Modeling Macronuclear DNA Regulation in Two Ciliates; *Paramecium tetraurelia* and *Tetrahymena thermophila*

Cassandra Elphinstone and Kyle Cheadle
Science One Program
The University of British Columbia
Vancouver, Canada
May 2013

Abstract

A revision to existing models for regulating the macronuclear DNA content of the ciliates *Paramecium tetraurelia* and *Tetrahymena thermophila* explains previously unresolved observations. Using an independent parameter to regulate ciliate macronuclear DNA content allows the mass of *P. tetraurelia* to be linked with DNA regulation. A similar parameterization of the *T. thermophila* model accounts for observed generation-dependent variations. Introducing controlled selection rules on macronuclear DNA content in modeled populations of *T. thermophila* results in evolving periodic distributions. The amount of unequal macronuclear division in the population is then directly proportional to the frequency of the resulting oscillation. Unequal division acts to restore the distribution opposing the selection displacement. Another parameter related to the replication's independence from macronucleus DNA content shows a critical value of $\sqrt{2}-1$ such that higher values result in periodic variations while lower values do not.

Introduction

A distinct feature of ciliates, including *Paramecium tetraurelia* and *Tetrahymena thermophila*, is the distribution of their genetic information into a small germinal micronucleus whose function is reproduction and a large (8 to 6575 times as much DNA [1]) somatic macronucleus involved in metabolic and protein synthesis. Because the ciliate macronucleus does not undergo mitotic division, different mechanisms are at work in regulating the genetic content across generations. These regulatory mechanisms are herein investigated for statistical populations of *P. tetraurelia* and *T. thermophila*. The ciliate cell cycle begins first with conjugation, then fission involving replication and division. Models for replication in *P. tetraurelia* and *T. thermophila* have previously been described by Berger and Schmidt [2] and Doerder and DeBault [3] respectively. Since the ciliate macronucleus does not contain cellular machinery to precisely govern division this is a major source of variance and regulatory mechanisms have developed to compensate.

Previous models [2,3] for ciliate macronuclei DNA regulation involving the input of mean DNA content across a population as a key parameter have successfully described cross-generational trends in the variance of this DNA content. Conceptually, it is difficult to reconcile how a global parameter such as average macronuclear DNA content can directly affect DNA regulation within an individual cell. How can a dynamic global mean be interpreted by a single cell in a single generation? Berger [4] suggested that the regulatory system may involve factors linked to cellular size. The involvement of cell size implies a relationship with the amount of resources the cell can devote to protein synthesis; there is no reason for a cell to possess DNA that transcribes more RNA than can be translated into proteins by cell machinery [5]. To address these concerns,

our model introduces simple changes to the two models for *P. tetraurelia* and *T. thermophila* to eliminate the global generation-dependent parameters that have previously been employed.

Effective Mean Model for DNA regulation

We modified the additive (model 1) and multiplicative (model 2) regulatory models of [2] in order to remove the nonlocal aspects of these models. Appendix A outlines the computer code used. Essentially, replication for *P. tetraurelia* (model 1) involves the addition of a constant DNA amount to arrive at the ith cell's postreplication (prefission) value (x_i'). In what follows, an x without subscripts represents the current generation's DNA content and an x_i represents the DNA content for generation i. While a population average was added at every generation in the original model 1, we instead added a weighted mean between an independent, external to the macronucleus, parameter (A) and the prereplication DNA content of the individual cell itself (x_i). The postreplication (prefission) DNA content of each cell is given by:

$$x' = x + \mu_1 \quad \text{where} \quad \mu_1 = \frac{(A+f\cdot x)}{1+f}. \tag{1}$$

f is a weighting factor biasing the effective average, μ_1, towards A ($f = 0$) or towards x for $f \to \infty$. A is normally chosen to be the same as the initial mean of the population, $\hat{x}$. For most cases, A and $\hat{x}$ are normalized to a value of 100.

In *T. thermophila* (model 2), the model is multiplicative and involves a postreplication increase of 1, 2 or 4 times the prereplication DNA content (x_i) dependent on thresholds consistent with known characteristics of *T. thermophila* [6], [7]. We modified the thresholds for replication which were previously based on a global population averages to again involve two components, an independent parameter, A, and the cell's individual prereplication content, x_i. The postreplication DNA content, x', the takes the form:

$$x' = x\cdot(b+1)\begin{cases} x < t_l & : b = 3 \\ t_l < x < t_u & : b = 1 \\ x > t_u & : b = 0. \end{cases}, \quad t_l = \mu_2\sqrt{2}, \; t_u = \frac{\mu_2}{\sqrt{2}}, \text{ where } \mu_2 = \frac{(h\cdot A + x)}{1+h}. \tag{2}$$

t_l and t_u are the lower and upper thresholds respectively and μ_2 is the effective mean between A and x. This effective mean varies between A (for $h\to\infty$) and x (for h=0). These effective mean values replace the average values used in previous models. Appendix B gives an explanation for why the threshold constant $\sqrt{2}$ minimizes the variance of the DNA content. The initial parameters for the model are the cell population N, the population's mean DNA content $\hat{x}$ and its s.d. σ (We shall denote a general standard deviation as s.d. in what follows). Appendix C outlines a possible physical interpretation for the parameter f and how, under circumstances associated with *T. thermophila*, the first additive replication model can give rise to the second threshold multiplicative replication model.

Macronuclei unequal division algorithm:

Table 1 *P. tetraurelia*: $\hat{d}$, θ ($\hat{d}$ from [2])

genotype	$\hat{d}$ (%)	θ (%)
Wild type	5	3.1
Am/am	8.2	5.1
tam A/tam A	31.1	19.5

Division algorithms were implemented such that macronucleus division could be either dependent or independent of macronuclei size. Division of the postreplication macronuclei (x' given by Eq. 1 and 2 above) results in the large and

small post-division daughter cells, L_i and S_i.

$L_i = \frac{x_i'}{2} + \gamma \frac{di}{2}$, $S_i = \frac{x_i'}{2} - \gamma \frac{di}{2}$, where $\gamma = \frac{(h_d \cdot A + x_i')}{1+h_d}$. (3)

d_i represents the fractional (relative to γ) difference between sisters and h_d is a weighting factor biasing the effective average towards x_i' ($h_d = 0$) or towards A ($h_d = \infty$). A is usually set to be the same as the initial normalized population mean $\hat{x} = 100$.

If d_i is normally distributed having a mean of 0 and s.d. of θ, then θ is related to the average fractional difference ($\hat{d}$) between sisters by $\hat{d} \cdot \sqrt{\frac{\pi}{2}}$ (See Appendix D for further information).

Observations of the mean inequality of macronuclear division expressed as a percentage of prefission DNA by Berger and Schmidt [2] ($\widehat{d}$) can then be used to derive θ as input to our model (see Table 1 and Appendix D). For any h_d from 0 to ∞, the resulting probability distributions of d_i versus x_i' are difficult to distinguish from each other for low populations and/or low σ (see Appendix D).

Model 1 Results and Discussion

The ratio of the coefficient of variation ($cv = 100 \times standard\ deviation/mean$) between the postreplication macronuclear DNA content and the prereplication content in a population provides important information about the ciliate regulatory mechanisms [2]. The coefficient of variation between the postreplication and prereplication DNA content for model 1, when $A = \hat{x}$, can be analytically calculated to be:

$$\frac{cv_{post}}{cv_{pre}} = \frac{1}{2} s \quad \text{where } s = \frac{2f+1}{1+f} \text{ (see Appendix C and E for more information).} \quad (4)$$

This generalizes from equivalent ratios ($\frac{1}{2}$and 1) found by Berger and Schmidt [2] for our model's limiting cases of $f = 0$ and $f = \infty$. Berger and Schmidt [2] also show that postreplication to prereplication coefficients of variation for *P. tetraurelia are* lowered by a ½ or more so that allowable values for *f* must be around 0. As the postreplication (G2) macronucleus divides into G1 sisters, there is an increased variance introduced by the unequal division of the macronucleus and the ratio of $\frac{\sigma_{G2}}{\sigma_{G1}}$ will evolve towards the equilibrium values. For $f = 0$ the ratio for a steady equilibrium is 1 and for $f \to \infty$ it is 2 similar to the results expected for model 1 and 2 in [2].

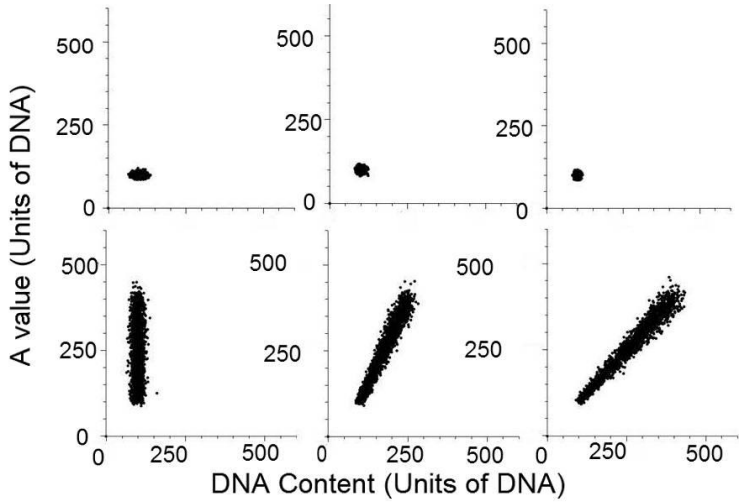

Figure 1. a. (top) Normal distribution of A values (no particular cell line has different A). Generations are 0 (left),1 (middle),30 (right) Model: f=0, h_d=0, mean A= 100 with fractional s.d.=0.05, initial $\hat{x} = 100$, σ =10.1, θ =0.031, N= 2000. **b.(bottom)** Changing A with cell line i (from 100 to 400), fractional s.d. =0.05.

We investigated how the independent parameter could be used to regulate DNA content. A normal distribution of A values (Mean of 100 and fractional s.d. of 0.22) was introduced throughout the population and the relationship between A and the DNA content for three generations 0, 3 and 34 is shown in Figure 1a. It is clear that no relationship is present. If instead we allow each cell line in our population to be associated with a different mean value of A (see Figure 1b) with the same fractional s.d. as before, then what initially (for generation 0) was a random relationship between A and the cell's DNA content changes into a strongly dependent one by generation 34. A was varied by a factor of 4 (from a normalized value of 100) to cover the range observed in *P. tetraurelia* cell mass [4]. The independent parameter A may now be understood as a parameter external to the DNA content of the macronuclei that nevertheless strongly regulates the DNA content generational evolution.

These general results showing a well-defined relationship between A and the DNA content are independent of our particular model parameter choices (excepting case ($f = \infty$) although the details of the relationship and generational development will vary. Our choice for A allowed us to reproduce the relationship found by Berger [4] between mean cellular mass and mean macronuclear content for separate cultures of *P. tetraurelia.* These results indicate that logically a link between cellular mass/size and the independent parameter A in the model could be made. The regulatory system of *P. tetraurelia* is capable of expandable regulation; it has the mechanisms necessary to regulate across greater disparities than normally found [4] and it may be that this regulatory mechanism developed in response to a need to regulate content that varied as a result of cellular size or mass. Berger [5] selected a population of *P. tetraurelia* with a mutation (tam/tam in Table 1) with extremely uneven macronuclear division and it was found that those cells with minimal gene dosages were limited in their rate of protein synthesis by the amount of macronuclear DNA present, while protein synthesis in cells receiving the greatest amount of DNA was limited above by overall cell mass. Thus, the lower limit is determined by the macronucleus and its capacity to transcribe RNA, while the upper limit is determined by the size of the cell and its associated amount of assembly available for synthesizing protein. Parameter A, initially established as a control for the population's "desired" average, could potentially be thought of as a ratio representing a maximized relationship between DNA and cell mass. Future studies using this new model may help clarify the mass/DNA content relationship.

Model 2 Results and Discussion

Although the original model [3] for *T. thermophila* yields observational values for the coefficient of variation and ratios for the $\frac{\sigma_{G2}}{\sigma_{G1}}$ ratio, problems arise in the model when unequal macronuclear division is introduced. As explained in Appendix B, the model works to confine the DNA content distribution to within upper and lower threshold values. With macronuclear division, the resulting distribution "leaks" out beyond the thresholds and since the thresholds are not symmetric about the mean, the average DNA content continuously increases with generation number. The parameter A in this model acts as an independent threshold mechanism that prevents a runaway mean DNA content. Doerder and DeBault [3] have shown in their Table 6,

ratios of $\left(\frac{\sigma_{G2}}{\sigma_{G1}}\right)^2$ indicating that $\frac{\sigma_{G2}}{\sigma_{G1}}$ typically lies between 1.88 and 2. An h_d parameter of 0 (emphasizing x_i) tends to give either a low ratio (1.66 at times) or an increasing coefficient of variation. This appears to be due to the production of a high DNA content tail to the probability distribution due to the unequal division biasing towards higher DNA content. A large h_d value gives reasonable values for both the mean values and the ratio $\frac{\sigma_{G2}}{\sigma_{G1}}$ (e.g from a start mean of 100, $\sigma = 0.13$, with h_d and h large, equilibrium exists with a mean of 102, ratio of 1.9 and G2 coefficient of variation of 18). These results give some initial support to the idea that the unequal division of *T. thermophila* may involve a mechanism independent of gene dosage.

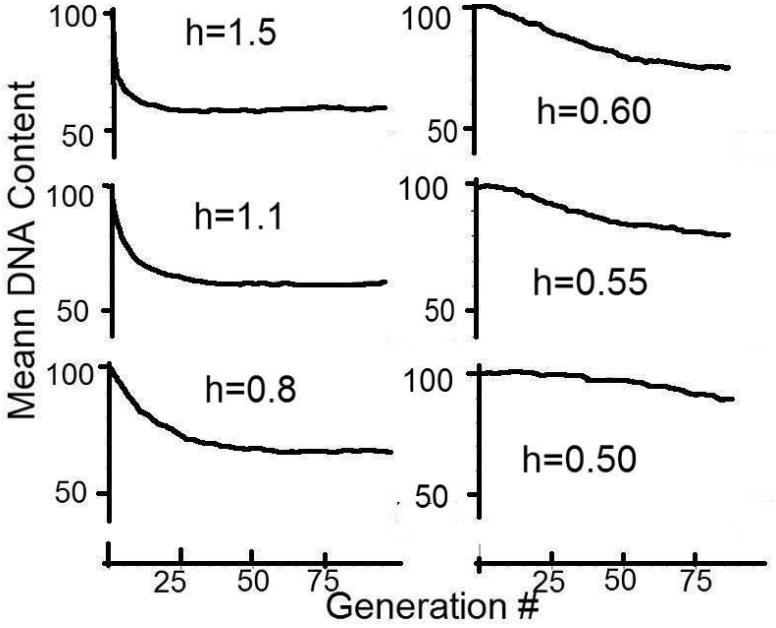

Figure 2. (mean DNA content vs. Generation): DNA content transition to A, varying with parameter h: for model 2; $A = 55$, $h_d = 0$ (division factor), $\theta = 0.13$, $N = 2000$. Initial $\hat{x}$ and σ = 100, 24 respectively.

Using the independent variable A in the thresholds also allows temporal changes to the ciliate's DNA content with the potential ability to regulate the time constant involved in invoking changes to the mean content. In *T. thermophila*, this is particularly important with "young cells" (with relatively few number of fissions since conjugation) and "old cells" (more fissions since conjugation) that have a different number of subunits associated with them. Older cells tend to have a baseline number of G1 subunits of around 45 while younger cells can have up to 164 G2 subunits (~64 G1 equivalent units). One strain maintains high content (130 G2 or 65 G1 subunits) for more than 50 generations while others more quickly (within 50 generations) revert back to a base level of about 90 G2 /45 G1 subunits) [3]. By choosing A to correspond to an "expected" DNA value (e.g. 45C in the case of *T. thermophila*) and beginning with a higher initial value ($\hat{x}$) for the mean DNA content of the population (for example 82), the model will evolve such that the mean population heads toward A. In Figure 2, the initial value is normalized to 100 and changing the parameter h in the effective mean μ_2 regulates how many generations it takes for the mean to shift to the base level.

Regulation occurs rapidly as $h \to \infty$ but can take much more than 50 generations if h falls below about 0.6. These results show that temporal regulation of *T. thermophila* DNA content may in part be controlled by a factor independent of the DNA content (A in this case) and that as that dependence increases so does the DNA's regulatory response time.

This was further supported when, instead of a random sister being chosen to continue the next generation, the larger sister instead was always chosen. The interaction of the thresholds, the unequal division of the macronucleus and the introduced selection rule resulted in a damped travelling wave moving through the probability distribution that had a wavelength (in DNA content) of $t_u - t_l$ and a period governed by how long the disturbing wave took to move through the distribution. As shown in Figure 3 (see also Appendix F), the wave frequency was proportional to the amount of unequal division introduced (θ) and the wave became better defined as N was increased. This result could be used to design experiments that could directly determine the regulation time constants associated with different strains of *T. thermophila*.

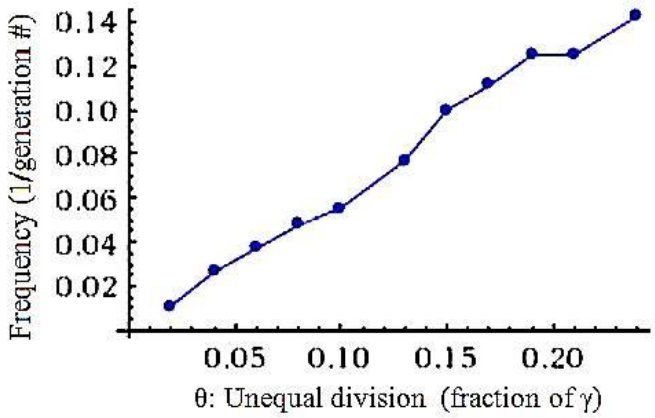

Figure 3. *T. thermophila* simulations undergoing large sister selection show a linear increase in frequency with increasing asymmetric division parameter *d.* $15000 > N > 5000. A = 100$, $sigma = 24$, $h = 50000$ (and see Appendix F).

Concluding Remarks

We have shown the importance of introducing an independent parameter into the regulatory models of both *P. tetraurelia (*model 1) and *T. thermophila* (model 2). In *P. tetraurelia,* the parameter allows a link to be established between external parameter such as the mass or the size of the cell while in *T. thermophila*, it can be used to regulate the temporal evolution of different strains. The model for *T. thermophila* responds under selection rules such that precise time constants for the DNA regulation could be determined given careful experimental design.

Acknowledgements

We would like to thank Dr. James Berger for sharing his extensive knowledge of ciliates, for always being available and open to questions, and for his detailed explanations of Vpython code. We would also like to thank Dr.Wayne Nagata, Linda Au, Jacelyn Shu, Daniel Bajj, Allen Yieh, and Rob Elphinstone for their help revising and improving our many drafts.

References

[1] Allen, S. and Gibson, I. Genome Amplification and Gene Expression in the Ciliate Macronucleus. *Biochemical Genetics* **6**, 293-313 (1972).

[2] Berger, J.D. and Schmidt, H.J. Regulation of Macronuclear DNA Content in Paramecium Tetraurelia. *Journal of Cell Biology* **76**, 116-126 (1978).

[3] Doerder, F.P. and DeBault, L.E. Life Cycle Variation and Regulation of Macronuclear DNA Content in Tetrahymena Thermophila. *Chromosoma* **69**, 1-19 (1978).

[4] Berger, J.D. Regulation of Macronuclear DNA Content in Paramecium tetraurelia. *Journal of Protozoology* **26**, 18-28 (1979).

[5] Berger, J.D. Effects of Gene Dosage on Protein Synthesis Rate in Paramecium Tetraurelia. *Experimental Cell Research* **141(2)**, 261-275 (1982).

[6] Doerder, F.P. Regulation of Macronuclear DNA Content in Tetrahymena Thermophila. *Journal of Protozoology* **26**, 28-35 (1979).

[7] Vorob'ev, V.I., Borschsenius, S.N., Belozerskaya, N.A., Merkulova, N.A., and Irilina, I.S. DNA replication in macronuclei of Tetrahymena pyriformis. *Experimental Cell Research* **93**, 253-260 (1975)

Appendix A: V python code for simulation

The simulation used Vpython 2.7 and was designed so that for a given simulation run two models (one associated with *P. tetraurelia* and the other with *T. thermophila)* could be analyzed together and compared. Model options include the original models of Berger and Schmidt [2], the effective mean models (see main paper) and a volume replicator model as outlined in Appendix C. Details of each run including standard deviations, coefficients of variations and means for various states of the macronucleus were printed. The code allowed for several plotting options including histogram time variations, parameter temporal variations and plots of sister postreplication and prereplication relationships. A "choosesister" flag allowed selection of a particular sister to be chosen in each generation. The code is about 50 pages long and so is not included here but is freely available upon request.

Appendix B: Understanding the threshold parameter $\sqrt{2}$

As shown in Berger and Schmidt [2], choosing $\sqrt{2}$ as the threshold constant in model appears to minimize the variance associated with the probability distribution of the population's DNA content but the reason for this was unknown. We give here a qualitative argument for why that occurs for model 2 from Berger and Schmidt [2] (using $\hat{x}$ instead of μ_2 in Eq. 2 above). It can be understood as the smallest factor that confines a symmetrically dividing population within the thresholds set out by model 2 in this paper. The lower threshold (t_l) requires values below it to increase in post replication by a factor of 4 while the upper threshold (t_u) requires the values above it to remain unchanged. When the postreplication cell divides in 2 during fission, the macronuclei in the central region (2 times replication) return to their previous DNA content. A step towards minimizing the variance of the distribution would be to attempt to confine the divided cell into the central region between the two thresholds. If C is the threshold constant, then for prereplication values just below the lower threshold to lie just below twice the upper threshold (postreplication) in the next generation requires:

$$4 \cdot t_l = 4 \cdot \frac{\hat{x}}{C} \leq 2 \cdot t_u = 2 \cdot C \cdot \hat{x} \ . \quad (1B)$$

So that $C^2 \geq 2$. Similarly for values just above the upper threshold, after replication they must lie just above twice the lower threshold (in postreplication):

$$t_u = C \cdot \hat{x} \ \geq 2 \cdot t_l = 2 \cdot \frac{\hat{x}}{C} \ . \quad (2B)$$

So that again $C^2 \geq 2$. (Note that probability values of a normal distribution just below the lower threshold and just above the upper threshold are the largest within their intervals and so are the most important to consider when trying to minimize the variance). In order to minimize the total variation within the thresholds (making the region $t_u - t_l$ as small as possible) the constant C should be as small as possible so that $C = \sqrt{2}$ in order to minimize the variation of the next generation.

Appendix C: Understanding the effective means μ_1 and μ_2 and a Possible Volume Replicator Model

In the paper, we used the effective mean μ_1(Eq. 1) as a weighted average that would vary between x and A depending on the weight factor f. This f value can be thought of as a ratio between two competing DNA production processes. One is governed by the macronucleus contents x producing at most A DNA units and the other by a certain number of subunits (S) always replicating at a certain rate. If $x = I + S$ then the I replicators will be associated with the production of the $\frac{A \cdot I}{x}$ units and the S replicators each replicate b DNA units. In this case some of the I replicators may not replicate so that A is not exceeded. In order to better understand model 2 where different replication rates are allowed we allow the S replicators to reproduce b subunits each (where $b = 0$, 1, or 3 as in Eq. 2). Then on average each subunit produces:

$$\frac{(I \cdot \frac{A}{x} + Sb)}{x} = \quad \frac{(I \cdot \frac{A}{x} + Sb)}{(I+S)} = \frac{(\frac{A}{x} + \frac{S}{I} b)}{(1+\frac{S}{I})} = \frac{(\frac{A}{x} + fb)}{(1+f)} \text{ where } f = \frac{S}{I}. \quad (1C)$$

And all of the x subunits produce $\frac{(A+fbx)}{(1+f)}$ so that the postreplication DNA content becomes:

$$x' = x + \frac{(A+fbx)}{(1+f)} \quad \text{and} \quad \mu = \frac{(A+fbx)}{1+f} \ , \ f = \frac{S}{I} \ . \quad (2C)$$

If the term bx dominates (i.e. $f \to \infty$) Eq. 2C becomes $x' = x \cdot (b+1)$ similar to the multiplicative Model 2 Eq. 2. In this case there are only S type replicators and the independent A process is unimportant. If $f = 0$ then there are no S type replicators and A units are always added (i.e. additive model 1).

We can then consider a possible scenario where the I and S type replicators represent the interior and surface respectively of the macronucleus itself. The interior replicators could then be associated with producing at most "*A*" DNA units and the S replicators simply replicate at a rate '*b*'. In *T. thermophila,* if the prereplication I subunits could produce in excess of the limit *A* that

the cell could support, that replication rate is too high and so just as in model 2, the replication threshold is reached and the rate reduced. The replication rate must change in order for all subunits in the macronucleus to replicate at the same rate. The rate *b* is then reduced for that generation.

In *T. thermophila* both S and I replicators can only replicate at $1X$ $(b = 0)$, $2X$ $(b = 1)$, or $4X$ $(b = 3)$ but the I type replicators must act like a switch changing the characteristic of the replication from $4X$ to $2X$ and then to $1X$ based on if the entire set of I replicators can produce *A* subunits of DNA content at the given replication rate. If the cell cannot support higher levels of production than *A* subunits, then it must switch to a lower replication rate when this production level is reached.

If the I subunits are spheres of radius r packed within the interior of the assumed spherical *T. thermophila* macronucleus with radius $r(g+1)$, and S subunits are a single layer on the surface of the sphere, then they can be related approximately (ignoring the packing of the spheres) by:

$$I\frac{4\pi}{3}r^3 = \frac{4\pi}{3}(rg)^3, \text{ and } x\cdot\frac{4\pi}{3}r^3 = (S+I)\frac{4\pi}{3}r^3 = \frac{4\pi}{3}(r(g+1))^3. \quad (3C)$$

So that:

$$I = g^3, \text{ and } x = (S+I) = (g+1)^3. \quad (4C)$$

One can calculate the value of *g* associated with the minimum number of I subunits needed to result in *A* extra units of DNA content at $4X$ $(b = 3)$ replication (i.e. $g = \sqrt[3]{\frac{A}{3}}$). x is then given by $x = \left(\sqrt[3]{\frac{A}{3}} + 1\right)^3$. We can then also ask what x will produce *A* units at $2X$ replication $(b = 1)$. This is just $x = A$. If these two conditions are met at the same time, then as the I replicators reduce their replication rate (because otherwise they will produce more than *A* units), the entire macronucleus (*x*) can still produce *A* at the lower replication rate. This condition is given by:

$$A = \left(\sqrt[3]{\frac{A}{3}} + 1\right)^3 \quad \text{or} \quad A = \left(\frac{1}{1-\sqrt[3]{\frac{1}{3}}}\right)^3 \cong 34.7 \ . \quad (5C)$$

If *A* was smaller then when *x* could produce *A* units at $2X$, there would not yet be enough I subunits to produce *A* at $4X$. If *A* was larger then when the I subunits could produce *A* at $4X$, there would not yet be enough *x* subunits to produce *A* at $2X$. A value of $A \cong 34.7$ ensures that if when the I subunits can produce their maximum (A) at $4X$, the macronucleus can switch to $2X$ replication and the whole macronucleus can still produce *A* subunits. A second limit will occur near where the I subunits will produce *A* subunits at $2X$ replication $(g = \sqrt[3]{A}.)$ The macronucleus can then switch to $1X$ replication $(b = 0)$. This condition will occur at:

$$x = \left(\sqrt[3]{A} + 1\right)^3 \cong 77. \quad (6C)$$

These geometric considerations then give natural reasons for thresholds for *T. thermophila.* A simulation using $A = 34.7, \theta = .13, h_d = 50000$ and an upper threshold of 77 results in a stable equilibrium for $\hat{x}$ of about 56 DNA subunits, a G2 coefficient of variation of 18.7 and G2/G1 cv ratio of 1.94. Lowering the value of *A* to 27 results in a stable equilibrium for $\hat{x}$ of about 45 DNA subunits and a G2 coefficient of variation of 19.2 and G2/G1 cv ratio of 1.97. This system also shows periodic variations in response to large sister selection.

The above considerations demonstrate a possible physical reason for the thresholds that exist in the *T. thermophila* model as well as outlining the physical basis for the f value used in the effective mean calculations.

Appendix D: Average difference between sisters and the unequal division between sisters

Let $F = \frac{\hat{d}}{a}$ be the fractional difference between sisters relative to *a*, DNA content (e.g. *A* or *x*). We show here the relationship between the s.d. θ of a normally distributed f population (mean=0) and the average difference $\hat{d}$ between these sisters is given by:

$$\theta = \frac{F}{2} \cdot \sqrt{\frac{\pi}{2}} \quad . \qquad (1D)$$

We assume that the large and small sisters are given by:

$$\text{L} = \frac{x}{2} + a f \quad , \text{S} = \frac{x}{2} - a f \quad . \qquad (2D)$$

The difference between sisters $= 2af$, and the average of this difference over all cells is :

$$\hat{d} = 2a \int_{-\infty}^{\infty} f \cdot p(f) df. \qquad (3D)$$

Where $p(f)$ is the probability distribution of the normally distributed *f* given by (Since the mean of f is 0 because the sum of $+ a f$ and $- a f$ is 0 for all pairs of sisters):

$$p(f) = \frac{1}{\theta\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{f}{\theta}\right)^2} . \qquad (4D)$$

The variance θ^2 is given by:

$$\theta^2 = \int_{-\infty}^{\infty} f^2 \cdot p(f) df. \qquad (5D)$$

Then the ratio $\frac{\hat{d}}{\theta}$ is given by:

$$\frac{\hat{d}}{\theta} = \frac{2a \int_{-\infty}^{\infty} f \cdot p(f) df}{\sqrt{\int_{-\infty}^{\infty} f^2 \cdot p(f) df}} . \qquad (6D)$$

Since F = $\frac{\hat{d}}{a}$ we can find θ as:

$$\theta = \frac{F\sqrt{\int_{-\infty}^{\infty} f^2 \cdot p(f) df}}{2 \int_{-\infty}^{\infty} f \cdot p(f) df} . \qquad (7D)$$

Integrating using Eq. 4D for $p(f)$ gives for the top $F \cdot \theta$ and the bottom $2 \cdot \theta \sqrt{\frac{2}{\pi}}$ so that:

$$\theta = \frac{F}{2} \cdot \sqrt{\frac{\pi}{2}} \quad . \qquad (8D)$$

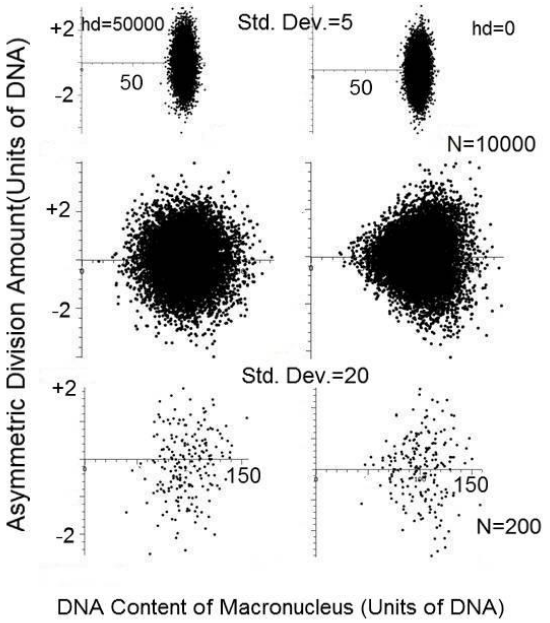

Figure D1. Division asymmetries for h_d=0 (right) and $h_d \rightarrow \infty$ (left) Top row, σ =5, middle row, σ =20 for N=10000, bottom row, σ =20, N=200: Model parameters θ =0.01, A=100.

In order to illustrate the differences between γ when it is independent of macronuclei content ($\gamma = A$ and $h_d = \infty$) and when it is not ($h_d = 0$), we have shown in Figure D1 the distribution of the parameter γf as a function of DNA content for these two extremes of h_d, for two different values of σ and two populations (*N*). Differences between the models begin to appear near σ >10 for $N = 10000$. For small populations ($N < 500$), it is difficult to determine which distribution is actually occurring. Since σ is about 24 for *T. thermophila* the model choice will be significant (the particular choice of θ does not affect the overall conclusion). The simulations also allowed us to verify that numerically Eq. 8D gave the correct relationship.

Appendix E: Model 1 Prereplication and Postreplication Variances and Sister Values

We show here the relation between prereplication and postreplication DNA content for model 1 using a population of *N* macronuclei. From Eq. 1 we can determine the postreplication mean value $\hat{x}'$ to be:

$$\hat{x}' = \hat{x} \cdot s + \frac{A}{1+f} \text{ where } s = 1 + \frac{f}{1+f} = \frac{2f+1}{1+f} \, . \qquad (1E)$$

When $A = \hat{x}$, $\hat{x}' = 2\hat{x}$. Using this, the postreplication s.d. σ' can be determined for the special case $A = \hat{x}$ by:

$$\sigma'^2 = \sum_{k=0}^{n} (x' - \hat{x}')^2 = \sigma^2 \cdot s^2 \quad . \qquad (2E)$$

And the postreplication coefficient of variation cv_{post} can also be found from $cv_{post} = 100\ \sigma'/\hat{x}'$ giving the result:

$$\frac{\mathrm{cv_{post}}}{\mathrm{cv_{pre}}} = \frac{1}{2}s \ . \qquad \text{(3E)}$$

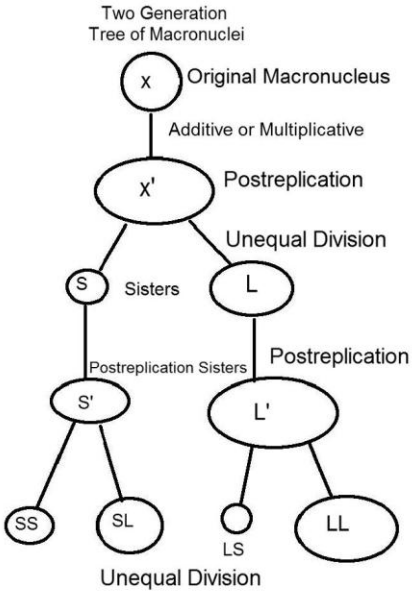

Figure E1. Schematic illustrating the replication/division process through two generations. S and L represent small and large daughter macronuclei respectively.

Some useful information can also be gained from calculating directly general pre and post replications values of sister macronuclei. For this we generalize further from the form of postreplication (Eq. 2E) by allowing the large and small sisters to have different b replication rates (b_L and b_S) and we use $h = 1/f$. The progression of two generations of replication is illustrated in Figure E1 beginning with an initial macronucleus x.

For simplicity, the large and small sisters are assumed to be given by L=$\left(\frac{x'}{2} + \frac{d}{2}\right)$, S=$\left(\frac{x'}{2} - \frac{d}{2}\right)$ where x' is given by:

$$x' = x + \frac{(A + fbx)}{(1+f)} \ . \qquad \text{(4E)}$$

The large (L') and small S' sister postreplication values are:

$$L' = L + \frac{(A + fb_L L)}{(1+f)}, \qquad S' = S + \frac{(A + fb_S S)}{(1+f)}. \qquad \text{(5E)}$$

The difference Δ between the sisters is d and the difference between the postreplication sisters Δ' is:

$$\Delta' = \frac{1+(b_L+1)f}{1+f}\Delta + \frac{f}{f+1}S \cdot (b_L - b_S) \ . \qquad \text{(6E)}$$

This then is a general relation between the postreplication differences between sisters as a function of the prereplication difference between sisters. Since for *T. thermophila* b_L and b_S are not always the same, a plot of these parameter should look very different than for *P. tetraurelia.* Note also that if the postreplication sisters were to divide, the sum of the two subsequent sisters (LL and LS in Figure E1) would return us to the postreplication values. Taking the difference between these might provide additional information about the division process and could be investigated in the future.
In the special case of $b_L = b_S$=b, Eq. 6E reduces to:

$$\Delta' = \frac{1+(b+1)f}{1+f}\Delta \quad \text{or} \quad \Delta' - \Delta = \frac{bf}{1+f} \cdot \Delta. \qquad (7E)$$

This shows a linear relation between the difference between the postreplication sister values and the difference between the prereplication sisters such that the slope is governed by the value of *f* and *b*.

Appendix F: Effect of selection rules on temporal variation

Imposing selection rules on which sister macronucleus survives to subsequent generations allows important information to be recovered about the time scale associated with DNA regulation in *T. thermophila*. The model was altered so that the large sister macronucleus went on to subsequent generations. We shall show that altering this simple feature allows us to clearly determine the time scales on which the regulation occurs. Although perhaps very difficult, if an experiment could be designed to impose this type of selection rule, then as we shall see below, considerable information could be gained about the regulation process. Using the large sister selection rule the simulation generated the oscillatory results as shown in Figure F1. The periodic nature of the result is fairly general and relies primarily on the decoupling of the thresholds from the DNA content itself. The effect disappears if the thresholds become dependent on the average DNA content.
As seen in Figure F1b, the probability distribution has a wave moving through it that translates to periodic variations in the $\hat{x}$ and cv_{post} parameters. To understand this oscillatory behaviour, it is worthwhile considering what occurs when the sister cells with small macronuclei are removed from the distribution. By removing cells with small macronuclei, a gap in the distribution is created, shifting it towards cells with larger macronuclei DNA content. Without a method to restore the balance, this upward shift would continue indefinitely. However, in the case of a multiplicative replication rate dependent on an upper and lower threshold involving an independent parameter such as A, the distribution refills within the thresholds from values outside the thresholds, moving to partially restore the old distribution.

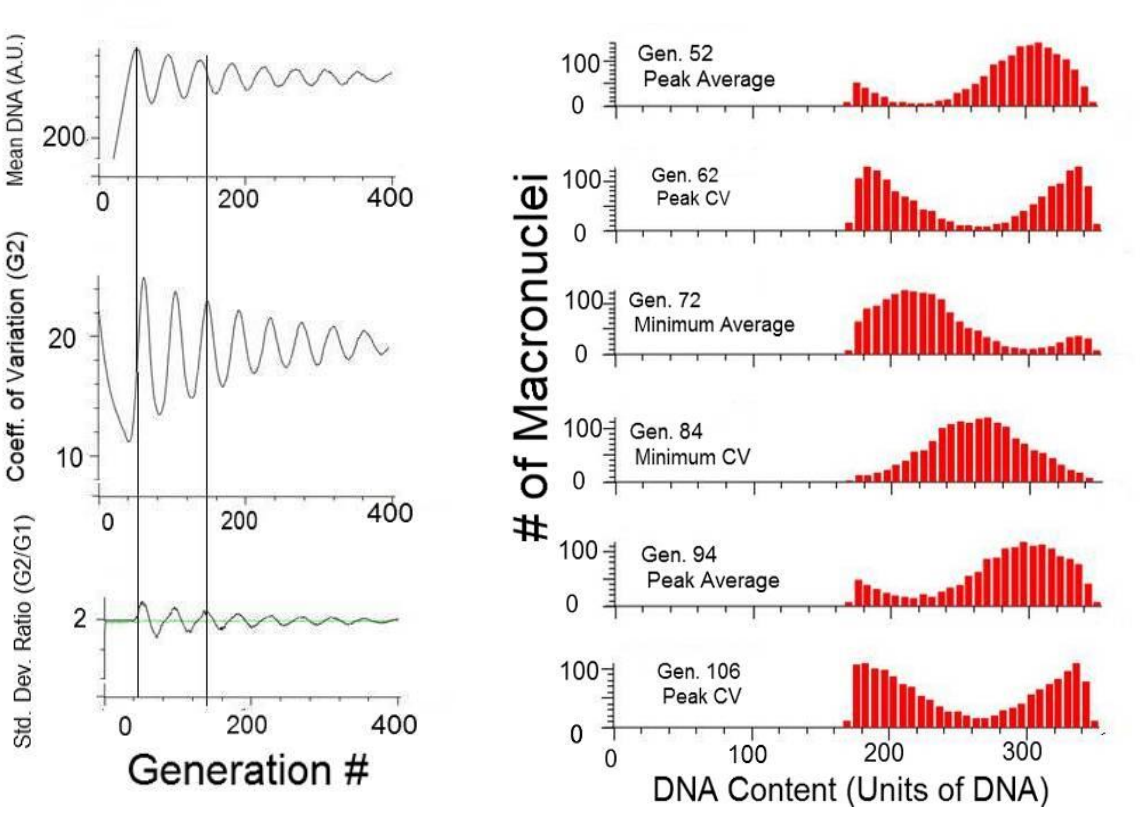

Figure F1a. *T. thermophila* model acting under selection rule that eliminates small sister cells each generation. The mean DNA content (top) and coefficient of variation (middle) and standard deviation ratio (G2/G1) for 400 generations are shown. The *cv* and G2/G1 ratio is $\pi/2$ out of phase with the mean values.
Figure F1b. Histogram distributions of DNA content illustrating a wave moving to the right. Generations shown represent minima and maxima in either the mean value or coefficient of variation. Model parameters: $h = 0.707$, $A = 100$, $h_d = 50000$ (division factor), $\theta = 0.1$, $N = 2000$, initial values of $\sigma = 24$, xˆ=100.

Somewhat remarkably, the result appears to be a wave pattern moving through the distribution causing temporal variations resembling a simple harmonic oscillator whose frequency depends directly on the amount of asymmetric division that is occurring (Figure F2). With more asymmetric division, the wave speed increases.

Critical Value of h ($\boldsymbol{h = \sqrt{2} - 1}$)

If *h* is larger than $\sqrt{2} - 1$ then periodic variations appear in both the coefficient of variation and the mean DNA content of the population. The period depends on both the amount of asymmetric division (inversely proportional) as well as being inversely proportional to $h - (\sqrt{2} - 1)$ (Figure F2). h values less than $\sqrt{2} - 1$) do not yield periodic behaviour.

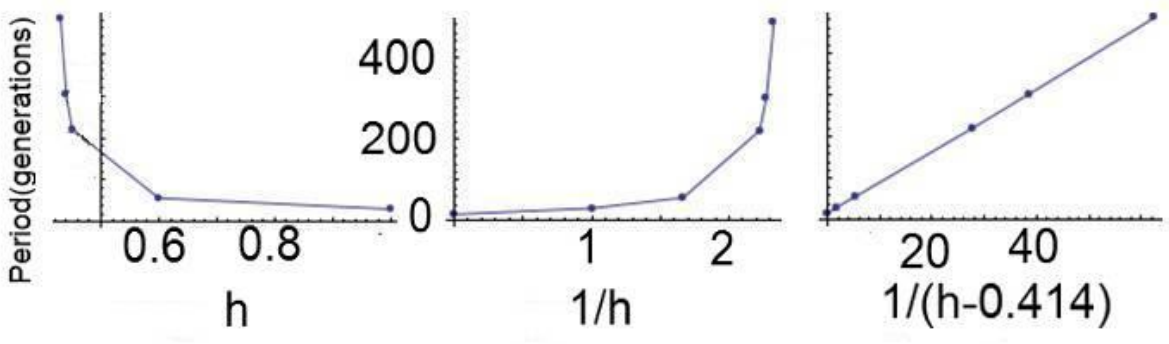

Figure F2. Relationships between the period T and weighted average parameter h. Simulations are for macronuclei undergoing large sister selection. $15000 > N > 5000$. $A = 100$, $\sigma = 24$, $\theta = 0.1$.

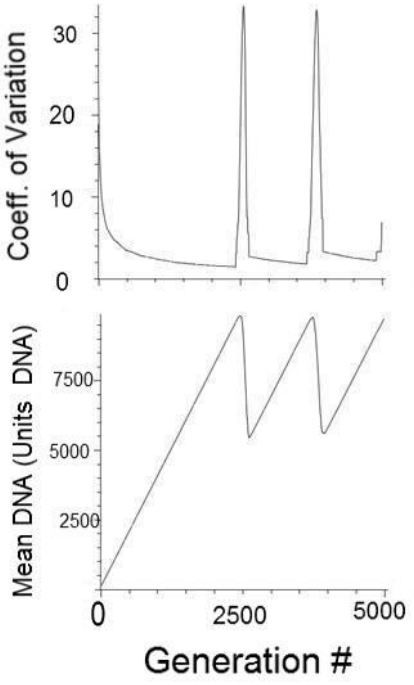

Figure F3. 5000 generations of the mean (top) and coefficient of variation (bottom) for model 2 simulation at $h = 0.42$ (just above the critical number $\sqrt{2} - 1$). Model parameters: $A = 100$, $\sigma = 24$, $h_d = 50000$, $\theta = 0.1$.

Near the critical point, the thresholds get further apart with time but the distribution continues with the same variance thereby becoming a local pocket of DNA content within the widely separated average thresholds. This pocket then moves to higher values until reaching the upper threshold where it splits and a portion moves to the lower threshold resulting in spike in the *cv*. Thus near the critical point $h = \sqrt{2}$ -1 the mean value ramps up linearly with time and then steps quickly down while the coefficient of variation spikes when the mean undergoes rapid change (see Figure F3).

The same thing occurs away from the critical point but in a less dramatic fashion. While this case is unlikely to occur in practise, it helps illustrate what is causing the periodicity.
The periodicity also becomes clearer as *N* increases. Large populations help establish more exactly the probability distribution. If there are not sufficient points to clearly define the distribution changes, then the periodicity is not as clear. Figure F4 shows how for small

populations the data generated looks more random and for older generations the periodicity is less clear.

Any system with generally similar constraints might react in this periodic way where the selection creates a gap that is then partially refilled with newly replicated values. The mean value models do not act in this manner as they do not have a control mechanism preventing them from drifting upward in response to the selection rule. If experiments could be designed to implement this form of selection on *T. thermophila* then we would have a powerful method to investigate the temporal aspects of DNA regulation.

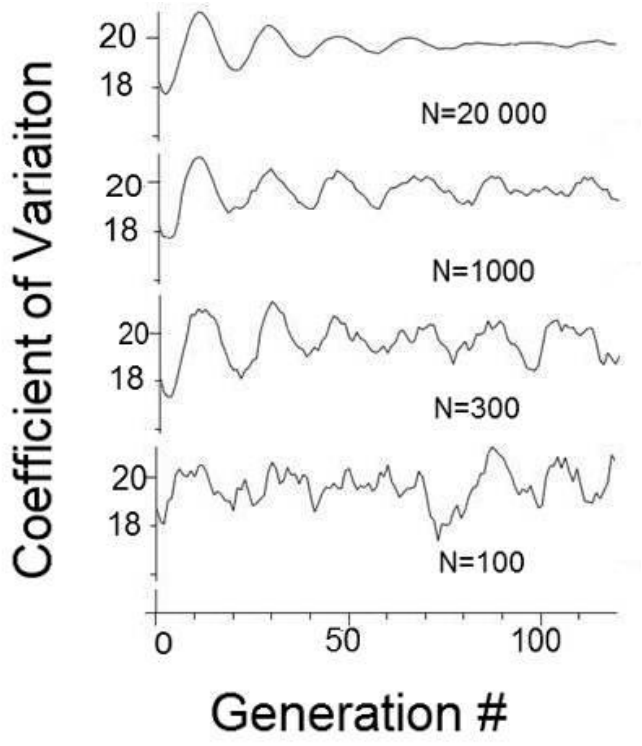

Figure F4. Illustrating the effect of changing the simulation population (N) from 100 (bottom) to 20000 (top). Model parameters are $h_d = 50000$, $h = 50000$, $A = 100$, $\sigma = 24$, $\theta = 0.1$, 120 generations.