Centriole Duplication Model for the 16 cell cyst in *Drosophila* Spermatogenesis

■ General description of the model and biological assumptions

We use the theory of multitype branching processes to evaluate the distribution of cells with a given number of centrioles. This work has some analogies with earlier modeling of plasmid numbers in bacteria[1] . While the model is very general, we consider here the situation of a cyst of 16 cells initiated by a gonial cell at time t=0 that duplicates by binary splitting. We assume that the initial gonial cell contains two centrioles in phase G_1 (wild type configuration). The centrioles duplicate in the cell between G_1 and G_2 phase and are distributed among the two daughter cells at cell division. We thus consider two distinct events in our model. The first event is the duplication of centrioles. We assume that the number of inherited centrioles determines the number of centrosomes in each cell. If a cell inherits only one centriole, it may duplicate or not, generating a centrosome with two or one centrioles, respectively. If a cell inherits a centrosome with two centrioles, they may duplicate or not, but in any instance they will separate and generate two centrosomes (with two or one centrioles, depending on whether centriole duplication occurred or not). This assumption results from the fact that we have never observed problems in centrosome separation (see main text). We also assume that a centriole is always needed to template the formation of new centrioles; if the cell inherits zero centrioles it will not be able to generate new centrioles. Centrioles within the same cell duplicate independently from each other with probability θ so that the number of centrioles in a cell in G_2 is a random variable. The second event is segregation of centrioles to two daughter cells in mitosis. This event brings the new cells into G_1 and to a new phase of centriole duplication. Partitioning is assumed to occur through inheritance of centrosomes. That is, a mutant cell in mitosis having one centrosome with two centrioles will produce one cell with one centrosome with two centrioles, and one cell with zero centrioles.

We construct two probability generating functions (pgfs). The first one corresponds to the whole cycle of duplication and segregation of centrioles to the daughter cells, from G_1 in one cell cycle to G_1 in the next cycle. Because experimental data was collected from 16 cell cysts in phase G_2, we had to generate another pgf that drives G_1 cells into G_2. The first pgf was denoted $f_{G_1}^{(i)}(s_0, s_1, s_2)$ and models the distribution of the number of cells with a given amount of centrioles (0,1,2) in phase G_1 produced by a cell with a number of centrioles i in G_1. The subscript j in the dummy variable s_j refers to the type of the progeny (thus, s_0, s_1, s_2 refer to cells with 0, 1 and 2 centrioles respectively). This pgf determines the distribution of the state of the cell lineage over time in phase G_1. We then construct the pgf denoted $f_{G_2}^{(i)}(s_0, s_1, s_{2s}, s_{2t}, s_3, s_4)$ of the distribution of cell types in phase G_2 that can be reached from a cell in phase G_1. We assume that a cell can be in five different states in phase G_2, that is with centriole number 0, 1, 2 in the same centrosome (dummy variable s_{2t}; corresponds to only one centrosome), 2 in different centrosomes (dummy variable s_{2s};corresponds to 2 centrosomes, 1 centriole per centrosome), 3 (corresponds to 2 centrosomes; one with 1 centriole the other with 2) and 4 (corresponds to 2 centrosomes, both with 2 centrioles). Those two pgfs allow us to evaluate the distribution of the number of cell types after four divisions, that is when there are 16 cells in the cyst, before they enter the meiotic divisions.

■ Generating function from phase G_1 to next phase G_1

A cell inheriting zero centrioles can only produce two daughter cells with zero centrioles. The pgf for such a cell thus reads

$$f_{G_1}^{(0)}(s_0) = s_0^2.$$

A cell inheriting one centriole will duplicate its centriole with probability θ, in which case it produces one daughter cell with two centrioles and another daughter cell with zero centrioles. With complementary probability 1-θ, the centriole does not duplicate and the cell produces one daughter cell with one centriole and the other cell with zero centrioles. Accordingly, the pgf of the distribution of the number of descendants of each type produced by such a cell is

$$f_{G_1}^{(1)}(s_0, s_1, s_2) = s_0((1-\theta)s_1 + \theta s_2).$$

Finally, a cell inheriting two centrioles duplicates both of its centrioles with probability θ^2 and produces two daughter cells with two centrioles. It duplicates only one centriole with probability $2\theta(1\text{-}\theta)$ and thus produces one cell with one centriole and the other cell with two centrioles. Finally, the cell duplicates none of its centrioles with probability $(1-\theta)^2$ and produces two daughter cells with one centriole. Then, the pgf of the distribution of cell types produces by a such cell reads

$$f_{G_1}^{(2)}(s_0, s_1, s_2) = ((1-\theta)s_1 + \theta s_2)^2.$$

With these three pgfs in hand, we can evaluate the pgf $F_{G_1}^{(i)}$ of the total distribution of cell types in the population initiated by a single cell of type i. This pgf evolves over time according to the recursion[2] :

$$F_{G_1(t+1)}^{(i)}(s_0, s_1, s_2) = F_{G_1(t+1)}^{(i)}(f_{G_1}^{(0)}, f_{G_1}^{(1)}, f_{G_1}^{(2)}).$$

Here, we follow the development of the cell line only over four generations, that is when the population comprises 16 cells. The pgf of the distribution of cell types at t=4 can then be evaluated explicitely and reads

$$F_{G_1(4)}^{(2)}(s_0, s_1, s_2) = a^2 b^2 ((1-\theta)s_0^2 + \theta a b)^2 ((1-\theta)s_0^4 + r a b ((1-\theta)s_0^2 + \theta a b))^2,$$

where

$$a = (1-\theta)s_1 + \theta s_2$$

$$b = (1-\theta)(s_0 + \theta s_1) + r^2 s_2.$$

■ Generating function from phase G_1 to phase G_2

Here we evaluate the pgf of the transition probabilities of a cell in state i in phase G_1, which reaches state j in phase G_2 at time t=4. We incorporate an error rate ϵ, which is the probability that a centrosome was missed by the experimentator, such an error is most likely due to the loss of centrosomes during manipulation of the cells during the preparation of the slides. The error rate ϵ thus represents an exogeneous factor that decreases the probability of observing centrosomes that are present. This rate is has been set to $\epsilon \equiv 0.075$, This value being estimated from data on wild type cysts of 16 cells. Below we detail the generating functions.

A cell with zero centrioles can only move to a state with zero centrioles, hence

$$f_M^{(0)}(s_0) = s_0.$$

A cell starting with one centriole ends up with one centrosome with one or two centrioles depending on whether duplication was successful or not. The centrosome will contain two centrioles with probability θ or remain with one centriole with complementary probability 1-θ. In both cases, there is a probability ϵ that the centrosome is not observed by the experimentator. Taking all events into account, the pgf for the distribution of cell types in phase G_2 of a cell initially with one centriole in phase G_1 reads

$$f_M^{(1)}(s_0, s_1, s_{2t}) = \epsilon\, s_0 + (1-\epsilon)\,((1-\theta)\, s_1 + \theta\, s_{2t}).$$

The situation becomes more complex for a cell inheriting two centrioles, as it can end up in five different states in phase G_2. After duplication, centrosomes can be missed by the experimentator. Both centrosomes are missed with probability (ϵ^2), one with probability $(2\,\epsilon(1-\epsilon))$ or none with probability $(1-\epsilon)^2$. Taking all possible different events of duplication and experimental error into account, we find that the pgf for the distribution of cell types in phase G_2 (including experimental error) of a cell with two centrioles in phase G_1 reads

$$\begin{aligned} f_M^{(2)}(s_0, s_1, s_{2s}, s_{2t}, s_3, s_4) &= \epsilon^2\, s_0 \\ &+ s_1\,(2\,\theta(1-\theta)\,(1-\epsilon)\,\epsilon + (1-\theta)^2\,2\,(1-\epsilon)\,\epsilon) \\ &+ s_{2s}\,(2\,\theta(1-\theta)\,(1-\epsilon)\,\epsilon + \theta^2\,2\,(1-\epsilon)\,\epsilon) \\ &+ s_{2t}\,(1-\theta)^2\,(1-\epsilon)^2 \\ &+ s_3\,2\,\theta(1-\theta)\,(1-\epsilon)^2 \\ &+ s_4\,\theta^2\,(1-\epsilon)^2 \end{aligned}$$

We can now evaluate the pgf of the distribution of cell types after four cell divisions and five rounds of centriole duplication and a final round of experimental error. Under the assumption that the 16 cells in the cyst were initiated by a single stem cell with two centrioles, we get

$$F_{M(4)}^{(2)}(s_0, s_1, s_{2s}, s_{2t}, s_3, s_4) = F_{G_1(4)}^{(2)}(f_M^{(0)}, f_M^{(1)}, f_M^{(2)}).$$

■ Dynamics of the mean number of centrioles

The mean number of cell type j in phase G_1 produced by a cell type i in phase G_1 is

$$\frac{\partial f_{G_1}^{(i)}}{\partial s_j}\Bigg|_{s_0=s_1=s_2=1},$$

from which we obtain the dynamic of the expected number of cells with one (n_1) and two (n_2) centrioles[2]

$$\begin{pmatrix} n_1 \\ n_2 \end{pmatrix}(t+1) = \begin{pmatrix} 1-\theta & 2(1-\theta) \\ \theta & 2\theta \end{pmatrix}\begin{pmatrix} n_1 \\ n_2 \end{pmatrix}(t).$$

We solve this system of equations with the initial conditions $n_1(0) = 0$ and $n_2(0) = 1$. Noting that the total number of cells in the population is given by 2^t, we find the explicit dynamics

$$\begin{aligned} n_0(t) &= 2^t - 2(1+\theta)^{t-1} \\ n_1(t) &= 2(1-\theta)(\theta+1)^{t-1} \\ n_2(t) &= 2\theta(\theta+1)^{t-1}. \end{aligned}$$

To evaluate the mean number of cells in phase G_2 we use the transition probabilities from phase G_1 to phase G_2 as described in the previous section and set $\epsilon \equiv 0$. The situation is graphed in Fig.4 in the main text with the value of the duplication rate $\hat{\theta}$ estimated from the data.

Statistics

The pgf $F_{G2(4)}^{(2)}$ contains the entire distribution of the number of cells with a given number of centrioles after four cell divisions, five rounds of centriole duplications and a a final round of experimental error. The joint probability distribution of the random number of cells ($X_0, X_1, X_{2t}, X_{2s}, X_3, X_4$) of a given type at the time of the observation can be extracted from the pgf according to the formula

$$P(X_0 = x_0, \ldots, X_4 = x_4) = \frac{1}{x_0! \ldots x_4!} \frac{\partial^{x_1+\ldots+x_4} F_{M4}^{(2)}}{\partial^{x_1} s_0, \ldots, \partial^{x_4} s_4}\Bigg|_{s_0=\ldots=s_4=0},$$

where we necessarily have

$$\sum_{i=0}^{5} x_i = 16.$$

The likelihood of the data is given by

$$L = \prod_{i=1}^{n} P_n(X_0 = x_0, \ldots, X_4 = x_4),$$

where P_n is the probability of the n's observation. The data analysed represents 22 replicates of SAK mutant cysts of 16 cells (Fig 4) and 15 replicates of wild type cysts of 16 cells (Fig 4). The likelihood function obtained given

the data is graphed in the inset of Fig.4 in the main text. The Maximum likelihood estimate $\hat{\theta}$ of the duplication rate is obtained from

$$\frac{\partial L}{\partial \theta}\Big|_{\theta=\hat{\theta}} = 0.$$

References

1. Seneta,E.& Tavaré S.Some stochastic models for plasmid copy number.Theor.Pop.Biol.23,241-256 (1983)

2. Caswell, H. Matrix Population Models: Construction, Analysis, and Interpretation (Sinauer, Sunderland, 2001).