm is assumed, figures 2A, 2C, and 2E, none of the mRNA stability scenarios produce a subset of protein half-life and total delay time parameter values that yield a period of 155 minutes. Under obstructed diffusion, figures 2B, 2D, and 2F, all mRNA stability scenarios produce a subset of protein half-life and total delay combinations that fully yield a period of $\sim$ 155 minutes, matching the known rate of somite segmentation in *A. mexicanum*.

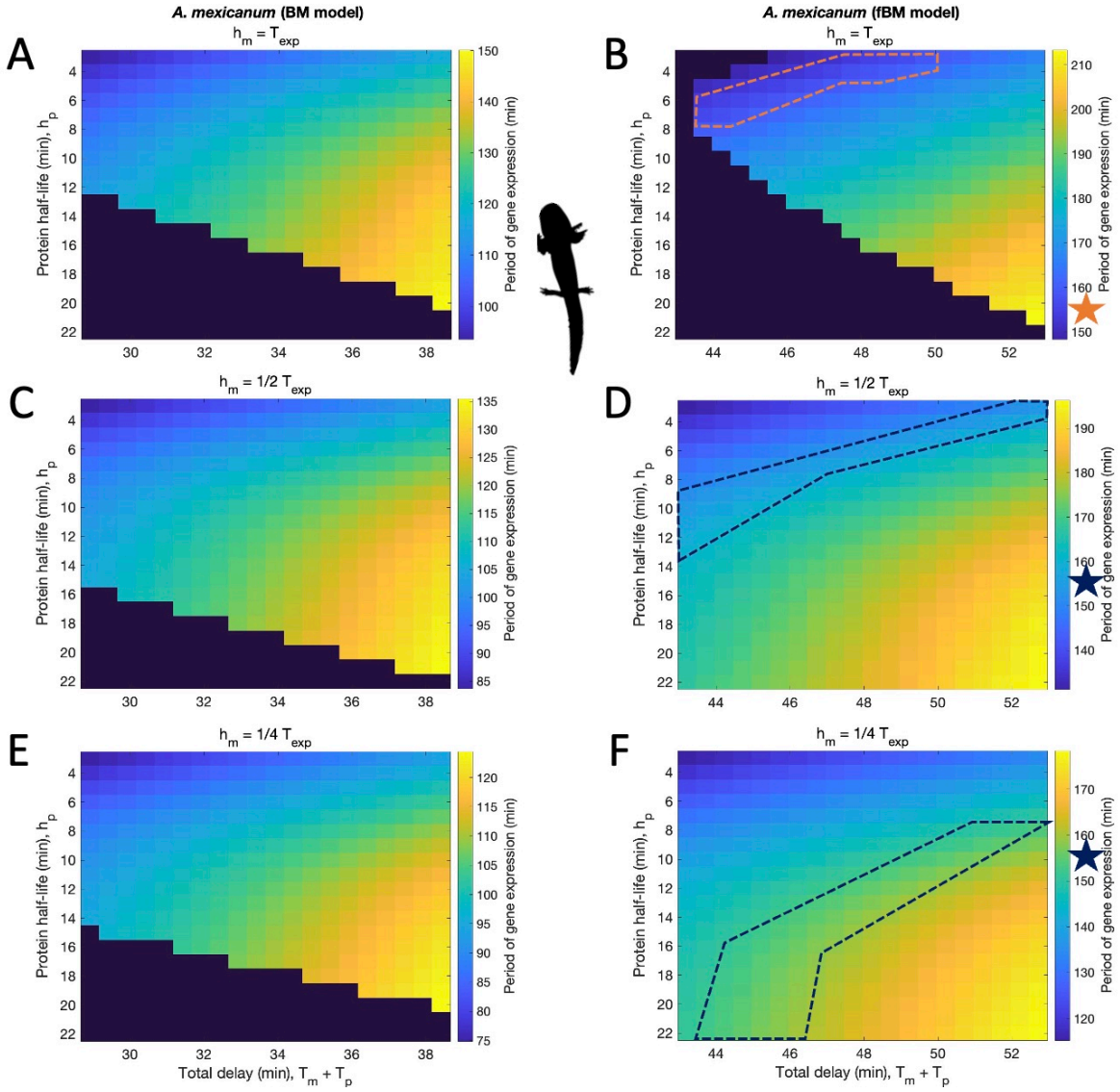

Figure 2: Resulting periods of gene expression for *A. mexicanum* models: A, C, E normal diffusion/Brownian Motion; B, D, F obstructed diffusion/fractional Brownian Motion.
mRNA half-life is held constant at: A, B diffusion-specific estimates for mRNA export delay; C, D half of estimated mRNA export delays; E, F a quarter of estimated mRNA export delays.
* Note: Color of star/outline is chosen for contrast and has no additional meaning.

Discussion

Previous research on developmental timing has pointed towards intron length and differences in biochemical characteristics, like degradation rates and network delays, as sources of species-specific tempo (Lázaro et al. 2023, Matsuda et al. 2020, Rayon et al. 2020, Swinburne and Silver

2008, Swinburne et al. 2008). In our study, we not only reconsider these points while contextualizing them within the scope of increasing genome and cell size, but we also make spatially explicit considerations that set our study apart from others.

In Table 2, we outline how species-specific parameters differ between *X. laevis* and *A. mexicanum*, and we note whether each change (based on our methods of estimation) captures increases in genome size or nuclear volume, or neither. Using these species- and diffusion-specific parameters, we first verified that our parameter estimates for *X. laevis* were able to recapitulate the ∼50 minute segmentation clock. This indicated that our methods of parameter estimation were reasonable, and we moved forward using the same methods for our *A. mexicanum* models, testing for parameter changes and combinations that recapitulate the ∼155 minute segmentation clock. The results of our models reveal the following potential mechanisms through which developmental tempo slows with increased genome and nuclear size.

Increasing intron length impacts developmental tempo through transcriptional delays

Increasing the sum of delays in the gene regulatory network has the most significant impact on the period of gene expression (see Table 3: Sensitivity Analysis). When we break total delay down into individual components, we see that while transcriptional delay, T_{tx}, is not the most significant contributor to total delay, $T_m + T_p$, it confirms an intuitive link between increasing genome size and the alteration of oscillatory gene expression kinetics. This is because transcriptional delay includes the time needed to transcribe the entire primary gene sequence, including intronic regions, which scale positively with genome size. Indeed, increasing intron size (in part due to transposable element insertions) is a major driver of genomic expansion in *A. mexicanum* (Nowoshilow et al. 2018), and while intronic regions only constitute ∼17% of the *hes5.7L* primary sequence in *X. laevis* (279 bp of intronic sequence), they make up ∼76% of the *hes7* primary se-

quence in *A. mexicanum* (6,307 bp of intronic sequence). This is a general pattern observed across orthologous genes in *A. mexicanum* and *X. laevis* (Nowoshilow et al. 2018). Interestingly, this pattern of intronic expansion seems to be constrained in developmental genes, where an ∼11-fold increase in average intron length is observed between *X. laevis* and *A. mexicanum* in contrast with an almost 20-fold increase seen in non-developmental genes (Nowoshilow et al. 2018). In Nowoshilow et al. [2018], the role of transcription in developmental timing is suggested as an explanation for this pattern. The time required to transcribe intronic regions of genes (i.e. "intron delay") has not only been identified as a potential mediator of developmental timing via its role in gene regulatory networks (Swinburne and Silver 2008), but has also been suggested to link genomic gigantism in particular with slowed development and regenerative abilities through its impact on gene expression kinetics (Sessions and Wake 2021).

However, according to our computational results, increased transcriptional delay alone cannot drive the vast differences in overall delay time between *X. laevis* and *A. mexicanum*. Our model therefore also implies that intron delay alone cannot drive slowed developmental tempo observed in the *A. mexicanum* segmentation clock relative to that of *X. laevis*. Instead, we see in Table 2 that the delay associated with increasing nuclear volume, T_{exp} is the most significant contributor to overall delay time across all species- and diffusion-specific models. Although the delay associated with intron-splicing, T_{in} is an important contributor to total delay, and the difference between species-specific values for T_{in} is in fact larger than that between species-specific values for T_{tx}, inter-species differences in splicing kinetics cannot be clearly linked with differences in intron or, by extension, genome size (Khodor et al. 2012).

mRNA export time increases with nuclear volume and has a pronounced impact on developmental tempo

Increasing nuclear volume between *X. laevis* and *A. mexicanum* is captured by the critical protein threshold value, p_{crit}, and mRNA export delay, T_{exp}, parameters. The increase in critical protein threshold accounts for the fact that in a larger nucleus, more protein molecules are needed to reach the same critical concentration of $10^{-9}M$ required for transcriptional repression to act "in earnest" (Lewis 2003). However, changes to p_{crit} have a very limited impact on the period of gene expression relative to other parameter changes. Meanwhile, the increase in mRNA export delay in *A. mexicanum* relative to *X. laevis* has a relatively pronounced impact on the period of gene expression.

Although the increase in the estimated radius of PSM nuclei in *A. mexicanum* relative to *X. laevis* may seem minimal, 5.5μm compared to 4μm, mRNA export delay still nearly doubles in *A. mexicanum* when normal diffusion is assumed and nearly triples when obstructed diffusion is assumed. The impact of increasing radius is greater when obstructed diffusion is assumed, and differences in estimated export times (between lengths and diffusion types) become more pronounced as radius increases. As a result, we would expect mRNA export delays to become more pronounced in gene regulatory networks across increasing nuclear volume, and we might also expect mRNA export to be the largest contributor to total delay in a gene regulatory network, especially in larger nuclei. This can be seen in Figure 3, where we plot simulated mean export times under both normal and obstructed diffusion for radii between 3 and 6 μm.

If we expand the range of radii past 6 μm, the impact of increasing radius and obstructed diffusion on export time becomes even more pronounced, suggesting that the nuclear center in species with the largest known genome sizes may have become "uninhabitable" for genes with

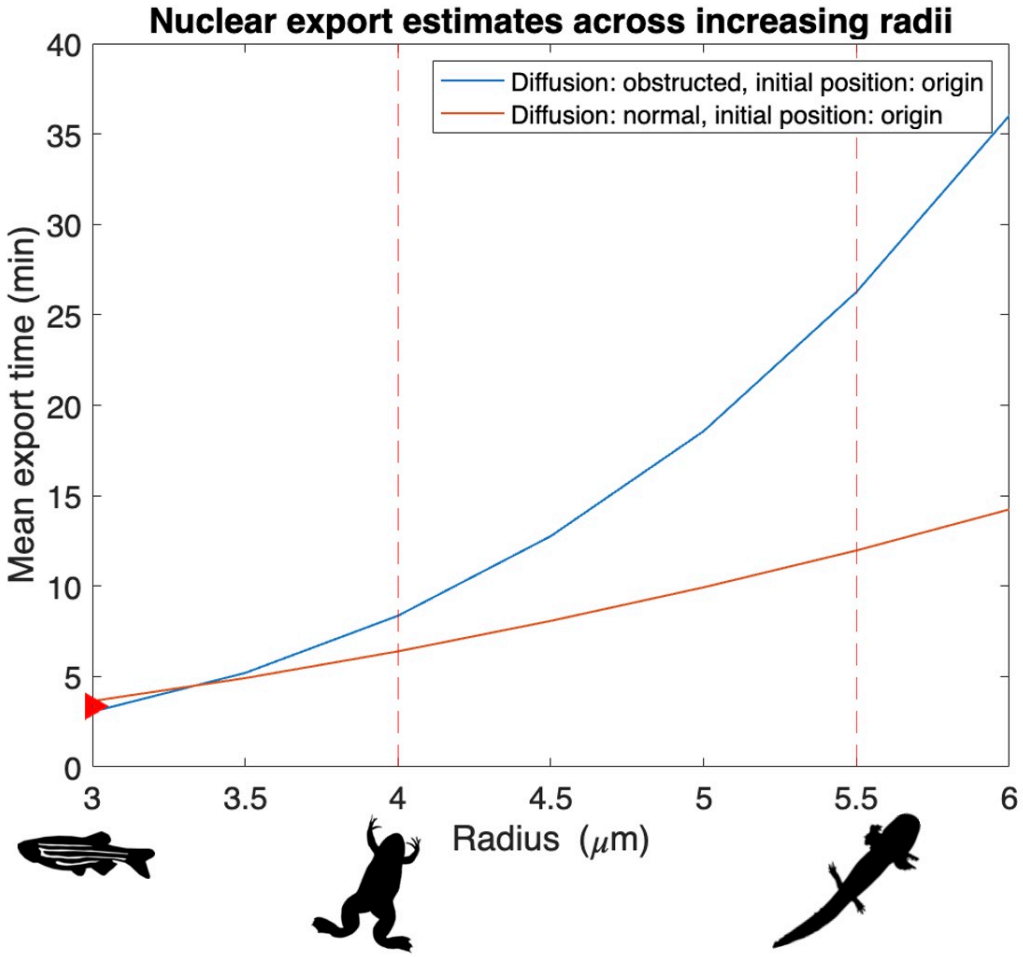

Figure 3: Nuclear export estimates under normal and obstructed diffusion across a range of radii that captures estimates for *X. laevis* and *A. mexicanum* PSM nuclei (shown by the red dashed lines). An initial position at the nuclear center is assumed. Trajectories are scaled such that a radius of 3μm corresponds to a mean export time of $\sim$3.36 minutes (shown by the red arrow) to match the reported export time of *her1 (hes7)* in zebrafish (Hoyle and Ish-Horowicz 2013). Note: There is a limit to how closely 3μm can be scaled with $\sim$3.36 minutes. Obstructed diffusion mean export times are slightly faster than normal diffusion for some radii $r < 3.5\mu$m. This is not biologically meaningful, and mean export times are quick to converge back to expectations.

dynamic expression patterns. If we introduce transcripts whose positions are drawn from a uniform distribution within the nucleus, as opposed to always starting from the origin/nuclear center, the impacts of increasing radius and obstructed diffusion on mean export time are reduced. The distribution of export times becomes more skewed towards lower values (because most transcript trajectories are starting closer to the periphery) (see Appendix: Supplemental

Material 3). In very large nuclei, we hypothesize that genes – especially those with dynamic expression patterns – may be constrained to occupy these locations away from the nuclear center, suggesting an overall effect of genome expansion on the organization of chromosomal territories.

Regardless of model type, assuming an intial position at the nuclear center, our simulations position mRNA export as the largest contributor to total delay time. This is consistent with observations that find mRNA export delays to be longer than both transcriptional and intron splicing delays in mice, chick, and zebrafish segmentation clocks (Hoyle and Ish-Horowicz 2013). As the largest contributor to total delay time, mRNA export is also the largest contributor to overall differences in *X. laevis* and *A. mexicanum* segmentation clocks, implying that spatially induced delays in gene regulatory networks play a significant role in the slowing of cellular and developmental processes. This suggests an important albeit underexplored spatial component in the mediation of developmental tempo across increasing cell and nuclear volumes.

Simulations position increasing chromatin density as a driver of slowed development

Across the tree of life, decreasing nucleus to cell volume ratios are observed as cells increase in volume (Malerba and Marshall 2021). In other words, larger cells have relatively smaller nuclei, though absolute nuclear volumes still increase with cell volume. Given the well-established positive correlation between genome size and cell volume, we might also state that nuclei become relatively smaller with increasing genome size. As a result, we would see an increasing amount of genetic material packed within relatively smaller nuclei, implying greater chromatin packing density as genome size and absolute nuclear volumes increase. Recall that chromatin structure density is a commonly cited source of obstructed diffusion observed in the nucleus (Ben-Ari et al. 2010, Ishihama and Funatsu 2009, Mor and Shav-Tal 2010, Oeffinger and Zenklusen 2012, Sheinberger and Shav-Tal 2013). Therefore, we propose that obstructed diffusion of transcripts in

organisms with large genomes and relatively smaller (but absolutely larger) nuclei is a reasonable expectation.

With these empirical possibilities in mind, we turn towards our theoretical results. In figures 1A and 1B, we see that both diffusion models are able to fully recapitulate the *X. laevis* segmentation clock, while obstructed diffusion (modeled by fractional Brownian Motion, shown in Figure 1D) must be assumed to fully recapitulate the slowed *A. mexicanum* segmentation clock. Similarly, in Figure 2, we have that obstructed diffusion still must be assumed, regardless of mRNA stability, in order to recapitulate the slowed *A. mexicanum* segmentation clock. Meanwhile, under normal Brownian Motion, all model scenarios fall short of fully recapitulating the *A. mexicanum* segmentation clock. Further increases in protein and/or mRNA stability do not fully resolve this issue (see Appendix: Supplemental Material 4), suggesting that normal diffusion does not introduce sufficient delays into the system. These results indicate that obstructed diffusion is sufficient but not necessary to recapitulate the segmentation clock when genome size is relatively small (*X. laevis*) yet becomes both sufficient and necessary when genome size is relatively large (*A. mexicanum*), supporting the hypothesis that chromatin packing density increases with genome size. Somitogenesis transcripts must exhibit obstructed diffusion in the nucleoplasm of *A. mexicanum* in order to slow the segmentation clock to an appropriate pace, thus positioning increasing chromatin density as a potential driver of slowed development.

Gene product stability also acts to mediate developmental tempo

Focusing on figures 2B, 2D, and 2F (fBM model results), we can see that as mRNA stability changes, so too does the range of protein half-life and total delay time combinations yielding a gene expression period of 155 minutes. This is in agreement with previous studies that point to gene product stability as a mediator of species-specific segmentation clock periods (Hirata et al.

2004, Lázaro et al. 2023, Matsuda et al. 2020, Rayon et al. 2020). However, some mRNA stability scenarios yield more reasonable results than others. In figures 2B and 2D, with mRNA half-lives of 26.27 and 13.14 minutes, the 155 minute period of oscillation is captured for protein half-lives that range from 3 to 8 and 3 to 13 minutes, respectively. As a result, we would have to assume that protein is as stable or less stable than mRNA, which is inconsistent with reported patterns for Hes7 in mice and humans (Matsuda et al. 2020). In Figure 2F, for which mRNA half-life is set to 6.54 minutes, we get a larger subset of parameter combinations that yield a period of 155 minutes. Across this subset, protein half-lives range from being only slightly more stable than mRNA at $\sim$8 minutes to almost 4-times as stable as mRNA at $\sim$23 minutes. The resulting relationship between mRNA and protein stability, that protein molecules exhibit increased stability relative to their transcripts, is more consistent with the patterns reported in Matsuda et al. [2020]. However, setting $h_m = \frac{1}{4}T_{exp}$ implies that a vast majority of mRNA transcripts, 87.5%, will degrade before exiting the nucleus. We might expect amplitude (i.e. number of transcripts in the cell) to be extremely low as a result, but this is not the case. In fact, mRNA degradation is related in complicated ways to not only the amplitude of mRNA expression, but to the amplitude of protein expression as well, resulting in counter-intuitive patterns of expression across different mRNA stability scenarios (see Appendix: Supplemental Material 5).

Considering again the period of gene expression, an interesting tradeoff exists between the possible mRNA stability scenarios described above: in order to mathematically recapitulate the slowed *A. mexicanum* segmentation clock, we must either assume that a vast majority of mRNA transcripts degrade within the nucleus, or that the protein molecules in this system are less stable relative to their transcripts. The assumption of obstructed diffusion necessitates some kind of deviation – either high rates of nucleoplasmic degradation, or low stability of protein molecules relative to their transcripts – from conventional gene expression kinetics as genome and nuclear

size increase between *X. laevis* and *A. mexicanum*. Based on reported patterns for Hes7 in mice and humans (Matsuda et al. 2020), we propose that widespread nucleoplasmic degradation of the somitogenesis transcript is more plausible than transcripts that are more stable than the corresponding protein molecules.

What remains the same across genome size and nuclear volume is that, as shown in previous implementations of the Lewis model (Hirata et al. 2004, Matsuda et al. 2020), the relative stabilities of different gene products interact to shape clock tempo. This is demonstrated by changing ranges of protein half-life corresponding to a period of 155 minutes as mRNA half-life decreases. However, gene product stability alone cannot capture slowed development across increasing genome size and nuclear volume. Simply increasing protein stability and/or mRNA stability for the Brownian Motion (normal diffusion) models, shown in figures 2A, 2C, and 2E, cannot fully recapitulate the slowed *A. mexicanum* clock (see Appendix: Supplemental Material 4 for an extended discussion). The additional delays introduced by obstructed nucleoplasmic diffusion are necessary to sufficiently slow the period of gene expression. Although gene product stability acts to mediate developmental tempo, our model results suggest that across increasing genome size and nuclear volume, gene product stability does not act alone to slow developmental tempo.

Conclusion

Our results suggest that spatial differences must be accounted for to recapitulate slowed development as a result of evolutionary increases in genome and cell size. In agreement with previous studies, our results also suggest an important role for gene product stability when it comes to mediating species-specific rates of development. Taken together, we show that the physical and spatial delays predicted by increased intron length and nuclear size, coupled with alterations to gene product stability that ensure the products persist for long enough to fulfill their molecu-

lar function, mathematically recapitulate the slow developmental tempo found in species with large genomes. However, we also find it necessary to assume obstructed diffusion in order to mathematically recapitulate the slowed *A. mexicanum* segmentation clock, indicating that we may expect to observe obstructed diffusion in organisms with absolutely larger genomes and relatively smaller nuclei.

Chapter 2: Stochastic gene expression model suggests increasing "search process noise" induced by large cell volumes

Introduction: Stochastic gene expression noise and cell volume

There are a host of evolutionary implications associated with genomic expansion. These include an increase in cell volume and a slowdown in cellular and developmental processes. In the previous section, we explored connections between increasing genome and cell size and the resulting slowdown in developmental processes. An interspecies comparison between *Xenopus laevis* and *Ambystoma mexicanum* suggested that increasing delays in both transcriptional time and nuclear export time, due to increasing genome size and nuclear volume, respectively, are key contributors to the slowing of the somitogenesis clock of the model salamander relative to the model frog. Although species-specific estimates for nuclear export delay are sourced from a random walk model, the system of delayed differential equations used to model the segmentation clock constitutes a deterministic model. Thus, we have so far ignored the impacts of stochastic noise on the progression of the segmentation clock. However, increases in cell volume pose interesting questions about a potential relationship between space and stochastic noise.

Let us imagine the biophysical processes operating behind the delayed differential equation model used in Chapter 1: mRNA is transcribed somewhere near the nuclear center, diffuses out to the nuclear pore, and then into the cytoplasm. There, mRNA molecules are recognized by ribosomes and the translational machinery assembled around them, after which translation ensues and the resulting proteins find their way back to the nucleus and associated promoter site to

which they bind, thus facilitating transcriptional repression. Intuitively, we expect this process to be noisier in larger cells where intracellular distances are longer, and where diffusing molecules have more space that can be explored via a random walk before reaching their target. Surprisingly, research exploring the effect of cell volume on stochastic noise is scarce. The research that does exist is limited to exploring stochasticity in cells that are growing in preparation for cell division (Gomez et al. 2014, Gonze 2013, Paijmans et al. 2017), but there is, to our knowledge, no comparison of stochastic noise between cells that exist at fundamentally different volumes, for example between a model organism cell and the cell of a genomic giant. To generate a working hypothesis on the relationship between stochastic noise and cell volume, we begin with exploring some of what is known about intracellular stochastic noise associated with gene expression.

A stochastic process is one that evolves in a random manner. Chemical reactions, for example, are intrinsically stochastic, as are many biological processes. Stochasticity in biological processes create "biological noise," and a wide range of context-dependent definitions of biological noise exist. Some notable examples include noise as variability in molecular copy number, cell decision making and/or cell fate, and cell differentiation between identical cells in a population (Mitchell and Hoffmann 2018, Puzović et al. 2023, Tsimring 2014). Developmental noise is often defined as the phenotypic variation between genetically identical individuals in population who are exposed to nearly identical environmental conditions (Yampolsky and Scheiner 1994). Also at the population level, we see demographic noise as a result of stochastic fluctuations in birth and death rates (Tsimring 2014). Tsimring [2014] provides a more comprehensive review of different forms of biological noise spanning across biological scales. From the list above, we notice that random variability and fluctuations are fundamental characteristics of biological noise.

Stochastic noise can be divided into two categories: intrinsic and extrinsic. Intrinsic noise is typically understood as arising from randomness that exists within a system, while extrinsic

noise is understood to be arising from fluctuations in environments and processes external to a system (Lei et al. 2015). In this chapter, we are interested primarily in intrinsic expression noise. Intrinsic expression noise is largely a product of the inherent randomness of molecular diffusion and reaction processes, and many point to expression noise as a fundamental source from which noise at larger biological scales originate (Munsky et al. 2012, Tsimring 2014). Indeed, we can imagine how even a small fluctuation in the timing or magnitude of gene expression at the intracellular level might have a ripple effect first across neighboring cells, then tissues and organs, and may even propagate out to the organismal level. It is this fundamental source of biological noise that we are most interested in, specifically at the single-cell level.

Stochastic noise in gene expression, or expression noise, leads to fluctuations in molecular copy numbers at the single cell level (Mitchell and Hoffmann 2018, Munsky et al. 2012, Pancaldi 2014, Puzović et al. 2023, Raser and O'Shea 2005). This implies that period and amplitude of a genetic oscillator, like the one driving the segmentation clock, could also exhibit high levels of variability (Tsimring 2014). Again, many studies of stochastic noise have focused on stochasticity in well-mixed systems. When we bring space explicitly into the system, we may also consider variability in the time needed for "search processes" in which molecules explore intracellular searching for a specific target site or region. For example, we may consider the variability in nuclear export delay, a component of gene regulatory networks that we assumed as deterministic in the first chapter, but a process for which variability might increase with respect to volume.

It is known that gene expression in general is "noisier," and deviates more from deterministic descriptions, when the underlying molecular copy numbers, specifically mRNA and protein, are low (Gomez et al. 2014, Hausser et al. 2019, Munsky et al. 2012, Puzović et al. 2023, Tsimring 2014), so we consider how proteomic concentration may influence expression noise in cells of different volumes. Similar to research gene expression noise, much of the existing research on

cell volume and proteomic concentration is limited to how concentrations change or are maintained throughout the cell cycle, or phases of growth and division in a single cell (Lanz et al. 2021, Lin and Amir 2018). Still, there are useful insights to be gleaned from understanding how molecular concentrations change or are maintained within a growing cell. First, it has been well documented that overall protein concentrations tend to be maintained by cells throughout the cell cycle (Lanz et al. 2021). This phenomenon is typically referred to as concentration homeostasis, and it necessitates that protein copy numbers increase proportionally with cell volume. Concentration homeostasis in growing cells is supported by increasing total RNA and protein synthesis, as well as increasing mRNA and protein stability (Lin and Amir 2018, Xie et al. 2022). Although data are scanty, there is evidence to suggest that proteomic concentrations are relatively similar across eukaryotes (Brown 1991). Data emerging from cells that have been experimentally manipulated to exceed typical ranges of volume reveals that cells lose functionality when proteomic concentrations are not maintained, and that cytoplasmic dilution contributes to senescence (Neurohr et al. 2019). However, this is not what we expect to see in large cells that exist in nature. We therefore hypothesize that even as cells evolve to be incredibly large, they maintain protein concentrations relatively similar to smaller eukaryotic cells in order to maintain basic cell functionality, but, without direct comparisons, we cannot definitively state that concentration homeostasis is maintained for all gene products.

Thus, we are left with two ways of conceptualizing how a relationship between cell volume and stochastic expression noise may emerge. First, we can imagine that increasing cell volume results in increasing levels of expression noise due to larger intracellular distances between specific target sites in the cell (i.e. between the promoter and nuclear membrane, and between the site of translation/protein emergence and the promoter). Second, we can also consider that if large cells are unable to scale proteomic concentration, gene expression noise may increase due

to a lower number of molecules per unit of volume (i.e. lower concentration), an under-explored but potential consequence of cytoplasmic dilution that may also contribute loss of cell functionality. In the following sections, we simulate both cases by comparing measures of expression noise related to oscillation period and search processes in small and large model cells. We test two large model cells, one for which protein concentration becomes dilute with increasing volume and another for which rates of mRNA synthesis and gene product (mRNA and protein) are increased to promote concentration homeostasis. For these simulations, we adapt a pre-existing spatial stochastic model of a hes gene regulatory network (Sturrock et al. 2013), thereby continuing to use the segmentation clock as our system of interest. In this chapter, we are less interested in mathematically recapitulating known segmentation clock periods, and we are more interested in comparing noise levels between cells for which specific model parameters have been altered with respect to volume and biosynthesis rates.

In the sections below, we focus first on whether or not our stochastic simulations produce robust oscillations in protein expression, and we consider the noisiness in possible patterns of oscillatory protein expression related primarily to the period of oscillation. We then turn to variability in the time needed for repression to first operate on the network (via protein-promoter site binding) as an additional measure of expression noise. We refer to the time needed for repression to first operate on the network as the "first repression event." The first repression event not only accounts for nuclear export delay but also for delays associated with the time between a transcript's emergence into the cytoplasm and translation, as well as the time needed for a protein molecule to return to and bind to the promoter site in the nucleus to facilitate autorepression. By measuring variability in the timing of the "first repression event," we are also able to measure noise associated with the cumulative "search process," and how what we will refer to as "search process noise" is impacted by increasing cell and nuclear volume.

Methods

Simulating cell autonomous stochastic gene expression

To test for an impact of cell volume and proteomic concentration on intracellular expression noise, we simulate stochastic gene expression in a model cell. We run spatial stochastic simulations in Smoldyn, a computer program that simulates zeroth, first, and second order reactions and Brownian Motion of molecules (Andrews et al. 2010). We set up a simple model cell in a 3 dimensional domain. Our model cell is spherical, with reflective boundaries – meaning that the molecules in the cell cannot diffuse out of the cell – and containing a spherical nucleus that is centered in the cell. The model nucleus is set to have transmissive boundaries – meaning that molecules diffuse freely in and out of the nucleus, since in this simplest model we are not interested in the kinetics of molecular diffusion and/or transport through the nuclear membrane. We also define a cytoplasm compartment as the volume contained within the cell but not within the nucleus. Doing so allows us to specify that the reaction modeling translation only occurs within the cytoplasm. Meanwhile degradation reactions may take place throughout the entire volume of the cell, and reactions related to transcription and repression take place at the nuclear center. We define 4 "species" within the cell: a free promoter site, an occupied promoter site, mRNA molecules, and protein molecules, and we simulate gene expression reactions in small and large model cells. In Table 4, we show the molecular species parameters that are held constant across all model cells (the diffusion constants and initial values associated with each species).

One simulation runs for $3,100$ time steps where we consider each time step to be 1 minute; this is in line with the run time in our Chapter 1 models, which was chosen based on the number of somites formed in *A. mexicanum* and *X. laevis*, and the amount of time needed to form each somite. We run 100 simulations for each model cell.

Table 4: Molecular species and size parameters

Molecular species	Diffusion coefficient	Initial value
mRNA	$4 * 10^{-1} \mu m^2/min$	0
Protein	$4 * 10^{-1} \mu m^2/min$	0
Free promoter (pF)	0	1
Occupied promoter (pO)	0	0

Parameter selection and comparison between different model cells

In the following section, we justify our model parameter selections, and we explain how some parameter values are used to define different versions of the small and large model cells we simulate. Model parameter values are given in Table 5, and parameters used to differentiate between model cells are in bold.

Model parameter selections

Diffusion coefficients, The mRNA diffusion coefficient shown in Table 4 is derived from normal Brownian Motion simulations carried out in the previous chapter (we use data from normal Brownian Motion as opposed to fractional Brownian Motion to match the simulation dynamics of our Smoldyn model). We take the mean nuclear export time associated with nuclear radii between 1 and 13 µm, increasing by increments of 0.5 µm, and we use the formula $D = \frac{MSD}{6t}$, where the mean squared displacement is the nuclear radius squared, $MSD = r^2$, and time, t, is the associated mean export time, to find an average diffusion coefficient, D, for each nuclear radius. We then average once more over all 26 diffusion coefficients to get $D \sim 0.42\ \mu m^2/min$ (standard deviation $\sim 0.02\ \mu m^2/min$). We therefore set the mRNA diffusion coefficient equal to $4 * 10^-1\ \mu m^2/min$ in our Smoldyn model. For simplicity, we also let the protein diffusion coefficient equal $4 * 10^-1\ \mu m^2/min$. Holding the mRNA and protein diffusion coefficients at the same value is in line with Sturrock et al., although we use a different diffusion coefficient. Finally, we set the diffusion coefficients associated with the free and occupied promoter molecules to 0 so

Table 5: Size and reaction parameters differentiate model cells

Size parameters	**Small model cells**	**Large model cells**
Cell radius (µm)	5	$5 * \sqrt[3]{10}$
Nuclear radius (µm)	3	$3 * \sqrt[3]{10}$

Reaction	**Small cell 1**	**Small cell 2**	**Large cell 1**	**Large cell 2**
$pF \xrightarrow{Tx} pF + mRNA$ Basal transcription (min^{-1})	$Tx = 3.0$	$Tx = 3.0$	$Tx = 3.0$	$\mathbf{Tx = 5.0}$
$pF + protein \xrightarrow{k_1} pO$ Protein-promoter association ($\mu m^3/molecules/min$)	$\mathbf{k_1 = 10/NA}$	$k_1 = 40/NA$	$k_1 = 40/NA$	$k_1 = 40/NA$
$pO \xrightarrow{Tx_{rep}} pO + mRNA$ Repressed transcription (min^{-1})	$Tx_{rep} = 0.3$	$Tx_{rep} = 0.3$	$Tx_{rep} = 0.3$	$\mathbf{Tx_{rep} = 0.5}$
$pO \xrightarrow{k_{-1}} pF + protein$ Protein-promoter dissociation (min^{-1})	$k_{-1} = 0.1$	$k_{-1} = 0.1$	$k_{-1} = 0.1$	$k_{-1} = 0.1$
$mRNA \xrightarrow{Tl} mRNA + protein$ Translation (min^{-1})	$Tl = 1.0$	$Tl = 1.0$	$Tl = 1.0$	$Tl = 1.0$
$mRNA \xrightarrow{\alpha_m} 0$ mRNA degradation (min^{-1})	$\alpha_m = 0.069$	$\alpha_m = 0.069$	$\alpha_m = 0.069$	$\boldsymbol{\alpha_m}$**= 0.02**
$protein \xrightarrow{\alpha_p} 0$ protein degradation (min^{-1})	$\alpha_p = 0.033$	$\alpha_p = 0.033$	$\alpha_p = 0.033$	$\boldsymbol{\alpha_p}$**= 0.015**

they maintain the same position throughout simulations. Diffusion coefficients are held constant across all models.

Initial values and placement, We set the initial values of mRNA, protein, free promoter and occupied promoter (shown in Table 4) to 0, 0, 1, and 0, respectively. The free promoter is placed at the center of the nucleus, and the promoter site, both in its free and occupied state, maintains this position throughout the simulations. Initial values and placement are also held constant across all models.

Size parameters, Model cell size parameters are given in Table 5. Small model cell and nuclei

are defined by radii of 5 µm and 3 µm, respectively. These numbers match size parameters chosen by Sturrock et al. [2013] but also fall within a range typical cell and nuclear radii seen in model mammalian cells (Li and Xie 2011). Large model cell and nuclear radii are set to achieve a 10-fold increase in volume (cell radius = $5 * \sqrt[3]{10}$); nuclear radius = $3 * \sqrt[3]{10}$). We choose to increase cell volume 10-fold to reflect a conservative observed difference in cell volume between "typical" vertebrate and genomic giants, like salamanders and lungfish (Gregory 2023).

Reaction parameters

We simulate 4 different model cells to compare measures of intracellular noise across increasing volume under various conditions. In Table 5, we provide the reaction system defined in Smoldyn, and the corresponding reaction parameters, some of which are used to differentiate between the 4 model cells. The reaction system we simulate is analogous to one presented in Sturrock et al. [2013] describing Hes1 auto-regulation in mice. The reaction system consists of transcription, translation, and mRNA and protein degradation reactions, but also includes a protein-promoter association reaction to describe transcriptional repression as an auto-regulatory mechanism. This is the same basis for the model in Chapter 1, but we now apply stochasticity to the system. All of the reactions and corresponding parameters described below are shown in Table 5.

Forward rate of protein-promoter binding, k_1 Although the protein-promoter association reaction is not listed first in Table 5, we start with justifications for selecting k_1 because model limitations compel us to model 2 versions of the small cell with different k_1 values. As a result, we are left with the 4 model cells listed in Table 5: 2 small and 2 large (one modeling dilution and the other for which biosynthesis rates are adjusted to promote concentration homeostasis). The 2 small cells differ only in the rate of protein-promoter association. Below, we explain the decision to model 2 small cells and justify their different rates of protein-promoter association.

We start with a theoretical protein-promoter binding rate of $k_1 = 10^9 M^{-1} min^{-1}$ (Tafvizi et al. 2011) that agrees well with empirical measurements, $k_1 = 7 * 10^9 M^{-1} sec^{-1}$, in a *lac* repressor-operator system (Riggs et al. 1970). We start with the theoretical measurement and convert to units of μm^3 and molecules, resulting in the following forward rate: $k_1 = 10/NA\ \mu m^3\ molecules^{-1} min^{-1}$ where NA is Avogadro's Number ($NA = 6.022 * 10^{23}$). To simulate biomolecular reactions, such as protein-promoter binding, Smoldyn computes a "binding radius" from the defined reactant diffusion coefficients, reaction rate constant, and defined time step, and the reaction is performed when two reactants diffuse closer together than the radius (Andrews et al. 2010). When we let $k_1 = 10/NA\ \mu m^3 molecules^{-1} min^{-1}$, the resulting radius is $0.75 \mu m$. This radius allows for regular occupation of the promoter site, about 40 times throughout one 3,100 minutes run, in the small model cell, but it is too small to facilitate regular promoter occupation in the larger cell (over 10-runs of the dilute large model cell, the average number of repression events is 1.2; over 10-runs of the model cell with increased biosynthetic capacity, the average number of repression events is 6).

When we let $k_1 = 10/NA\ \mu m^3 molecules^{-1} min^{-1}$ in the small model cell, there is a binding radius-to-nuclear radius ratio of 0.25 (0.75 µm/3.00 µm = 0.25), and we choose a reaction rate for the large model cells such that the resulting binding radius yields the same ratio. A forward binding rate of $k_1 = 40/NA\ \mu m^3 molecules^{-1} min^{-1}$ yields a binding radius of ∼1.6 µm and therefore a binding-to-nuclear radius ratio of 0.25 (1.6 µm/6.46 µm ∼ 0.25), so we set this as our forward rate of protein-promoter binding in both large model cells. We therefore define 2 different small model cells, one for which the relative binding radius is held constant across cell sizes at a binding-to-nuclear radius, and another for which the absolute binding radius is held constant across cell sizes. Excluding the rate of protein-promoter association, k_1, all parameters are held constant across both small model cells as well as the dilute large model cells.

Basal and repressed transcription rates, Tx and Tx_{rep} In both small cell models and the dilute large cell model, basal transcription is set to $Tx = 3.00\ molecules/min$ (Sturrock et al. 2013). In line with empirical observations (Sun et al. 2020, Xie et al. 2022), we ramp up the rate of transcription to promote concentration homeostasis in the second large model cell.

While increasing the rate of transcription, we noticed a trade-off between the rate of basal transcription and promoter repression activity: when the rate of transcription is too high (at or above $\sim$ 6.00 molecules/min), protein-promoter binding either never or only rarely happens over the 3,100 minute run. We believe this is a model limitation arising from the dependence of both reactions (basal transcription and protein-promoter binding) on pF (the free promoter) as a reactant. Nonetheless, transcription repression is an integral component of the autoregulatory negative feedback loop thought to underpin the segmentation clock and therefore of any mathematical model of the segmentation clock as well, so it is necessary to keep the rate of basal transcription low enough to allow for regular protein-promoter binding reactions. Letting $Tx = 5.00$ molecules/min allows for regular protein-promoter binding reactions (i.e. regular transcription repression), while reflecting an almost 70% increase in the rate of basal transcription relative to the small and dilute model cells, so we set $Tx = 5.00$ molecules/min in the second large model cell to promote protein concentration homeostasis. For all cells, repressed transcription is set equal to 10% of the basal transcription rate: $T_{rep} = 0.3$ molecules/min in the small and large dilute model cells, and $T_{rep} = 0.5$ molecules/min in the additional large model cell.

Translation rate, Tl Following the rate of translation used in Sturrock et al. [2013], we set $Tl = 1.00 molecules/min$. We hold this rate constant across all model cells with the vast amount of literature pointing towards transcriptional control as a primary means of maintaining appropriate protein concentration levels in mind (Sun et al. 2020, Xie et al. 2022). However, we note that there is evidence to suggest an important role for translational control over protein levels as well (Lanz

et al. 2021, Liu et al. 2021), so we consider an increase in Tl in the large homeostasis cell as a possible future addition to our model.

Rate of protein-promoter dissociation, k_{-1} Following Sturrock et al. [2013], we set $k_{-1} = 0.1$ molecules/min. As the authors in Sturrock et al. [2013] point out, there is very little known about rates of protein-promoter dissociation, so we hold this rate constant across all model cells.

Degradation rates, α_m and α_p In the small model cells and large dilute model cell, we set the rates of mRNA and protein degradation to $\alpha_m = 0.069$ and $\alpha_p = 0.033$ corresponding to half-lives of $hl_m \sim 10$ and $hl_p \sim 21$ minutes, respectively (given $\alpha_m = ln(2)/h_m$ and $\alpha_p = ln(2)/h_p$). We choose these degradation rates because they complement the range of mRNA and protein stability tested in Chapter 1. These stabilities also match empirical observations of Hes7 in mice (Matsuda et al. 2020, Takashima et al. 2011), so they make plausible reaction rates for a segmentation clock reaction system.

In line with empirical observations that gene product stability, in addition to rates of synthesis, act to support the maintenance of concentration homeostasis in growing cells (Lanz et al. 2021, Liu et al. 2021, Sun et al. 2020, Xie et al. 2022), we increase the stability of both mRNA and protein molecules in the second large model cell. Similar to how we choose an increased basal rate of transcription in the large homeostasis model cell, we test out different rates of mRNA and protein degradation and pick reasonable rates (i.e. rates that fall within the range of observed Hes/Her gene stability) that facilitate regular protein-promoter occupation. We then checked to see if copy numbers were greater in test runs of large homeostasis cell than test runs in the small cell runs, and they were by 100s of molecules. However, we weren't able to verify if concentration homeostasis was truly achieved until all model cells were simulated 100-times. True concentration homeostasis would be achieved if protein copy numbers in the second large cell were, on average, 10-times greater than copy numbers in the small model cells.

Reading in data to Matlab for analyses

Smoldyn outputs data into a text file with columns corresponding to time and molecular copy numbers over time. We save the text file associated with each of the 100 runs per model cell, and we then read each text file into Matlab. In Matlab, we save the protein trajectories for each model cell as a 3,101 by 100 matrix with rows corresponding to each time step (starting at 0 and going to 3100) and columns protein counts over time for each of the 100 runs. We also save protein-promoter binding information into a 3,101 by 100 matrix where each column is comprised of 0s and 1s representing time steps for which the promoter is unoccupied or occupied, respectively.

Fourier and Continuous Wavelet Transforms to test for oscillations

For each simulated data set, we randomly select 20 protein trajectories and save them as columns in a matrix. To test for oscillations, we perform both a Fourier and Continuous Wavelet Transform (CWT) on each column of each matrix, and we show representative plots from the 4 model cells. We perform a Fourier Transform in Matlab using the Fast Fourier Transform `fft` function performed on the protein time series data subtracted by the mean protein copy number. We limit the range of interest to frequencies between 0 and 0.1 cycles/min, corresponding to periods greater than or equal to 10 mins (the closer the frequency is to 0, the longer the period; a frequency of 0.1 cycles/min corresponds to a period of 10 minutes). We perform a CWT with the Morlet wavelet function in Matlab using the Matlab toolkit WAVOS (Harang et al. 2012) by sending our data to CWT visualization, and plotting a CWT heatmap. For our simulated datasets, we start by setting the minimum and maximum periods between 1 and 500 minutes, respectively, to explore a wide range of periods that captures rates of somite segmentation known to be relatively short, like the $\sim$ 30 minute zebrafish segmentation clock (Hoyle and Ish-Horowicz 2013, Keskin et al. 2018, Lewis 2003), and ones known to be relatively long, like the human segmentation clock (Matsuda

et al. 2020). However, we noticed that dominant periods exhibited by the large cell for which biosynthesis rates were increased tended to trend towards the maximum period of 500 minutes (although it did not always reach 500 minutes), so for these CWT, we increased the maximum period to 1,000 minutes to be certain that we weren't missing important patterns beyond the range of 1 to 500 minute periods.

Comparison against empirical data from cell autonomous oscillator

We also download a time series data set corresponding to experimental observations of Her1 reporter signal intensity in cultured isolated zebrafish tailbud cells (Webb et al. 2016). We run the Fourier and Continuous Wavelet analyses above on a subset of the experimental dataset from Webb et al. [2016], and we compare simulated data transform results.

For CWTs of the empirical data set, we set the minimum and maximum periods to 1 and 100 minutes, respectively, since the authors report an average period of oscillation of 42.5 $\pm$11.4 min., and previous reports point towards a 30 minute segmentation clock period in zebrafish *in vivo* (Hoyle and Ish-Horowicz 2013, Keskin et al. 2018). We keep the range of frequency in the Fourier Transform plots the same (between 0 and 0.1 cycles/min).

"First repression events" as a measure of noise

To compare noise levels between our model cells, we identify and store the first-time step at which the promoter site becomes occupied and stays occupied for at least 4 subsequent time steps (i.e. at least 5 minutes total) – we call this the "first repression event", and it serves as the proxy for the time-delay in repression that is required for the operation of the segmentation clock. We identify the "first repression event" for all 100 simulation runs associated with each model cell. For each model cell, we calculate the average and standard deviation of the "first repression

event" over all 100 runs, and we use these two measurements to calculate the coefficient of variation as a measure of noise associated with the occurrence of a "first repression event" in each model cell.

Results

Fourier and Continuous Wavelet transforms reveal noisiness of simulated relative to empirical time series data

In Figure 4, we show representative plots of both Fourier and Continuous Wavelet Transforms for each of the 5 sets of time series data (4 from model cell simulations and 1 from experimental observations). We also give the average maximum magnitude of fit achieved by each data set across the 20 randomly selected protein trajectories that the CWT is performed on.

Comparing first the Fourier Transform plots between groups, we see that clear peak frequencies can be identified in the FTs of empirical data from Webb et al. [2016]. However, we cannot identify clear dominant frequencies in any of the model cell simulations, indicating that simulated protein expression is much noisier than empirical data tracking protein expression in autonomous PSM cells. We also compare between CWTs for each data set and find that while all time series data exhibit certain dominant periods of gene expression over time, the simulated data sets tend to lack the consistency in dominant period trends over time. For example, in Figure 4 F, we see that the dominant period of oscillation over time (shown in red) comes in patches rather than a thick line, like what we see in Figure 4 J, and that these patches don't always to correspond to the same period. There are also important differences in the magnitude of fit between the CWTs of the different data-sets. To the left of all the plots, we provide the average maximum magnitude of fit across the 20 CWT plots for each data set. While the small and dilute model cells exhibit relatively low maximum magnitudes of fit (Small cell 1: 6.74; Small

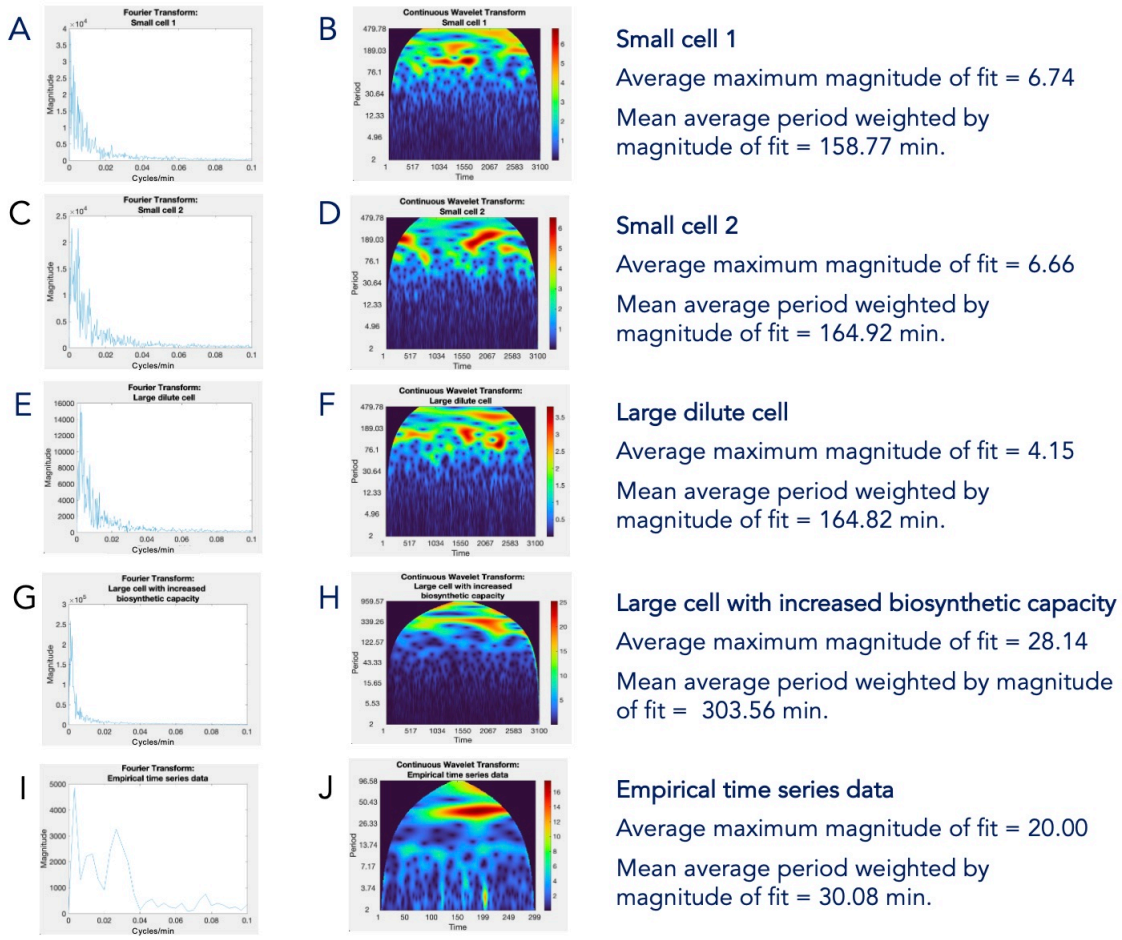

Figure 4: A, C, E G, I: Representative Fourier Transform plots and B, D, F, H, J: representative Continuous Wavelet Plots for each model cell and empirical time series data from Webb et al. [2016]. Average maximum magnitude of fit and mean average period weighted by magnitude of over 20 plots from each data set are given to the right of the corresponding transform plots.

cell 2: 6.66; Large dilute cell: 4.15), CWTs of the second large model cell and empirical time series data exhibit relatively high magnitudes of fit (Large cell 2: 28.14; Empirical: 20.00). This is a kind proxy for measuring expression noise; lower magnitudes of fit imply higher levels of variability in the observed patterns, and vice versa. CWTs of the large model cell with increased biosynthetic capacity, and empirical time series data also seem to exhibit more consistency in the dominant period of gene expression across the measured time range.

Also to the left side of the CWT plots in Figure 4, we provide a mean weighted average period over the 20 CWT plots for each data set. We notice that while the small and dilute model cells

Table 6: Protein levels, repression, and search process noise in different model cells

	General measures		Search process time measures		
	Average protein copy number	Average no. repression events	Mean (min.)	SD (min.)	CV
Small cell 1	144.15	43.66	131.75	91.80	0.70
Small cell 2 (40/na)	122.05	92.36	53.90	31.30	0.58
Dilute cell	47.83	12.10	436.74	358.29	0.82
Homeostasis cell	998.01	17.40	310.32	231.60	0.75

tend to exhibit relatively similar average periods of oscillation, albeit with relative poor fits, the mean weighted average for the large cell with increased biosynthetic capacity almost doubles relative to the smaller cells.

Model cells exhibit different protein levels and frequency of repression events

In Table 6, we include general measures of protein expression kinetics observed across 100-simulations of each model cell. We provide the average protein copy number and the average number of repression events measured over 100 simulations of 3,1000 minutes in each model cell. Though the magnitudes of fit exhibited by the different versions of the small model cell are relatively similar – implying also relatively similar levels of expression noise – we note that the average protein copy number and number of repression events (shown in Table 6) are quite different between the small model cells. An increase in the binding radius (resulting from an increase in the forward rate of protein-promoter association) results in an increase in the average number of repression events, and a decrease in average protein levels.

We also see that an increase in cell volume results in a decrease in the average number of repression events identified across 100-runs, but that the average number of repression events is lowest in the dilute large model cell. Similarly, the average protein copy number is also lowest in the dilute large model cell. The average protein copy number over 100-runs of the second large

model cell does not achieve complete concentration homeostasis of protein relative to the smaller cells (this would require a 10-fold increase in average protein copy number, and therefore between $\sim 1,200$ and $1,400$ protein molecules). Instead, we end with an average of $\sim 1,000$ protein molecules, suggesting that model parameters need additional adjustments to model concentration homeostasis in the larger cell.

Search process noise increases with volume in model cells and is especially pronounced when biosynthetic capacity is not increased

Also in Table 6, we provide descriptive statistics associated with search process noise, which serves as our primary measure of expression noise in this chapter. Again, the statistics provided for each model cell are measured over 100-runs. We first see that the average time needed for a "first repression event" to occur, or, alternatively, the average total search process time between the emergence of an mRNA molecule and the occurrence of transcriptional repression induced by protein-promoter binding, decreases with cell volume. Comparing the small model cells, the average and standard deviation of search process time decreases with an increased protein-free promoter binding radius. Comparing between the large model cells, the average and standard deviation of search process time decrease with increasing protein concentration.

The coefficient of variation acts as our measure of search process noise. Comparing between the two small model cells, search process noise decreases as the forward rate of protein-promoter binding, and therefore binding radius, increases. We also see that, regardless of which small cell we take as our baseline comparison, expression noise increases with cell volume. However, the level of search process noise is at its highest when we model increasing cell volume without an accompanying increase in biosynthetic capacity.

Discussion

Simulations and empirical data suggest that cell-autonomous oscillators are noisy, yet developmental outcomes remain relatively homogeneous

The Fourier and Continuous Wavelet Transforms plotted in Figure 4 reveal that our simulated time series data demonstrate high levels of expression noise related to oscillation period, specifically for small and dilute model cell simulations. Although the large homeostasis cell demonstrates low levels of period noise and more consistency in dominant period trends relative to the other model cells (see CWT plots), it remains difficult to identify dominant frequencies in the corresponding Fourier Transform plots. We therefore cannot conclude that our model cell simulations exhibit robust mathematical oscillations as we initially expected. This remains an ongoing area of exploration.

It is interesting to note that FTs of the homeostasis cell simulation data appear to be less noisy than data from the large dilute cell as well as both small cell simulations. CWT plots corraborate this observation. The representative large homeostasis cell CWT plots exhibit more consistent and better supported patterns of dominant gene expression period over time than the small and dilute model cells. It is also interesting to note that the magnitudes of fit associated with dominant periods of gene expression are higher for CWTs of the homeostasis large cell compared relative to empirical data. This result however can be explained by the increased protein numbers in the homeostasis cell, and the fact that the analysis of the period naturally depends upon long time trajectories. At these longer times the system behaves as a well mixed system.

Looking specifically at transforms of empirical time series data from Webb et al., we notice that both the Fourier and Continuous Wavelet transforms imply relatively low levels of expression noise compared to simulated time series data. The Fourier Transform plots in particular

reveal this pattern. In contrast to the Fourier Transforms on our simulated time series data, one can easily identify a few dominant frequencies associated with the empirical time series data. This contrast holds true even when comparing to large homeostasis cell simulation results, which is not necessarily revealed by the CWT.

Furthermore, the second highest peak in panel I, located at a frequency of ~ 0.027 cycles/min, corresponds to a period of ~ 37 minutes (Period = min/cycle). This agrees well with empirical observations that put the zebrafish segmentation clock at around 30 minutes (Hoyle and Ish-Horowicz 2013, Keskin et al. 2018). CWTs taken of the empirical time series data exhibit similarly low levels of noise compared to the small and dilute model cells, specifically, and dominant trends located between ~ 26 and ~ 50 minute periods, also agree well with previous measurements of the zebrafish segmentation clock. However, looking back at the FT shown in panel I, the dominant frequency peaks at ~ 0.003 cycles/min, corresponding to a period of about 333 minutes, and there is an additional peak between ~ 0.003 and ~ 0.027 cycles/min that is also relatively prominent. These results suggest that there is still a notable amount of noise impacting the expression of Her1 in cultured cells, and that empirical results are still noisier than one might expect given the existing segmentation clock literature.

For several decades the segmentation clock was modeled as a delay-differential equation system, that yielded smooth and regular oscillations. However, more recently the advent of powerful imaging techniques have led to the realization that the segmentation clock in vivo shows significant stochastic noise (Delaune et al. 2012). Experiments that studied the single-cell oscillator that underlie the clock and wavefront model have revealed that the single-cell oscillator is indeed a very noisy oscillator (Keskin et al. 2018, Webb et al. 2016). Thus, we are left with the question of how PSM cells cope with high levels of expression noise to produce a homogenous developmental outcome. Some explanations already exist. In fact, cell autonomous oscillations

tend to exhibit more variability in expression period and amplitude than PSM cells in vivo (Webb et al. 2016), suggesting an additional noise damping role for the cell-to-cell signaling pathways that sync oscillations of neighboring cells. This additional function has already been proposed in the literature, and it is well supported by empirical observations (Keskin et al. 2018).

Average periods of oscillation suggested by Continuous Wavelet Transforms of simulated data could imply that volume alone does not slow oscillations

When comparing the mean weighted average periods of oscillations between different model cells, as shown in Figure 4, we notice the small and dilute model cells exhibit relatively similar average periods of oscillation (between 155 and 165 minutes). Meanwhile, the average period across 20 CWTs of simulated data from the large model cell with increased biosynthetic capacity almost doubles relative to the average periods of oscillation measured in the small model cells. This result could imply that increasing volume alone does not slow the period of oscillation, but that biosynthetic capacity must also be adjusted to slow oscillations. This is in line with results from Chapter 1 that suggest important roles for mRNA and protein stability in the mediation of developmental tempo. However, it is worth noting that comparisons between different model cell CWTs must be made with caution since the levels of noise so vastly differ between the large cell for which biosynthetic capacity is increased and all other model cells.

Search process and repression patterns relative to binding radius size, protein concentration, and cell volume complement results from chapter 1

In Table 6, we show the average protein copy number and average numbers of repression events across 100-runs of each model cell in addition to measurements related to search process time and noise. We observe intuitive patterns related to protein copy number, frequency of repression

events, and search process time relative to cell volume and protein concentration. These patterns are helpful checks that, in general, our model is exhibiting expected patterns, which helps to provide support for less intuitive and more novel results related to search process noise. We start with general observations across increasing cell volume that hold true regardless of which small and large model cell variations are compared. First, we observe that the average search process time increases with cell volume. This is intuitive simply because we would naturally expect the cumulative journeys of an mRNA molecule from the nuclear center to a ribosome in the cytoplasm, and a protein molecule from the site of translation back to the nuclear center to take longer in a larger cell when diffusion coefficients are held constant, as they are here. For similar reasons, we observe that when cell volume increases, the total occurrence of repression events (i.e. sustained protein-promoter binding) over an equal time span (3,100 minutes) decreases. Relative levels of mRNA and protein stability with respect to cell volume may also contribute to this phenomena if mRNA and protein molecules are unable to persist long enough to leave the nucleus and return to the promoter site, respectively, in a dramatically larger cell. This also explains differences observed between large model cells when degradation rates (i.e. species stability) are not adjusted (dilute cell), and when they are adjusted (homeostasis cell). When comparing the dilute and homeostasis large cells, we see that increased transcriptional rate and gene product stabilities act to decrease the average cumulative search process time, likely due to the fact that more proteins persist long enough to return to the promoter region in the homeostasis relative to the dilute cell. Additionally, increased transcriptional rate and gene product stabilities act to increase the number of protein molecules (and therefore also the concentration of protein) within the cell, as well as the number of repression events that take place over the same time period.

In contrast to differences observed between the two large model cells, which are primarily driven by differences in both mRNA and protein stability, and the rate of mRNA synthesis,

differences observed between the two small model cells reflect the impact of binding reaction rates and the resulting binding radius. When the rate of protein-promoter association, k_1 and binding radius increase, we observe an increase in the average number of repression events and a decrease in the average time needed for repression to first happen, both likely due to the fact that the promoter region is more quickly found when it is bigger. Likely as a result of increasing frequency of transcriptional repression, we also see lower protein levels when the binding rate and radius are increased.

Results indicate that "search process noise" increases when protein concentration homeostasis is not maintained

Consider the "search process" undergone by mRNA as it leaves the promoter site, travels to the nuclear membrane, and then to a ribosome after exiting the nucleus. This is followed by a subsequent search process undergone by the protein molecule as it travels through the cytoplasm back to the nucleus, through the nuclear membrane, and then to the promoter site where it binds to act as a transcriptional repressor. These back-to-back search processes culminate in the occupation of the promoter site. When we set initial mRNA and protein values equal to 0, measuring the "first repression event" as defined by us is also a measure of the total delay between mRNA production and transcriptional repression induced by protein-promoter binding. In-line with results from Chapter 1, we find that the average search process time increases with cell volume. We also find that the average search process takes longer in the large dilute relative to the large homeostasis cell. This is likely due to the lack of increased gene product stability to complement increasing cell volume in the large dilute cell. If protein molecules are less stable, it makes sense that less of them will return to the promoter site, which would increase the average time needed for the first (sustained) protein-promoter binding reaction to occur. The lack of

protein stability likely also contributes to the increased coefficient of variation measured in the large dilute cell relative to the homeostasis and small cells. The larger coefficient of variation in the large homeostasis cell relative to the small model cells is mainly due to increasing volume, but since perfect concentration homeostasis was not achieved, there is possibly a small contribution of noise coming from remaining dilution in the system. The average protein copy number observed falls between an almost 7 and slightly over an 8 fold change relative to small model cells 1 and 2, which falls short of the 10-fold increase in copy number needed to match the 10-fold increase in volume to maintain exact protein concentration homeostasis.

Model limitations and future considerations

Throughout this chapter, we have touched on a handful of model limitations. In this section, we expand upon limitations that have already been touched on, and we introduce a few more. We also discuss future considerations and directions for the spatial stochastic model.

We first recognize that the rate of translation contributes to overall biosynthetic capacity, and that translational control has been proposed as a means of maintaining proteomic concentration homeostasis. This is not a model limitation per se, as we could easily increase the rate of translation, we simply started by only increasing the rate of transcription. As shown in Table 6, doing so revealed that increasing the transcription rate and gene product stabilites alone is no sufficient to achieve full proteomic concentration in the larger model cell. In the future, we would be interested in increasing the rate of translation in the larger model cell to see if this yields full proteomic concentration homeostasis, and to see the resulting impact of achieving full concentration homeostasis on search process noise in the large model cell.

A true model limitation that we discuss in the Methods section is the tradeoff that exists between the rate of transcription and protein-promoter binding affinity. We believe this tradeoff

is due to the fact that both reactions are dependent on the free promoter, so increasing the basal rate of transcription decreases the ability of the free promoter to be involved in other reactions. This puts a cap on how high we can currently set the basal rate of transcription while still maintaining the crucial auto-repressive component of the negative feedback underlying the segmentation clock gene regulatory network, which could become problematic if we have a biologically realistic increased rate of basal transcription that is higher than the model imposed cap. It will be important to continue to explore this potential tradeoff, and to find a work around, if necessary.

Finally, we would like to propose as a future direction adding some layers of complexity to our very simple single cell model. For example, it would be relatively easy to add ribosomes dispersed throughout the cytoplasm, and to adjust the translation reaction accordingly. This is a simple addition, but it could yield informative changes in the model gene expression kinetics. Additionally, it is also well known that many genes have multiple promoter sites from which transcription and transcriptional repression can originate (Guisoni et al. 2016). It would also be interesting to incorporate multiple promoter sites into the model. Increasing the number of promoter sites could potentially alleviate the tradeoff between transcriptional rate and protein-promoter association discussed above.

Conclusion

In this chapter, we have shown that both simulated and empirical time series gene expression data are noisier than we would expect given a clock-and-wavefront somitogenesis model that is underpinned by a cell-autonomous segmentation clock. Instead, Fourier and Continuous Wavelet Transforms performed on simulated and empirical datasets suggest first that, in a cell autonomous system, true robust oscillations may not exists, and second that if cell autonomous

oscillations do exist, they are subject to high levels expression noise. This second implication of our results supports a previously proposed role for cell-to-cell signaling pathways as a means of damping noise associated with cell-autonomous oscillatory gene expression. Finally, we use measure ”search process” kinetics to show that an increase in cell volume slows down the process of auto-repression, and increases intracellular expression noise, and that the second effect is especially pronounced when transcriptional rates and gene product stabilities are not adjusted for volume. To our knowledge, this is a novel finding.

Concluding Remarks

In the previous chapters, we have explored how developmental tempo and stochastic gene expression noise interact with genome size and nuclear volume, and with cell volume, respectively. Throughout both chapters, use the same model system: the segmentation clock component of somitogenesis, a developmental process conserved across vertebrates that operates in species-specific ways, yet we adopt different approaches in each chapter to answer specific questions. In Chapter 1, we use a deterministic model of the segmentation clock to better understand potential mechanisms that drive slowed developmental rate across increasing genome size and nuclear volume. In Chapter 2, we use a stochastic model of the auto-regulatory gene network thought to underlie the segmentation clock to explore a potential relationship between cell volume and intracellular gene expression noise. Some important general conclusions remain the same across both models. Both models suggest that increasing intracellular distances between target sites in cells of larger volumes slow fundamental "search processes," like nuclear export and auto-repression of transcription, down which in turn act to slow developmental processes as well. Both models also suggest an important role for gene product, i.e. mRNA and protein, stability when it comes to mediating development tempo and expression noise across increasing cell volume. However, we also see that adding stochasticity to a model of gene expression poses questions about the functionality of cell autonomous oscillators that are taken for granted by the deterministic model. We find that while, individually, each model proposes novel findings, they offer a more complete view together.

References

S. S. Andrews, N. J. Addy, R. Brent, and A. P. Arkin. Detailed simulations of cell biology with Smoldyn 2.1. *PLoS Computational Biology*, 6(3), 2010. ISSN 15537358. doi: 10.1371/journal.pcbi.1000705.

J. B. Armstrong and A. C. Graveson. Progressive patterning precedes somite segmentation in the mexican axolotl (Ambystoma mexicanum). *Developmental Biology*, 126(1):1–6, 1988. ISSN 00121606. doi: 10.1016/0012-1606(88)90232-1.

A. Aulehla and O. Pourquié. Signaling gradients during paraxial mesoderm development. *Cold Spring Harbor Perspectives in Biology*, 2(2), 2010. ISSN 19430264. doi: 10.1101/cshperspect.a000869.

S. Banfi, L. Monti, F. Acquati, G. Tettamanti, M. de Eguileor, and A. Grimaldi. Muscle development and differentiation in the urodele Ambystoma mexicanum. *Development Growth and Differentiation*, 54(4):489–502, 5 2012. ISSN 00121592. doi: 10.1111/j.1440-169X.2012.01338.x.

Y. Ben-Ari, Y. Brody, N. Kinor, A. Mor, T. Tsukamoto, D. L. Spector, R. H. Singer, and Y. Shav-Tal. The life of an mRNA in space and time. *Journal of Cell Science*, 123(10):1761–1774, 5 2010. ISSN 00219533. doi: 10.1242/jcs.062638.

R. Ben-Yishay and Y. Shav-Tal. The dynamic lifecycle of mRNA in the nucleus. *Current Opinion in Cell Biology*, 58:69–75, 6 2019. ISSN 18790410. doi: 10.1016/j.ceb.2019.02.007.

Y. Bessho, R. Sakata, S. Komatsu, K. Shiota, S. Yamada, and R. Kageyama. Dynamic expression

and essential functions of Hes7 in somite segmentation. *Genes and Development*, 15(20):2642–2647, 10 2001. ISSN 08909369. doi: 10.1101/gad.930601.

A. Bizhanova and P. D. Kaufman. Close to the edge: Heterochromatin at the nucleolar and nuclear peripheries. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1864(1), 1 2021. ISSN 18764320. doi: 10.1016/j.bbagrm.2020.194666.

R. Blewitt. Enhancer of split-related-2 mRNA shows cyclic expression during somitogenesis in Xenopus laevis. *Bioscience Horizons*, 2(1):22–31, 3 2009. ISSN 17547431. doi: 10.1093/biohorizons/hzp006.

S. Boyle, S. Gilchrist, J. M. Bridger, N. L. Mahy, J. A. Ellis, and W. A. Bickmore. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Human Molecular Genetics*, 10(3):211–219, 2001.

G. C. Brown. Total Cell Protein Concentration as an Evolutionary Constraint on the Metabolic Control Distribution in Cells. *Journal of Theoretical Biology*, 153:195–203, 1991.

C. Cadart, J. Bartz, G. Oaks, M. Z. Liu, and R. Heald. Polyploidy in Xenopus lowers metabolic rate by decreasing total cell surface area. *Current Biology*, 33(9):1744–1752, 5 2023. ISSN 18790445. doi: 10.1016/j.cub.2023.03.071.

J. Cooke and E. C. Zeeman. A clock and wavefront model for control of the number of repeated structures during animal morphogenesis. *Journal of Theoretical Biology*, 58(2):455–476, 1976.

K. L. Curran, L. Allen, B. B. Porter, J. Dodge, C. Lope, G. Willadsen, R. Fisher, N. Johnson, E. Campbell, B. VonBergen, D. Winfrey, M. Hadley, and T. Kerndt. Circadian genes, xBmal1 and xNocturnin, modulate the timing and differentiation of somites in Xenopus laevis. *PLoS ONE*, 9(9), 2014. ISSN 19326203. doi: 10.1371/journal.pone.0108266.

L. A. Dali, J. Gustin, K. Perry, C. R. Domingo, and C. Domingo. Signals that instruct somite and myotome formation persist in Xenopus laevis early tailbud stage embryos. *Cells Tissues Organs*, 172:1–12, 2002. URL www.karger.com.

R. L. Davis, D. L. Turner, L. M. Evans, and M. W. Kirschner. Molecular targets of vertebrate segmentation: Two mechanisms control segmental expression of Xenopus hairy2 during somite formation. *Developmental Cell*, 1:553–565, 2001.

E. A. Delaune, P. François, N. P. Shih, and S. L. Amacher. Single-Cell-Resolution Imaging of the Impact of Notch Signaling and Mitosis on Segmentation Clock Dynamics. *Developmental Cell*, 23(5):995–1005, 11 2012. ISSN 15345807. doi: 10.1016/j.devcel.2012.09.009.

M. Diaz-Cuadros, T. P. Miettinen, O. S. Skinner, D. Sheedy, C. M. Díaz-García, S. Gapon, A. Hubaud, G. Yellen, S. R. Manalis, W. M. Oldham, and O. Pourquié. Metabolic regulation of species-specific developmental rates. *Nature*, 613(7944):550–557, 1 2023. ISSN 14764687. doi: 10.1038/s41586-022-05574-4.

W. L. Eckalbar, E. Lasku, C. R. Infante, R. M. Elsey, G. J. Markov, A. N. Allen, J. J. Corneveaux, J. B. Losos, D. F. DeNardo, M. J. Huentelman, J. Wilson-Rawls, A. Rawls, and K. Kusumi. Somitogenesis in the anole lizard and alligator reveals evolutionary convergence and divergence in the amniote segmentation clock. *Developmental Biology*, 363(1):308–319, 3 2012. ISSN 1095564X. doi: 10.1016/j.ydbio.2011.11.021.

J. Faber and P. Nieuwkoop, editors. *Normal Table of Xenopus Laevis (Daudin)*. Garland, New York, 1994.

S. Gibb, M. Maroto, and J. K. Dale. The segmentation clock mechanism moves up a notch.

Trends in Cell Biology, 20(10):593–600, 2010. ISSN 09628924. doi: 10.1016/j.tcb.2010.07.001. URL http://dx.doi.org/10.1016/j.tcb.2010.07.001.

D. Gomez, R. Marathe, V. Bierbaum, and S. Klumpp. Modeling stochastic gene expression in growing cells. *Journal of Theoretical Biology*, 348:1–11, 5 2014. ISSN 00225193. doi: 10.1016/j.jtbi.2014.01.017.

D. Gonze. Modeling the effect of cell division on genetic oscillators. *Journal of Theoretical Biology*, 325:22–33, 5 2013. ISSN 00225193. doi: 10.1016/j.jtbi.2013.02.001.

T. R. Gregory. The bigger the C-value, the larger the cell: Genome size and red blood cell size in vertebrates. *Blood Cells, Molecules, and Diseases*, 27(5):830–843, 2001. ISSN 10799796. doi: 10.1006/bcmd.2001.0457.

T. R. Gregory. Animal Genome Size Database, 2023. URL http://www.genomesize.com.

N. Guisoni, D. Monteoliva, and L. Diambra. Promoters architecture-based mechanism for noise-induced oscillations in a single-gene circuit. *PLoS ONE*, 11(3), 3 2016. ISSN 19326203. doi: 10.1371/journal.pone.0151086.

E. S. Haag and J. R. True. Developmental system drift. In L. Nuño de la Rosa and G. B. Müller, editors, *Evolutionary Developmental Biology*, pages 99–110. Springer, 2021.

L. Hamilton. The formation of somites in Xenopus. *Journal of Embryology and Experimental Morphology*, 22(2):253–264, 1969.

R. Harang, G. Bonnet, and L. R. Petzold. WAVOS: A MATLAB toolkit for wavelet analysis and visualization of oscillatory systems. *BMC Research Notes*, 5, 2012. ISSN 17560500. doi: 10.1186/1756-0500-5-163.

J. Hausser, A. Mayo, L. Keren, and U. Alon. Central dogma rates and the trade-off between precision and economy in gene expression. *Nature Communications*, 10(1), 12 2019. ISSN 20411723. doi: 10.1038/s41467-018-07391-8.

U. Hellsten, R. M. Harland, M. J. Gilchrist, D. Hendrix, J. Jurka, V. Kapitonov, I. Ovcharenko, N. H. Putnam, S. Shu, L. Taher, I. L. Blitz, B. Blumberg, D. S. Dichmann, L. Dubchak, E. Amaya, J. C. Detter, R. Fletcher, D. S. Gerhard, D. Goodstein, T. Graves, I. V. Grigoriev, J. Grimwood, T. Kawashima, E. Lindquist, S. M. Lucas, P. E. Mead, T. Mitros, H. Ogino, Y. Ohta, A. V. Poliakov, N. Pollet, J. Robert, A. Salamov, A. K. Sater, J. Schmutz, A. Terry, P. D. Vize, W. C. Warren, D. Wells, A. Wills, R. K. Wilson, L. B. Zimmerman, A. M. Zorn, R. Grainger, T. Grammer, M. K. Khokha, P. M. Richardson, and D. S. Rokhsar. The genome of the western clawed frog Xenopus tropicalis. *Science*, 328(5978):633–636, 4 2010. ISSN 10959203. doi: 10.1126/science.1183670.

M. Hidalgo, C. Sirour, V. Bello, N. Moreau, M. Beaudry, and T. Darribère. In vivo analyzes of dystroglycan function during somitogenesis in Xenopus laevis. *Developmental Dynamics*, 238 (6):1332–1345, 6 2009. ISSN 10588388. doi: 10.1002/dvdy.21814.

H. Hirata, Y. Bessho, H. Kokubu, Y. Masamizu, S. Yamada, J. Lewis, and R. Kageyama. Instability of Hes7 protein is crucial for the somite segmentation clock. *Nature Genetics*, 36(7):750–754, 2004. ISSN 10614036. doi: 10.1038/ng1372.

S. A. Holley, D. Jülich, G.-J. Rauch, R. Geisler, and C. Nüsslein-Volhard. her1 and the notch pathway function within the oscillator mechanism that regulates zebrafish somitogenesis. *Development*, 129(5):1175–1183, 2002.

N. P. Hoyle and D. Ish-Horowicz. Transcript processing and export kinetics are rate-limiting steps in expressing vertebrate segmentation clock genes. *Proceedings of the National Academy of Sciences of the United States of America*, 110(46), 2013. ISSN 00278424. doi: 10.1073/pnas.1308811110.

Y. Ishihama and T. Funatsu. Single molecule tracking of quantum dot-labeled mRNAs in a cell nucleus. *Biochemical and Biophysical Research Communications*, 381(1):33–38, 3 2009. ISSN 0006291X. doi: 10.1016/j.bbrc.2009.02.001.

W.-C. Jen, V. Gawantka, N. Pollet, C. Niehrs, and C. Kintner. Periodic repression of Notch pathway genes governs the segmentation of Xenopus embryos. *Genes and Development*, 13(11): 1486–1499, 1999. URL www.genesdev.org.

J.-H. Jeon, A. V. Chechkin, and R. Metzler. First passage behavior of multi-dimensional fractional Brownian Motion and application to reaction phenomena. In R. Metlzer, G. Oshanin, and S. Redner, editors, *First-Passage Phenomena and Their Applications*, chapter 8, pages 175–202. World Scientific, 5 2014.

E. L. Jockusch. An evolutionary correlate of genome size change in plethodontid salamanders. *Proceedings of the Royal Society of London Series B*, pages 597–604, 1997. URL https://royalsocietypublishing.org/.

R. Kageyama, T. Ohtsuka, and T. Kobayashi. The Hes gene family: Repressors and oscillators that orchestrate embryogenesis. *Development*, 134(7):1243–1251, 4 2007. ISSN 09501991. doi: 10.1242/dev.000786.

S. Keskin, G. S. Devakanmalai, S. B. Kwon, H. T. Vu, Q. Hong, Y. Y. Lee, M. Soltani, A. Singh, A. Ay, and E. M. Özbudak. Noise in the vertebrate segmentation clock is boosted by time delays but tamed by notch signaling. *Cell Reports*, 23(7):2175–2185, 2018. ISSN 22111247. doi: 10.1016/j.celrep.2018.04.069.

Y. L. Khodor, J. S. Menet, M. Tolan, and M. Rosbash. Cotranscriptional splicing efficiency differs

dramatically between Drosophila and mouse. *RNA*, 18(12):2174–2186, 12 2012. ISSN 13558382. doi: 10.1261/rna.034090.112.

B. T. Kile, K. E. Hentges, A. T. Clark, H. Nakamura, A. P. Salinger, B. Liu, N. Box, D. W. Stockton, R. L. Johnson, R. R. Behringer, A. Bradley, and M. J. Justice. Functional genetic analysis of mouse chromosome 11. *Nature*, 425(6953):81–85, 9 2003. ISSN 00280836. doi: 10.1038/nature01931.

J. Klepstad and L. Marcon. A clock and wavefront self-organizing model explains somitogenesis in vivo and in vitro. *bioRxiv*, 2023. doi: 10.1101/2023.01.18.524516. URL https://doi.org/10.1101/2023.01.18.524516.

A. J. Krol, D. Roellig, M. L. Dequéant, O. Tassy, E. Glynn, G. Hattem, A. Mushegian, A. C. Oates, and O. Pourquié. Evolutionary plasticity of segmentation clock networks. *Development*, 138 (13):2783–2792, 7 2011. ISSN 09501991. doi: 10.1242/dev.063834.

A. Kuretani, T. Yamamoto, M. Taira, and T. Michiue. Evolution of hes gene family in vertebrates: the hes5 cluster genes have specifically increased in frogs. *BMC Ecology and Evolution*, 21(1), 12 2021. ISSN 14726785. doi: 10.1186/s12862-021-01879-6.

T. J. Lampo, S. Stylianidou, M. P. Backlund, P. A. Wiggins, and A. J. Spakowitz. Cytoplasmic RNA-protein particles exhibit non-Gaussian subdiffusive behavior. *Biophysical Journal*, 112(3): 532–542, 2 2017. ISSN 15420086. doi: 10.1016/j.bpj.2016.11.3208.

M. C. Lanz, E. Zatulovskiy, M. P. Swaffer, L. Zhang, I. Ilerten, S. Zhang, D. S. You, G. Marinov, P. McAlpine, J. E. Elias, and J. M. Skotheim. Increasing cell size remodels the proteome and promotes senescence. *bioRxiv*, page 2021.07.29.454227, 2021. URL https://www.biorxiv.org/content/10.1101/2021.07.29.454227v3%0Ahttps:

//www.biorxiv.org/content/10.1101/2021.07.29.454227v3.abstract%0Ahttps://www.biorxiv.org/content/10.1101/2021.07.29.454227v3%0Ahttps://www.biorxiv.org/content/10.1101/2021.07.29.454227v3.

J. Lázaro, M. Costanzo, M. Sanaki-Matsumiya, C. Girardot, M. Hayashi, K. Hayashi, S. Diecke, T. B. Hildebrandt, G. Lazzari, J. Wu, S. Petkov, R. Behr, V. Trivedi, M. Matsuda, and M. Ebisuya. A stem cell zoo uncovers intracellular scaling of developmental tempo across mammals. *Cell Stem Cell*, 30(7):938–949, 7 2023. ISSN 19345909. doi: 10.1016/j.stem.2023.05.014. URL https://linkinghub.elsevier.com/retrieve/pii/S1934590923002102.

X. Lei, W. Tian, H. Zhu, T. Chen, and P. Ao. Biological Sources of Intrinsic and Extrinsic Noise in cI Expression of Lysogenic Phage Lambda. *Scientific Reports*, 5, 9 2015. ISSN 20452322. doi: 10.1038/srep13597.

J. Lewis. Autoinhibition with transcriptional delay: A simple mechanism for the zebrafish somitogenesis oscillator. *Current Biology*, 13:1398–1408, 2003. doi: 10.1016/S.

G. W. Li and X. S. Xie. Central dogma at the single-molecule level in living cells, 7 2011. ISSN 00280836.

Y. Li, U. Fenger, C. Niehrs, and N. Pollet. Cyclic expression of esr9 gene in Xenopus presomitic mesoderm. *Differentiation*, 71(1):83–89, 2003. ISSN 03014681. doi: 10.1046/j.1432-0436.2003.700608.x. URL http://dx.doi.org/10.1046/j.1432-0436.2003.700608.x.

J. Lin and A. Amir. Homeostasis of protein and mRNA concentrations in growing cells. *Nature Communications*, 9(1), 2018. ISSN 20411723. doi: 10.1038/s41467-018-06714-z. URL http://dx.doi.org/10.1038/s41467-018-06714-z.

S. Liu, C. Tan, C. Melo-Gavin, K. G. Mark, M. B. Ginzberg, R. Blutrich, N. Patel, M. Rape,

R. Kafri, and . Affiliations. Large cells activate global protein degradation to maintain cell size homeostasis. *bioRxiv*, 2021. doi: 10.1101/2021.11.09.467936. URL https://doi.org/10.1101/2021.11.09.467936.

M. E. Malerba and D. J. Marshall. Larger cells have relatively smaller nuclei across the Tree of Life. *Evolution Letters*, pages 1–9, 2021. ISSN 2056-3744. doi: 10.1002/evl3.243.

M. Matsuda, H. Hayashi, J. Garcia-Ojalvo, K. Yoshioka-Kobayashi, R. Kageyama, Y. Yamanaka, M. Ikeya, J. Toguchida, C. Alev, and M. Ebisuya. Species-specific segmentation clock periods are due to differential biochemical reaction speeds. *Science*, 369(6509):1450–1455, 2020. ISSN 10959203. doi: 10.1126/SCIENCE.ABA7668.

R. Mayer, A. Brero, J. von Hase, T. Schroeder, T. Cremer, and S. Dietzel. Common themes and cell type specific variations of higher order chromatin arrangements in the mouse. *BMC Cell Biology*, 6, 12 2005. ISSN 14712121. doi: 10.1186/1471-2121-6-44.

S. Mitchell and A. Hoffmann. Identifying noise sources governing cell-to-cell variability. *Current Opinion in Systems Biology*, 8:39–45, 4 2018. ISSN 24523100. doi: 10.1016/j.coisb.2017.11.013.

A. Mor and Y. Shav-Tal. Dynamics and kinetics of nucleo-cytoplasmic mRNA export. *Wiley Interdisciplinary Reviews: RNA*, 1(3):388–401, 11 2010. ISSN 17577004. doi: 10.1002/wrna.41.

A. Mor, S. Suliman, R. Ben-Yishay, S. Yunger, Y. Brody, and Y. Shav-Tal. Dynamics of single mRNP nucleocytoplasmic transport and export through the nuclear pore in living cells. *Nature Cell Biology*, 12(6):543–552, 6 2010. ISSN 14657392. doi: 10.1038/ncb2056.

B. Munsky, G. Neuert, and A. Van Oudenaarden. Using Gene Expression Noise to Understand Gene Regulation. *Science*, 336(6078):183–187, 2012. URL https://www.science.org.

G. E. Neurohr, R. L. Terry, J. Lengefeld, M. Bonney, G. P. Brittingham, F. Moretto, T. P. Miettinen, L. P. Vaites, L. M. Soares, J. A. Paulo, J. W. Harper, S. Buratowski, S. Manalis, F. J. van Werven, L. J. Holt, and A. Amon. Excessive Cell Growth Causes Cytoplasm Dilution And Contributes to Senescence. *Cell*, 176(5):1083–1097, 2019. ISSN 10974172. doi: 10.1016/j.cell.2019.01.018. URL https://doi.org/10.1016/j.cell.2019.01.018.

S. Nowoshilow, S. Schloissnig, J. F. Fei, A. Dahl, A. W. Pang, M. Pippel, S. Winkler, A. R. Hastie, G. Young, J. G. Roscito, F. Falcon, D. Knapp, S. Powell, A. Cruz, H. Cao, B. Habermann, M. Hiller, E. M. Tanaka, and E. W. Myers. The axolotl genome and the evolution of key tissue formation regulators. *Nature*, 554(7690):50–55, 2018. ISSN 14764687. doi: 10.1038/nature25458.

A. C. Oates and R. K. Ho. Hairy/E(spl)-related (Her) genes are central components of the segmentation oscillator and display redundancy with the Delta/Notch signaling pathway in the formation of anterior segmental boundaries in the zebrafish. *Development*, 129(12):2929–2946, 2002.

M. Oeffinger and D. Zenklusen. To the pore and through the pore: A story of mRNA export kinetics. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1819(6):494–506, 6 2012. ISSN 18749399. doi: 10.1016/j.bbagrm.2012.02.011.

J. Paijmans, D. K. Lubensky, and P. Rein Ten Wolde. Robustness of synthetic oscillators in growing and dividing cells. *Physical Review E*, 95(5), 5 2017. ISSN 24700053. doi: 10.1103/PhysRevE.95.052403.

V. Pancaldi. Biological noise to get a sense of direction: An analogy between chemotaxis and stress response. *Frontiers in Genetics*, 5(MAR), 2014. ISSN 16648021. doi: 10.3389/fgene.2014.00052.

N. Puzović, T. Madaan, and J. Y. Dutheil. Being noisy in a crowd: Differential selective pressure on gene expression noise in model gene regulatory networks. *PLoS Computational Biology*, 19 (4), 4 2023. ISSN 15537358. doi: 10.1371/journal.pcbi.1010982.

J. M. Raser and E. K. O'Shea. Noise in Gene Expression: Origins, Consequences, and Control. *Science*, 309(5743):2010–2013, 2005. URL https://www.science.org.

T. Rayon, D. Stamataki, R. Perez-Carrasco, L. Garcia-Perez, C. Barrington, M. Melchionda, K. Exelby, J. Lazaro, V. L. Tybulewicz, E. M. Fisher, and J. Briscoe. Species-specific pace of development is associated with differences in protein stability. *Science*, 369(6509), 9 2020. ISSN 10959203. doi: 10.1126/SCIENCE.ABA7667.

A. D. Riggs, S. Bourgeois, and M. Cohn. The Zac Repressor-Operator Interaction. *Journal of Molecular Biology*, 53:401–417, 1970.

G. Roth and W. Walkowiak. The influence of genome and cell size on brain morphology in amphibians. *Cold Spring Harbor Perspectives in Biology*, 7(9), 9 2015. ISSN 19430264. doi: 10.1101/cshperspect.a019075.

E. W. Sayers, E. E. Bolton, J. R. Brister, K. Canese, J. Chan, D. C. Comeau, R. Connor, K. Funk, C. Kelly, S. Kim, T. Madej, A. Marchler-Bauer, C. Lanczycki, S. Lathrop, Z. Lu, F. Thibaud-Nissen, T. Murphy, L. Phan, Y. Skripchenko, T. Tse, J. Wang, R. Williams, B. W. Trawick, K. D. Pruitt, and S. T. Sherry. Database resources of the national center for biotechnology information. *Nucleic Acids Research*, 50(D1):D20–D26, 1 2022. ISSN 13624962. doi: 10.1093/nar/gkab1112.

J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Toman-

cak, and A. Cardona. Fiji: An open-source platform for biological-image analysis. *Nature Methods*, 9(7):676–682, 7 2012. ISSN 15487091. doi: 10.1038/nmeth.2019.

S. Schloissnig, A. Kawaguchi, S. Nowoshilow, F. Falcon, L. Otsuki, P. Tardivo, N. Timoshevskaya, M. C. Keinath, J. J. Smith, S. R. Voss, and E. M. Tanaka. The giant axolotl genome uncovers the evolution, scaling, and transcriptional control of complex gene loci. *Proceedings of the National Academy of Sciences of the United States of America*, 118(15), 2021. ISSN 10916490. doi: 10.1073/pnas.2017176118.

S. K. Sessions. Evolutionary cytogenetics in salamanders. *Chromosome Research*, 16(1):183–201, 3 2008. ISSN 09673849. doi: 10.1007/s10577-007-1205-3.

S. K. Sessions and A. Larson. Developmental correlates of genome Ssize in Plethodontid salamanders and their implications for genome evolution. *Evolution*, 41(6):1239, 1987. ISSN 00143820. doi: 10.2307/2409090.

S. K. Sessions and D. B. Wake. Forever young: Linking regeneration and genome size in salamanders. *Developmental Dynamics*, 250(6):768–778, 2021. ISSN 10970177. doi: 10.1002/dvdy.279.

Y. Shav-Tal, X. Darzacq, S. M. Shenoy, D. Fusco, S. M. Janicki, D. L. Spector, and R. H. Singer. Dynamics of single mRNPs in nuclei of living cells. *Science*, 304(5678):1797–1800, 6 2004. ISSN 00368075. doi: 10.1126/science.1099754.

J. Sheinberger and Y. Shav-Tal. The dynamic pathway of nuclear RNA in eukaryotes. *Nucleus*, 4 (3):195–205, 2013. ISSN 19491042. doi: 10.4161/nucl.24434.

J. Shinga, M. Itoh, K. Shiokawa, S. Taira, and M. Taira. Early patterning of the prospective midbrain-hindbrain boundary by the HES-related gene XHR1 in Xenopus embryos. *Developmental Cell*, 109(2):225–239, 2001. URL www.elsevier.com/locate/modo.

B. M. Smalec, R. Ietswaart, K. Choquet, E. Mcshane, E. R. West, and L. S. Churchman. Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle. *bioRxiv*, 2022. doi: 10.1101/2022.08.21.504696. URL https://doi.org/10.1101/2022.08.21.504696.

J. J. Smith, N. Timoshevskaya, V. A. Timoshevskiy, M. C. Keinath, D. Hardy, and S. R. Voss. A chromosome-scale assembly of the axolotl genome. *Genome Research*, 29(2):317–324, 2 2019. ISSN 15495469. doi: 10.1101/gr.241901.118.

M. Sturrock, A. Hellander, A. Matzavinos, and M. A. Chaplain. Spatial stochastic modelling of the hes1 gene regulatory network: Intrinsic noise can explain heterogeneity in embryonic stem cell differentiation. *Journal of the Royal Society Interface*, 10(80), 3 2013. ISSN 17425662. doi: 10.1098/rsif.2012.0988.

X. M. Sun, A. Bowman, M. Priestman, F. Bertaux, A. Martinez-Segura, W. Tang, C. Whilding, D. Dormann, V. Shahrezaei, and S. Marguerat. Size-Dependent Increase in RNA Polymerase II Initiation Rates Mediates Gene Expression Scaling with Cell Size. *Current Biology*, 30(7): 1217–1230, 4 2020. ISSN 18790445. doi: 10.1016/j.cub.2020.01.053.

I. A. Swinburne and P. A. Silver. Intron delays and transcriptional timing during development. *Developmental Cell*, 14(3):324–330, 2008. ISSN 15345807. doi: 10.1016/j.devcel.2008.02.002.

I. A. Swinburne, D. G. Miguez, D. Landgraf, and P. A. Silver. Intron length increases oscillatory periods of gene expression in animal cells. *Genes and Development*, 22(17):2342–2346, 9 2008. ISSN 08909369. doi: 10.1101/gad.1696108.

A. Tafvizi, L. A. Mirny, and A. M. Van Oijen. Dancing on DNA: Kinetic aspects of search pro-

cesses on DNA. *ChemPhysChem*, 12(8):1481–1489, 6 2011. ISSN 14397641. doi: 10.1002/cphc.201100112.

Y. Takashima, T. Ohtsuka, A. González, H. Miyachi, and R. Kageyama. Intronic delay is essential for oscillatory expression in the segmentation clock. *Proceedings of the National Academy of Sciences of the United States of America*, 108(8):3300–3305, 2011. ISSN 00278424. doi: 10.1073/pnas.1014418108.

L. S. Tsimring. Noise in biology. *Reports on Progress in Physics*, 77(2), 2 2014. ISSN 00344885. doi: 10.1088/0034-4885/77/2/026601.

A. Verdugo and R. Rand. Hopf bifurcation in a DDE model of gene expression. *Communications in Nonlinear Science and Numerical Simulation*, 13(2):235–242, 3 2008. ISSN 10075704. doi: 10.1016/j.cnsns.2006.05.001.

M. Watanabe, Y. Yasuoka, S. Mawaribuchi, A. Kuretani, M. Ito, M. Kondo, H. Ochi, H. Ogino, A. Fukui, M. Taira, and T. Kinoshita. Conservatism and variability of gene expression profiles among homeologous transcription factors in Xenopus laevis. *Developmental Biology*, 426(2):301–324, 6 2017. ISSN 1095564X. doi: 10.1016/j.ydbio.2016.09.017.

A. B. Webb, M. Lengyel, D. J. Jö Rg, G. Valentin, F. Jü Licher, L. G. Morelli, and A. C. Oates. Persistence, period and precision of autonomous cellular oscillators from the zebrafish segmentation clock. *eLife*, 2016. doi: 10.7554/eLife.08438.001.

G. A. Wyngaard, E. M. Rasch, N. M. Manning, K. Gasser, and R. Domangue. The relationship between genome size, development rate, and body size in copepods. *Hydrobiologia*, 532:123–137, 2005.

S. Xie, M. Swaffer, and J. M. Skotheim. Eukaryotic Cell Size Control and Its Relation to Biosyn-

thesis and Senescence. *Annual Review of Cell and Development Biology*, 38:291–319, 2022. doi: 10.1146/annurev-cellbio-120219. URL https://doi.org/10.1146/annurev-cellbio-120219-.

L. Y. Yampolsky and S. M. Scheiner. Developmental noise, phenotypic plasticity, and allozyme heterozygosity in Daphnia. *Evolution*, 48(5):1715–1722, 1994. URL https://academic.oup.com/evolut/article/48/5/1715/6870049.

M. C. Zody, M. Garber, D. J. Adams, T. Sharpe, J. Harrow, J. R. Lupski, C. Nicholson, S. M. Searle, L. Wilming, S. K. Young, A. Abouelleil, N. R. Allen, W. Bi, T. Bloom, M. L. Borowsky, B. E. Bugalter, J. Butler, J. L. Chang, C. K. Chen, A. Cook, B. Corum, C. A. Cuomo, P. J. De Jong, D. DeCaprio, K. Dewar, M. FitzGerald, J. Gilbert, R. Gibson, S. Gnerre, S. Goldstein, D. V. Grafham, R. Grocock, N. Hafez, D. S. Hagopian, E. Hart, C. H. Norman, S. Humphray, D. B. Jaffe, M. Jones, M. Kamal, V. K. Khodiyar, K. LaButti, G. Laird, J. Lehoczky, X. Liu, T. Lokyitsang, J. Loveland, A. Lui, P. Macdonald, J. E. Major, L. Matthews, E. Mauceli, S. A. McCarroll, A. H. Mihalev, J. Mudge, C. Nguyen, R. Nicol, S. B. O'Leary, K. Osoegawa, D. C. Schwartz, C. Shaw-Smith, P. Stankiewicz, C. Steward, D. Swarbreck, V. Venkataraman, C. A. Whittaker, X. Yang, A. R. Zimmer, A. Bradley, T. Hubbard, B. W. Birren, J. Rogers, E. S. Lander, and C. Nusbaum. DNA sequence of human chromosome 17 and analysis of rearrangement in the human lineage. *Nature*, 440(7087):1045–1049, 4 2006. ISSN 14764687. doi: 10.1038/nature04689.

Appendix: Supplemental Material

Supplement 1: Differences between periods of mRNA and protein expression are minimal

Table S1: Resulting periods of mRNA and protein expression given parameter values corresponding to *X. laevis* and *A. mexicanum* Brownian Motion models

Parameter set	**Resulting period of mRNA expression** (minutes)	**Resulting period of protein expression** (minutes)
$a = 4.5$, $k = 33$, $p_{crit} = 161$, $h_m = 3$, $h_p = 3$, Total delay = 13.17 (*X. laevis* BM model)	43.03	43.04
$a = 4.5$, $k = 33$, $p_{crit} = 420$, $h_m = 3$, $h_p = 3$, Total delay = 33.91 (*A. mexicanum* BM model)	84.97	85.03

Supplement 2: Analytical conditions for the emergence of oscillations

Following the derivation given in Verdugo and Rand [2008], we derive formulae that give analytical conditions for the emergence of oscillations for the system of delayed differential equations (DDE) modelling the segmentation clock as proposed by Lewis [2003].

We begin with Lewis' DDE system, using $\dot{p}$ and $\dot{m}$ derivative notation in place of $\frac{dp}{dt}$ and $\frac{dm}{dt}$, respectively, for convenience:

$$\dot{p} = am(t - T_p) - bp(t) \tag{4}$$

$$\dot{m} = \frac{k}{1 + (\frac{p(t-T_m)}{p_{crit}})^2} - cm(t) \tag{5}$$

and we start by rescaling the system.

Rescaling We rescale, following Verdugo and Rand, and define the following variables:

$$x = \frac{m}{k}, \quad y = \frac{p}{ka}, \quad \Upsilon_0 = \frac{p_{crit}}{ka} \tag{6}$$

We also use the following notation: $m_d = m(t - T_p)$ and $p_d = p(t - T_m)$, and we define x_d and y_d similarly. From the variables defined in Equation (3), we have:

$$\dot{x} = \frac{1}{1 + (y_d/\Upsilon_0)^2} - cx, \tag{7}$$

$$\dot{y} = x_d - by \tag{8}$$

Steady state solution At the steady state (x^*, y^*) we have $x_d = x^*$ and $y_d = y^*$. Setting $\dot{x} = 0$ and $\dot{y} = 0$ yields the following steady state equations:

$$\frac{1}{1 + (y^*/\Upsilon_0^2)} - cx^* = 0 \tag{9}$$

$$x^* - by^* = 0 \tag{10}$$

Eqn. (7) implies $x^* = by^*$. Upon substituting in Eqn. (8) we get the following cubic equation for y^*:

$$(y^*)^3 + \Upsilon_0^2(y^*) - \Upsilon_0^2/cb = 0 \tag{11}$$

This equation can be solved by Mathematica. It has one real root given by,

$$y^* = -\frac{2^{1/3}Y_0^2}{D} + \frac{D}{3 * 2^{1/3}} \tag{12}$$

where,

$$D = \left[\frac{27Y_0^2}{bc} + 3\sqrt{\frac{81Y_0^4}{b^2c^2} + 12Y_0^6}\right]^{1/3} \tag{13}$$

Linearize about a fixed point To linearize about a fixed point we define the following deviations from the steady state (x^*, y^*),

$$\xi = x(t) - x^*, \quad \eta = y(t) - y^* \tag{14}$$

and we use the subscript d to signify the lagged variable such that $\xi_d = \xi(t - T_m)$ and $\eta_d = \eta(t - T_p)$. Note also that $\xi_d = x(t - T_m) - x^*$ and $\eta_d = y(t - T_p) - y^*$.

It follows that,

$$\dot{\xi} = \dot{x} = \frac{1}{1 + \frac{(\eta_d + y^*)^2}{Y_0^2}} - c(\xi + x^*) \tag{15}$$

and,

$$\dot{\eta} = \dot{y} = \xi_d + b\eta \tag{16}$$

We now expand Eqn. 12 for small η_d. To linear order we get,

$$\dot{\xi} = -c\xi - cx^* + \frac{Y_0^2}{(y^*)^2 + Y_0^2} - \frac{2y^*(Y_0)^2}{((y^*)^2 + Y_0^2)^2} * \eta_d \tag{17}$$

Note that the second and third terms on the right hand side of Eqn. (14) sum to zero.

In order to compare the results in Eqn. (14) to Verdogu and Rand, we rewrite the coefficient

of η_d in terms of $\beta = y^*/Y_0$ and define:

$$K = \frac{2y^*(Y_0)^2}{((y^*)^2 + Y_0^2)^2} = \frac{2\beta^2}{y^*(1+\beta^2)^2} \tag{18}$$

We therefore end up with the following linearized equations:

$$\dot{\xi} = -x\xi - K\eta_d \tag{19}$$

$$\dot{\eta} = \xi_d - b\eta \tag{20}$$

which are analogous to those presented in Verdugo and Rand

Oscillatory solution In a supercritical Hopf bifurication, a stable fixed point becomes unstable and is surrounded by a stable limit cycle. Close to the bifurcation, the amplitude of the limit cycle is very small and can be approximated by cosine functions. Therefore, we assume that Eqns. 16 and 17 have solutions given by,

$$\xi(t) = Bcos(\omega t + \phi), \xi_d = Bcos(\omega(t - T_p) + \phi) \tag{21}$$

$$\eta(t) = Acos(\omega t), \eta_d = Acos(\omega(t - T_m)) \tag{22}$$

We can now substitute Eqns. (18) and (19) into the linearized equations, Eqns. (16) and (17), above to solve for ω in terms of model parameters. The condition for oscillation is thus that ω is a real, positive, non-zero number. After some tedious trigonometry and algebra, we find that this condition is satisfied when $K > bc$. In other words the geometric mean of the degradation

constants has an upper bound given by the following.

$$\sqrt{bc} < \sqrt{K} \tag{23}$$

The frequency of oscillation ω is given by:

$$\omega = \left[\frac{-(c^2 + b^2) + \sqrt{(c^2 + b^2)^2 + 4K^2 - 4b^2c^2}}{2} \right]^{1/2} \tag{24}$$

We may also ask if there exists a minimum total delay, $T = T_p + T_m$, required for the emergence of oscillations. Solving for such a T, we get:

$$T = \frac{1}{\omega} ArcSin \left[\frac{\omega(c + b)}{K} \right] = T_{crit} \tag{25}$$

and we define this T to be a critical total delay, T_{crit}, because it is the time delay required for an oscillatory solution just at the bifurcation. It follows that a second condition for oscillation can be defined, that is:

$$T_m + T_p > T_{crit} \tag{26}$$

Note that even when the conditions given in (20) and (23) are met, a finite cut-off in numerical simulations may lead a solution to be classified as non-oscillatory.

In figures S1 and S2, we plot h_p against T_{crit} for every model shown in figures 1 and 2 in the main text, respectively (note that the axes are swapped relative to their corresponding plots in the main text, to clearly demonstrate the positive relationship between protein half-life and critical total delay). T_{crit} is dependent on protein and mRNA degradation rates, b and c, respectively, as well as K. K is dependent on β and y^*, both of which are dependent on Y_0 which is defined as $Y_0 = p_{crit}/(ka)$. As a result, we can say that T_{crit} is dependent on only three parameters from

the Lewis model: b, c, and p_{crit}. Therefore, we need only generate multiple species-specific plots when h_m is changing between models (recall that $b = ln(2)/h_p$ and $c = ln(2)/h_m$) because the range of h_p and value of p_{crit} will remain the same across all species-specific models, regardless of diffusion type. This is why there are only two plots in figure S1 (corresponding to species- and diffusion-specific for which mRNA half-life is held constant at $h_m = 3$), but there are six plots in figure S2 (corresponding to diffusion-specific *A. mexicanum* models for which h_m is adjusted to different scaling relationships with diffusion-specific t_{exp} values).

There is a clear and consistent trend across all plots in both figures. That is, as protein half-life increases, the total delay time required for oscillations to emerge, T_{crit}, also increases. Analytical analysis therefore confirms that total delay time must increase with protein stability to ensure the emergence of oscillations

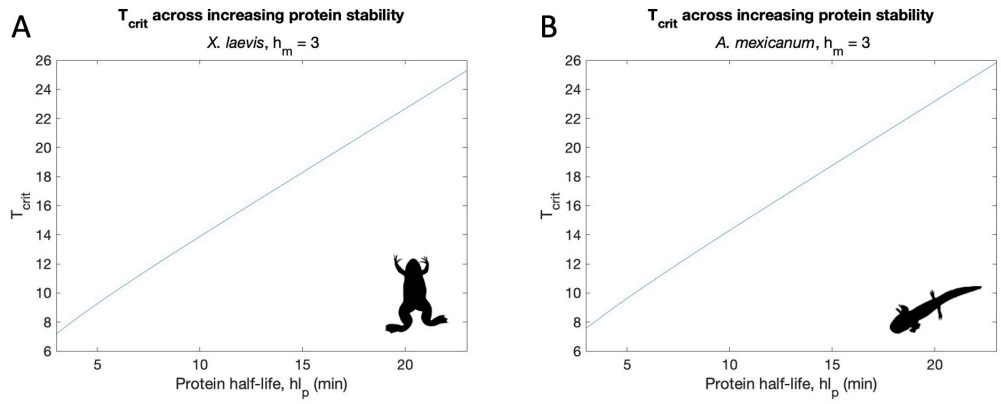

Figure S1: T_{crit} plotted across increasing protein stability corresponding to a range of half-lives between 3 and 23 minutes. A *X. laevis* BM and fBM models for which $h_m = 3$, as shown in figures 1A and 1B. B *A. mexicanum* BM and fBM models for which $h_m = 3$, as shown in figures 1C and 1D.

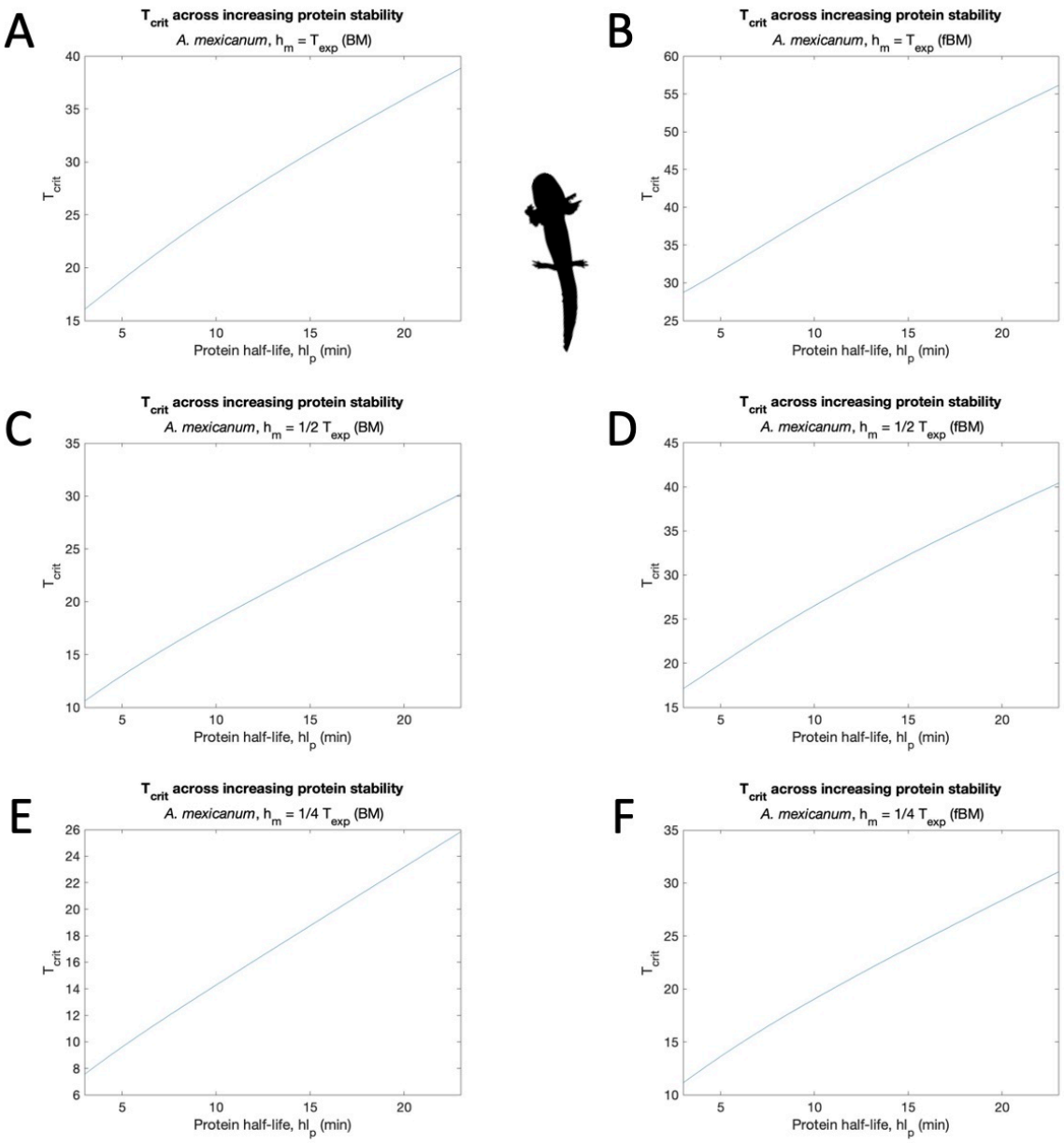

Figure S2: T_{crit} plotted across increasing protein stability corresponding to a range of half-lives between 3 and 23 minutes. A, C, E *A. mexicanum* BM models with $h_m = T_{exp}$, $h_m = \frac{1}{2}T_{exp}$, and $h_m = \frac{1}{4}T_{exp}$, respectively. B, D, F *A. mexicanum* fBM models with $h_m = T_{exp}$, $h_m = \frac{1}{2}T_{exp}$, and $h_m = \frac{1}{4}T_{exp}$, respectively. All panels correspond with their counterparts from Figure 2.

Supplement 3: Extended nuclear export simulation results

Extending nuclear export simulations across a larger range of radii reveals that the impact of obstructed diffusion is pronounced in larger nuclei, but it is reduced by drawing initial positions from a uniform distribution.

In figure S3A we show nuclear export simulation results across a range of interest that captures estimates for nuclear radii of presomitic mesoderm tissue cells in *X. laevis* and *A.mexicanum*, and in figure S3B we show results across a range of ∼0.5 to 13 μm, based on the minimum and maximum nuclear volumes reported in the dataset used by Malerba and Marshall [2021] (https://doi.org/10.5061/dryad.vq83bk3ss) while assuming a spherical volume $V = \frac{4}{3}\pi r^3$. It is important to note that while the results in figure S3B give us an idea of how nuclear export delay increases across nuclear radii for different types of diffusion and chromatin addresses, these estimations are still scaled based on observed nuclear export time in zebrafish (Hoyle and Ish-Horowicz 2013), so we cannot make definitive conclusions about nuclear export times for all transcripts in all nuclei of a certain size.

In figure S4, we plot the distribution of nuclear export times for all four diffusion models shown in figure S3 (BM with initial position at the origin, BM with initial positions drawn from a uniform distribution, fBM with initial position at the origin, and fBM with initial positions drawn from a uniform distribution), and for three different nuclear radii (3.5, 6, and 13 μm; most species have nuclear radii well below 6 μm). We can see that nuclear export distributions are skewed more towards the left (i.e. towards smaller times) when initial positions are drawn from a uniform distribution relative to when initial position is always at the origin or nuclear center. This pattern holds across nuclear size.

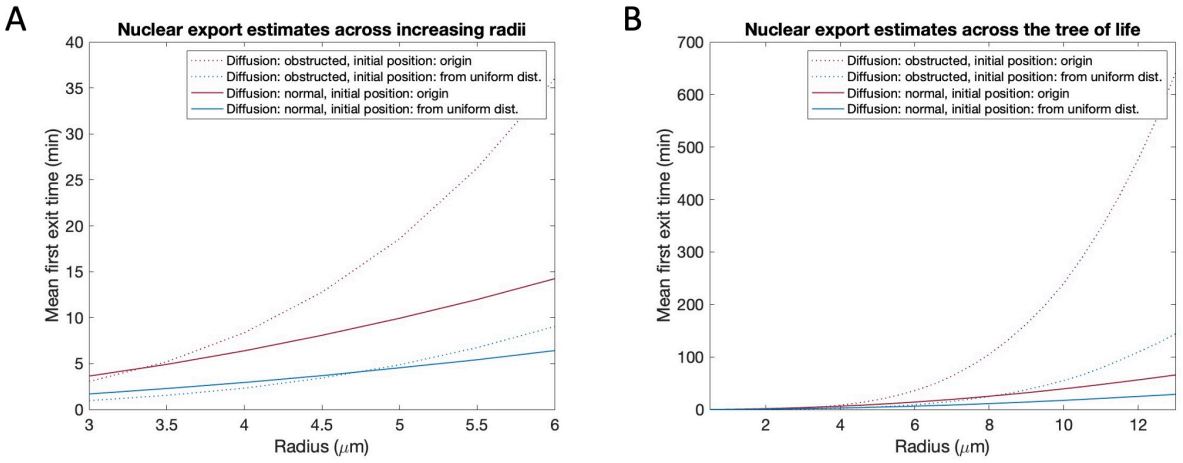

Figure S3: Nuclear export simulation results. A Mean nuclear export time across nuclear radii in *D. rerio*, *X. laevis*, and *A. mexicanum* PSM cells. B Mean nuclear export time across a wider range of nuclei that reflect what has been observed across the tree of life. Simulation results are shown for Brownian Motion(BM)/normal diffusion (solid lines) and fractional Brownian Motion(fBM)/obstructed diffusion (dashed lines), and for initial positions at the origin (shown in maroon) and drawn from a uniform distribution (shown in blue).

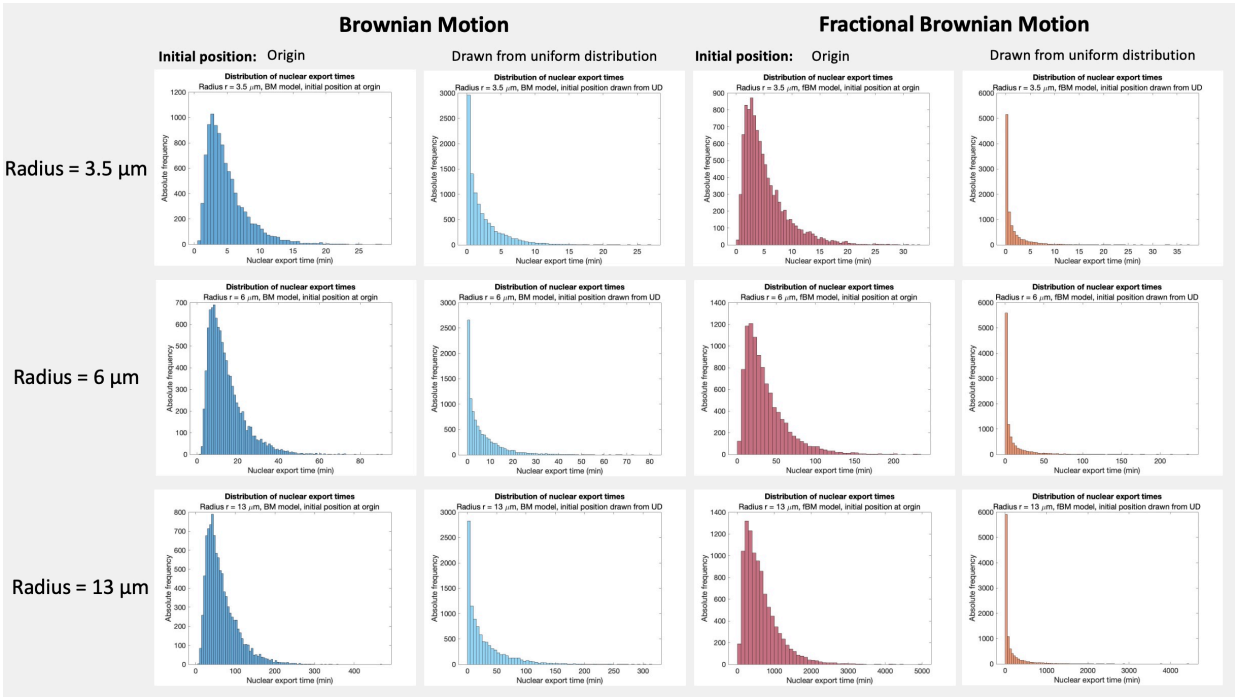

Figure S4: Nuclear export distributions for nuclei with radius 3.5, 6, and 13 μm, and across different diffusion and initial position models.

Supplement 4: Increasing gene product stability for *A. mexicanum* Brownian Motion models

Figure 2A corresponds to the Brownian Motion *A. mexicanum* model that is closest to achieving a period of gene expression to match the species-specific segmentation rate of 155 minutes. For this model, we have that $h_m = T_{exp} = 11.97$. Looking at Figure 2A, we can imagine that increasing the range of protein stability might yield a period of gene expression slow enough to match the known rate of somite segmentation in *A. mexicanum*, and that the subset of parameters to yield a period of 155 minutes would correspond to protein stability that is relatively high compared to mRNA stability. Furthermore, we would have a scenario in which only 50% of mRNA transcripts are expected to degrade before leaving the nucleus, $h_m = T_{exp}$. To test this, we take the model shown in Figure 2A, and we consider a new range of protein stability from 15 to 35 minutes (as opposed to from 3 to 23 minutes). All other parameters are held constant. As shown in figure S5A, solely increasing protein stability while holding all other parameters does not yield the 155 minute rate of somite segmentation.

Increasing mRNA stability by 25%, $h_m = 14.96$, while holding protein stability at its higher range, yields one parameter combination with a period of $\sim$154 minutes (figure S5B). Increasing mRNA stability by 50% relative to $h_m = 17.96$, yields only one parameter combination for which oscillations emerge, and its period is $\sim$153 minutes (figure S5C). We can only begin to recapitulate the *A. mexicanum* segmentation clock (i.e. we are within 2 minutes of the known segmentation rate) under a Brownian Motion/normal diffusion model if both mRNA and protein stability are increased. However, we still fail to fully recapitulate a period of 155 minutes. Without increases in the total delay time, $T_m + T_p$, there is an upper limit on how stable gene products can become before the emergence of oscillations is no longer possible; instead of slowing the segmentation

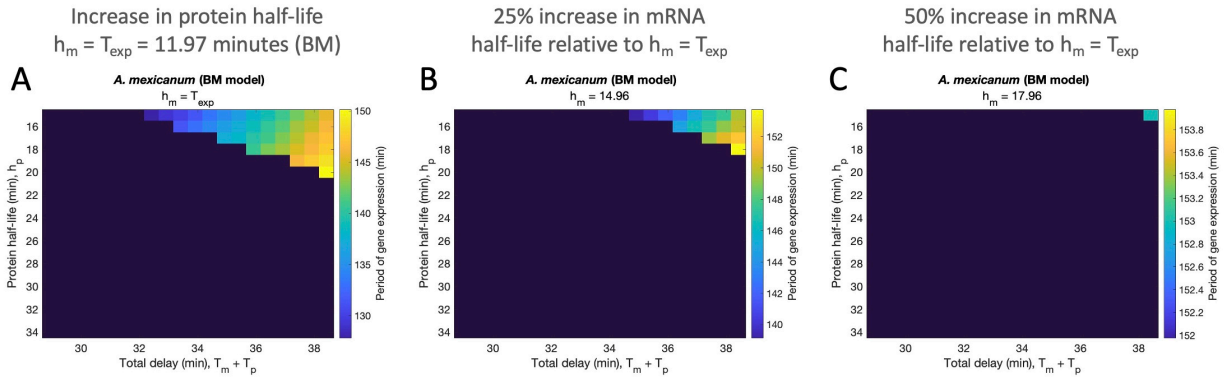

Figure S5: Further increasing gene product stability in the *A. mexicanum* BM model. A We keep mRNA half-life equal to the species- and normal diffusion-specific estimated nuclear export time, and we set protein stability to a higher range corresponding to half-lives of 15 to 35 minutes. B We keep this high range of protein stability constant and we increase mRNA stability by 25% relative to $h_m = T_{exp}$. C We increase mRNA stability by 50% relative to $h_m = T_{exp}$ while holding the high range of protein stability constant.

clock down, we "break" it altogether by damping oscillations. Additionally, results in figure S5B and S5C suggest that to achieve a period close to 155 minutes protein molecules are either only slightly more stable than their transcripts (S5B), or less stable than their transcripts (S5C). This is an issue discussed in the main text.

Supplement 5: Amplitude plots

In figures S6 and S7, we plot the amplitudes of mRNA and protein expression, respectively, corresponding to the models plotted in Figure 2, for which mRNA stability is set to scale with estimated export time, $h_m = T_{exp}$. In figure S6, we can see that when mRNA stability is decreased from $h_m = T_{exp}$ by 50% to $h_m = \frac{1}{2}T_{exp}$, we see an overall increase in the amplitude of mRNA expression, yet when mRNA stability is further decreased to $h_m = \frac{1}{4}T_{exp}$ there is an overall decrease in amplitude. This general pattern is seen across both models (normal and fractional Brownian Motion), and in figure S7, we see this pattern extend to protein as well. The results shown in figures S6 and S7 are counter intuitive. First, we might intuit that higher rates of mRNA degradation (i.e. lower levels of stability) would result in low transcript numbers, but

we see the opposite when mRNA stability decreases from $h_m = T_{exp}$ to $h_m = \frac{1}{2}T_{exp}$. One might reason that increased transcript degradation would lead to less translation resulting in lower protein numbers and therefore less transcriptional repression, and that this would explain an increase in transcript numbers following a decrease in mRNA stability. However, we also have increasing protein amplitude when mRNA stability decreases from $h_m = T_{exp}$ to $h_m = \frac{1}{2}T_{exp}$, so this explanation does not hold. Furthermore, the increase in both mRNA and protein amplitude is followed by a decrease when mRNA stability is further decreased from $h_m = \frac{1}{2}T_{exp}$ to $h_m = \frac{1}{4}T_{exp}$. Taken together, we have that there is a complex relationship between mRNA stability and the amplitude of mRNA and protein expression.

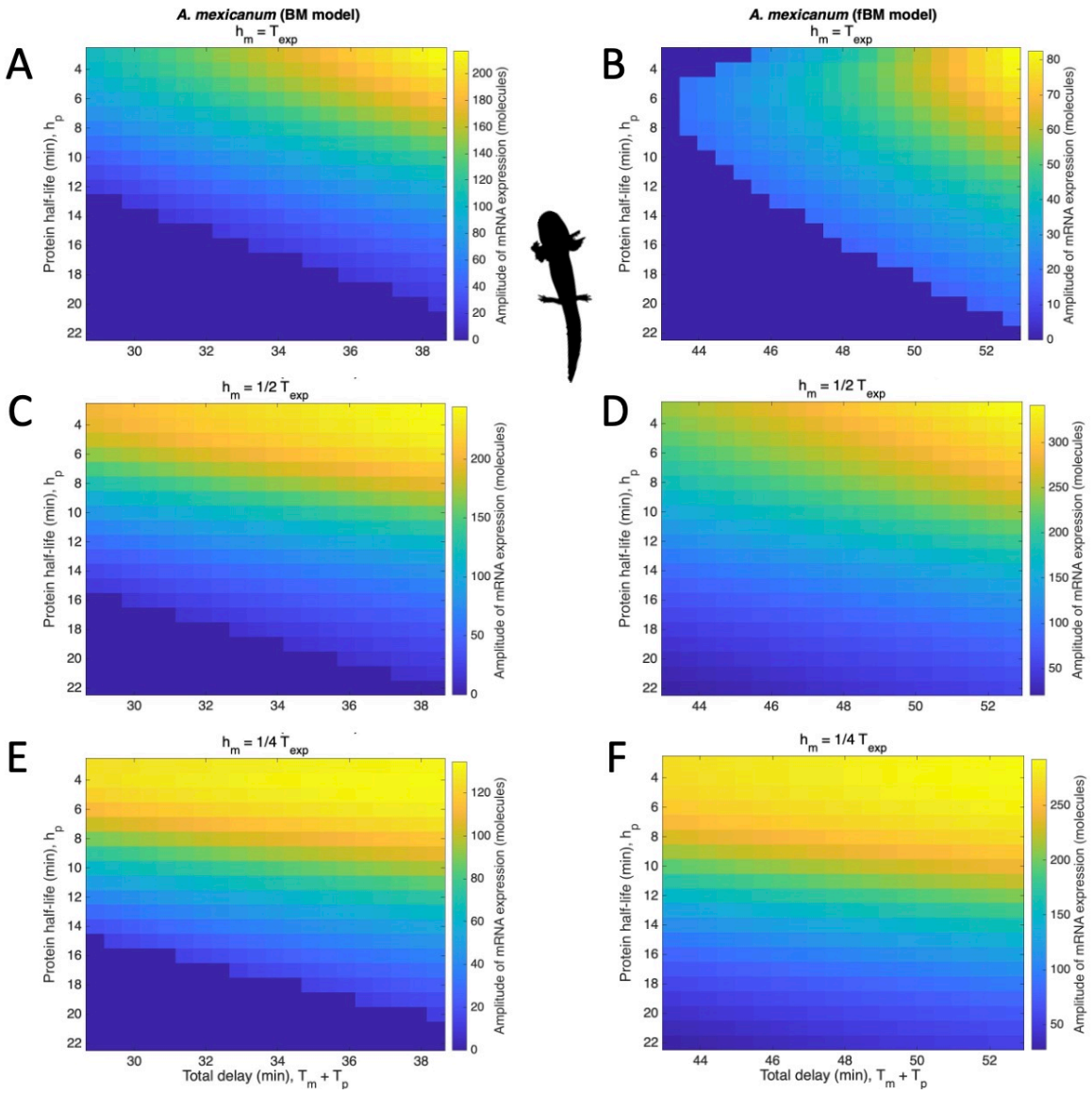

Figure S6: Resulting amplitudes of mRNA expression for *A. mexicanum* models corresponding to plots in Figure 2. A, C, E normal diffusion/Brownian Motion model results; C, D, F obstructed diffusion/fractional Brownian Motion model results. mRNA half-life is held constant at: A, B diffusion-specific estimates for mRNA export delay; C, D half of estimated mRNA export delays; E, F a quarter of estimated mRNA export delays.

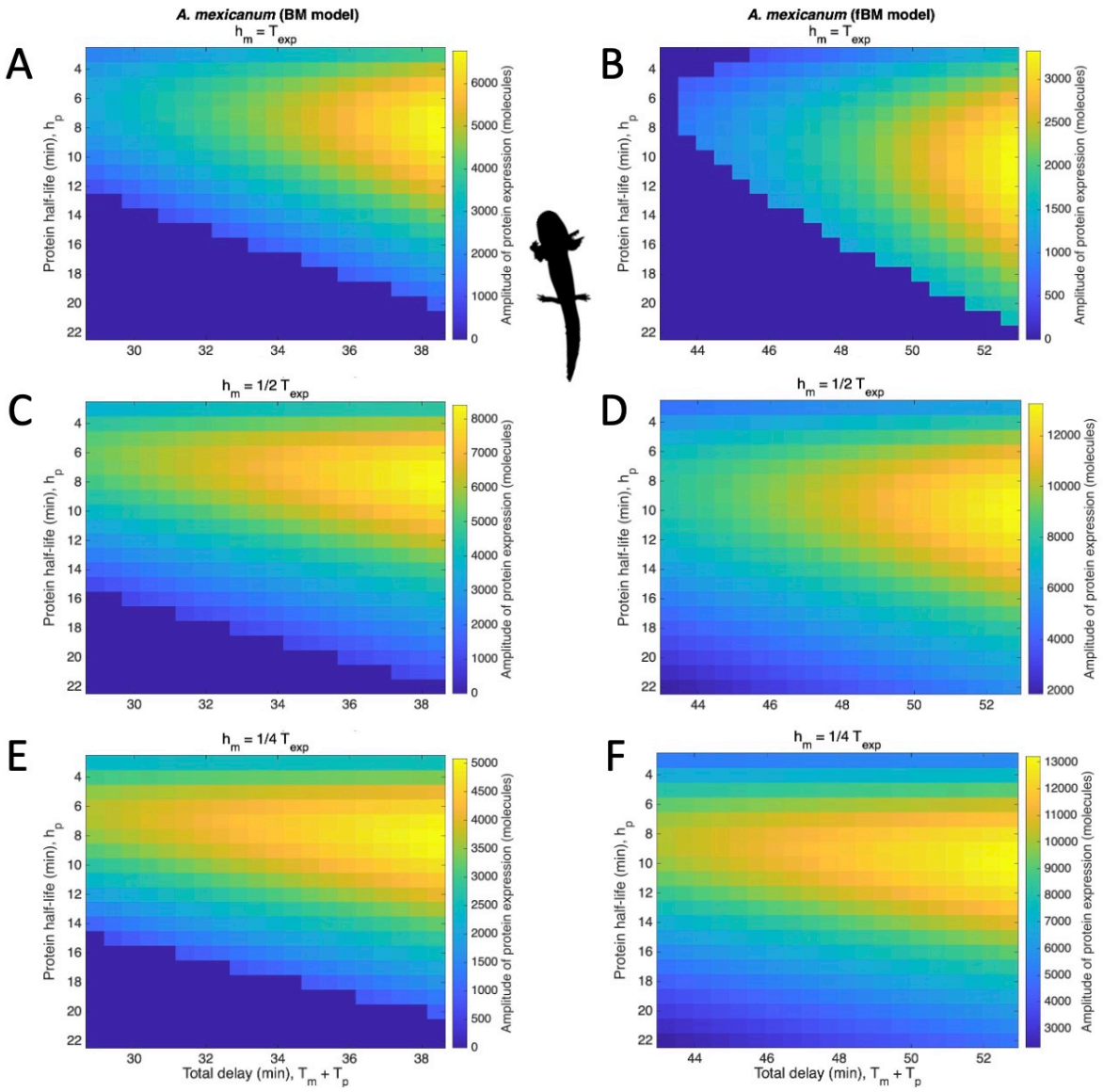

Figure S7: Resulting amplitudes of protein expression for *A. mexicanum* models corresponding to plots in Figure 2. A, C, E normal diffusion/Brownian Motion model results; C, D, F obstructed diffusion/fractional Brownian Motion model results. mRNA half-life is held constant at: A, B diffusion-specific estimates for mRNA export delay; C, D half of estimated mRNA export delays; E, F a quarter of estimated mRNA export delays.