

Stochasticity and robustness in DNA replication

Physical Review Letters **108**, 058101 (2012).

Alessandro Moura¹

Jens Kerschau^{1,2}, Julian Blow²

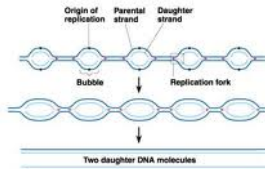
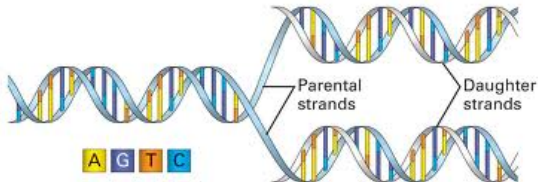
¹Institute for Complex Systems and Mathematical Biology, University of Aberdeen, UK

²Cancer Research UK Chromosome Replication Research Group, University of Dundee, UK

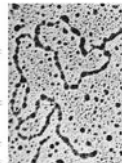
July 18, 2012



DNA replication



© 1999 Addison Wesley Longman, Inc.



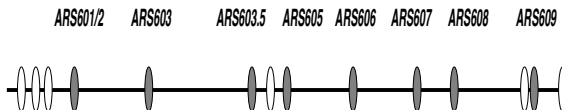
(b)

0.25 μ m

Basic facts about DNA replication

- Replication starts at precise locations on the DNA, called replication origins.
- Origin activation is stochastic, and different origins may have different activation time distributions.
- If a fork reaches a not yet activated origin, the origin can no longer fire, and it is said to be passively replicated.
- In yeast and some other organisms, origin locations are fixed, and defined by specific DNA sequences.
- In *Xenopus* eggs, in contrast, origins have no preferential location.

Map of origin locations in Chromosome VI of *S. cerevisiae*

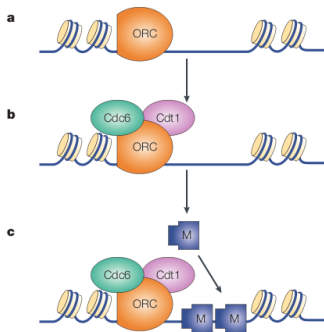


Important questions

- What is the optimum placement of origins from the point of view of replication efficiency?
- Has natural selection optimized the origin distribution?

Not all origins are capable of firing

Origins must be licensed before S phase



Definition of origin potential

The potential p of an origin is the probability that it has been licensed by the start of S phase (Moura, Retkute, Hawkins, Nieduszynski, *Nucleic Acid Research*, 2010; Retkute, Nieduszynski, Moura, *Physical Review Letters*, 2011).

Simple case: two origins with same potential p

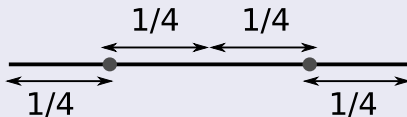
Same activation time

Average replication time is given by

$$T_{\text{rep}} = (1 - p)^2 T_0 + 2p(1 - p) T_1 + p^2 T_2. \quad (1)$$

Origins located at $x = 1/4$ and $x = 3/4$

$$T_d = (1 - p)^2 T_0 + \frac{3}{2}p - \frac{5}{4}p^2 \quad (2)$$



Simple case: two origins with same potential p

Same activation time

Average replication time is given by

$$T_{\text{rep}} = (1 - p)^2 T_0 + 2p(1 - p) T_1 + p^2 T_2 \quad (1)$$

Both origins located at $x = 1/2$

$$T_s = (1 - p)^2 T_0 + p - \frac{1}{2} p^2 \quad (3)$$



Probability-dependent optimal origin localization

configuration where origins are together is optimal, if $T_s < T_d$:

$$p - \frac{1}{2}p^2 < \frac{3}{2}p - \frac{5}{4}p^2,$$

which gives:

$$\boxed{p < \frac{2}{3}}. \quad (4)$$

It is optimal for origins with efficiency less than $\approx 67\%$ to cluster.

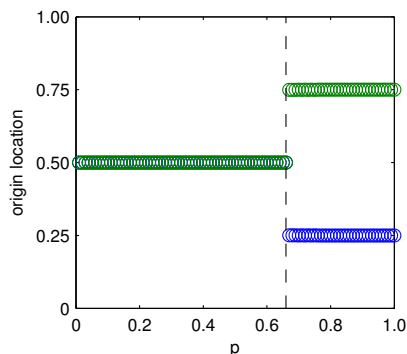
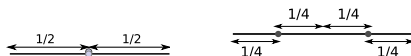
Probability-dependent optimal origin localization

configuration where origins are together is optimal, if $T_s < T_d$:

$$p - \frac{1}{2}p^2 < \frac{3}{2}p - \frac{5}{4}p^2,$$

which gives:

$$\boxed{p < \frac{2}{3}}. \quad (4)$$



It is optimal for origins with efficiency less than $\approx 67\%$ to cluster.

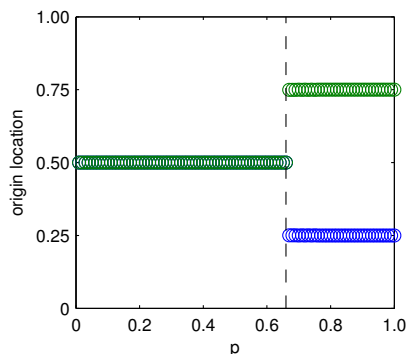
Probability-dependent optimal origin localization

configuration where origins are together is optimal, if $T_s < T_d$:

$$p - \frac{1}{2}p^2 < \frac{3}{2}p - \frac{5}{4}p^2,$$

which gives:

$$\boxed{p < \frac{2}{3}}. \quad (4)$$



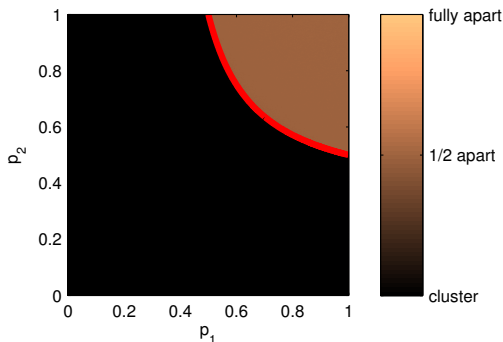
It is optimal for origins with efficiency less than $\approx 67\%$ to cluster.

2 origins with different probabilities

Origins will cluster

if $T_{x=1/2} < T_{x=1/4}$:

$$p_2 > \frac{p_1}{3p_1 - 1} \quad (5)$$



For origins with potential less than 50%, it is always better to cluster with other origins

Clustering for n origins

Case for n origins having the same probability:

Effective cluster probability:

$$p_{\text{eff}} = 1 - (1 - p)^m. \quad (6)$$

Clusters of n origins break up symmetrically into 2 clusters of $n/2$. For the condition $p > p_c$:

$$p_{\text{eff}} = 2/3,$$

with: $m = n/2$

$$p_c = 1 - \frac{1}{\sqrt[n]{9}}. \quad (7)$$

Clustering for n origins

Case for n origins having the same probability:

Effective cluster probability:

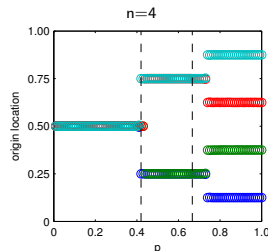
$$p_{\text{eff}} = 1 - (1 - p)^m. \quad (6)$$

Clusters of n origins break up symmetrically into 2 clusters of $n/2$. For the condition $p > p_c$:

$$p_{\text{eff}} = 2/3,$$

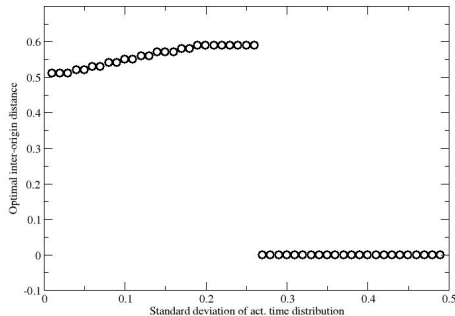
with: $m = n/2$

$$p_c = 1 - \frac{1}{\sqrt[n]{9}}. \quad (7)$$



Optimal inter-origin distance in a 2-origin chromosome

For a simple 2-origin system, we find a sharp transition from clustered to isolated origins as the standard deviation in origin activation time increases.



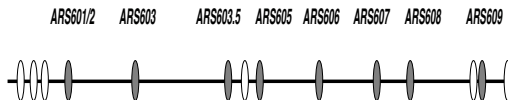
Extension for more realistic firing dynamics

The clustered-isolated transition is robust!

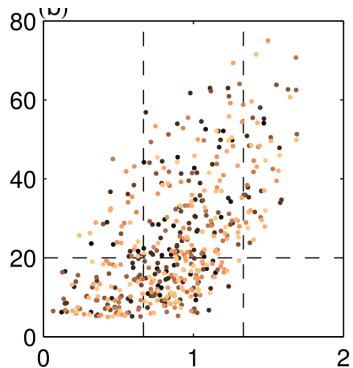
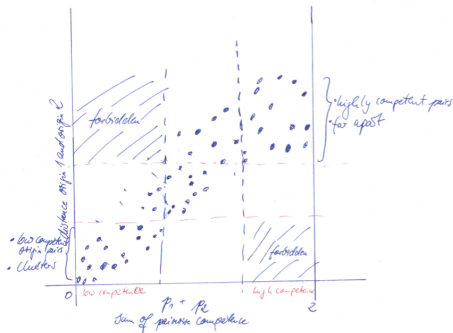
- So far we assumed that the origins fire simultaneously at fixed times; but in reality origin firing is stochastic.
- Monte-Carlo simulations reveal the same sharp transition from clustered to isolated origins in stochastically firing origins as their potential increases. Stochasticity shifts the transition to higher potentials.
- This phenomenon is also present if the average activation times of the origins differ. In this case, the isolated regime is no longer symmetrical.

Conclusions so far

- In the optimal situation, highly efficient origins are isolated.
- Less efficient origins tend to cluster.
- origins tend to be separated, if $p_{eff} \gtrsim 70\%$.
- We therefore expect to see a trend in yeast origin placement, with isolated origins being more likely to be highly efficient, and inefficient origins being found in clusters.



Comparison with “data”



DNA replication in *Xenopus*

- In *Xenopus* eggs, licensing can take place at any point in the DNA, independently of position.
- A licensing complex can load more than one Mcm ring; so clusters of MCM can be created during licensing.
- Each Mcm (pair) can give rise to replication forks, so having clustered Mcm molecules is equivalent to having clustered origins.
- It is believed that most Mcm's on *Xenopus* egg extracts are organized in clusters of about 10 Mcm's, separated by an average distance of 10 kb.

Question

What is the optimal distribution for the Mcm rings? Does this explain the observed clusters?

Optimality in *Xenopus* replication

Origins tend to cluster within a chromosome length of d when $2v\sigma$ becomes comparable to d ; v is the fork velocity.

Using $v = 1.5\text{kb/min}$, this predicts clustering within a length of about 10kb, for σ in the range 5–10 minutes, which is the best estimate. This agrees with current knowledge.

Optimal clustering in Xenopus

The optimal number of origins in a cluster predicted by the model is found to be 5–10, which is close to the current estimates.

