

# A stochastic hybrid model of DNA re-replication across a complete genome

Maria Anna Rapsomaniki, Stella Maxouri, Manuel Ramirez Gar-  
rastacho, Patroula Nathanailidou, Nickolaos Nikiforos Giakoumakis,  
Stavros Taraviras, John Lygeros and Zoi Lygerou

## S1 Model Description

### S1.1 Model states and dynamics

Let  $L_i, i = 1, \dots, n$  denote the genomic location of each origin and  $S_j, j = 1, \dots, m$  denote the DNA strand it is located on. Each origin copy, whether ancestral or offspring, can be identified by its location and strand index and is denoted as  $\text{Ori}_{i,j}$ . Initially, all origins are located on the parental strand,  $j = 1$ .

The discrete dynamics of the model are associated with the discrete origin states. Let  $q_{i,j}(t) \in \{\text{PreR}, \text{RB}, \text{RL}, \text{RR}, \text{PassR}, \text{PostR}\}$  denote the state of  $\text{Ori}_{i,j}$  at time  $t$ . It is:

1. **PreR**: Pre-replicative state, if the origin has not yet fired or been passively replicated.
2. **RB**: If the origin has fired, and the replication forks it has initiated in the process are still active in both directions.
3. **RR/RL**: The left/right replication fork initiated by the firing of the origin has met a replication fork from a different origin moving in the opposite direction. The genome between the two origins has been completely replicated on this particular strand, and only the right/left fork initiated by the origin in question is still active.
4. **PassR**: If the origin has been passively replicated by a replication fork initiated by an adjacent origin on the same strand.
5. **PostR**: Both the right and left replication forks initiated by the firing of the origin have met forks initiated by adjacent origins moving in the opposite direction.

Contrary to the normal replication model, the total number of origins in re-replication increases over time, starting with  $n$  individual origins on strand  $j = 1$ . The continuous dynamics model the movement of the replication forks. The continuous dynamics are deterministic and captured by two differential equations whose states represent the number of bases replicated as the forks emanating from an active origin move to the left and right. Let  $RF_{i,j}(t), LF_{i,j}(t)$  denote the position of the right and left fork of  $\text{Ori}_{i,j}$  at time  $t$  and  $v$  represent the replication fork speed, measured in bases per minute (see Fig.S1).

It is:

$$\frac{d}{dt}RF_{i,j}(t) = \begin{cases} v(RF_{i,j}(t)) & q_{i,j}(t) \in \{\text{RB}, \text{RR}\} \\ 0, & \text{otherwise} \end{cases} \quad (\text{S1})$$

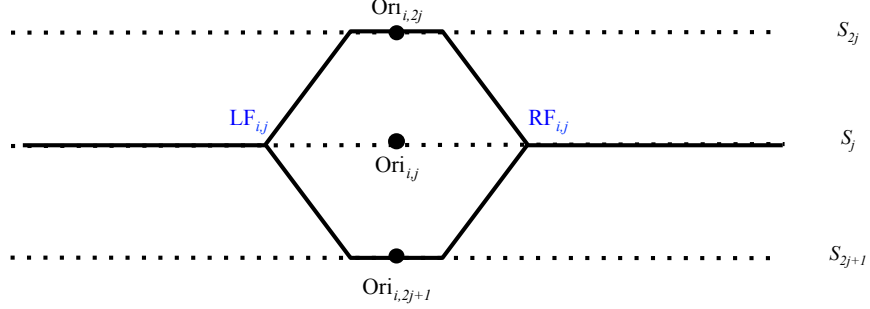


Figure S1: **Indices of origin and fork locations.** Black dots represent origins on different strands, solid lines represent synthesized DNA and dashed horizontal lines different strands. Origin indices are given in black and fork indices in blue.

$$\frac{d}{dt}LF_{i,j}(t) = \begin{cases} -v(RF_{i,j}(t)) & q_{i,j}(t) \in \{RB, RL\} \\ 0, & \text{otherwise} \end{cases} \quad (\text{S2})$$

Uncertainty plays a vital role in re-replication, and it is represented by modeling the time and location of origin firing and re-firing as stochastic events. This is achieved by associating each origin with an intrinsic probability to fire, also referred to as firing propensity, derived from experimentally defined firing efficiencies. Let  $FE_{i,j}$  denote the experimentally determined firing efficiency (i.e. the percentage of cells in which an origin fires during one cell cycle) and  $TF_{i,j}$  the firing time. The probability that  $\text{Ori}_{i,j}$  has fired by time  $t$  is:

$$P[TF_{i,j} \geq t] = e^{-\lambda_{i,j}t}$$

where  $\lambda_{i,j}$  denotes the firing propensity of  $\text{Ori}_{i,j}$ , i.e. the probability to fire in a unit time. It is:

$$\lambda_{i,j} = -\frac{1}{T_c} \ln(1 - FE_{i,j}) \quad (\text{S3})$$

where parameter  $T_c$  denotes the completion time of a normal S phase replication. Using the above equation, the firing propensities  $\lambda_{i,j}$  can be derived from the experimentally defined firing efficiencies  $FE_{i,j}$  in a straightforward manner.

## S1.2 Discrete state transitions

Transitions between discrete states, depicted as arrows in Fig.1B, depend on both the continuous and stochastic dynamics of the system. Origin firing is represented through the spontaneous  $\text{PreR} \rightarrow \text{RB}$  transition and is modeled as a stochastic event, governed by an exponentially distributed random variable whose rate depends on each origin's firing propensity. The rest of the discrete state transitions are deterministic and governed by guards, logical statements that depend on the evolution of the continuous state.

The  $\text{RB} \rightarrow \text{RR}|\text{RL}$  transitions take place when the left/right replication fork emanating from an active origin meets the right/left replicating fork of the nearest active neighbor to its left/right on the same strand (merging events). Similarly, the  $\text{RB}|\text{RR}|\text{RL} \rightarrow \text{PostR}$  transitions occur when replication forks have merged in both directions. The  $\text{PreR} \rightarrow \text{PassR}$  transition occurs when, before an origin has fired, it is passively replicated by a fork emanating from another active origin on the same strand, to the left or right. In contrast to normal DNA replication, origins that have already fired or have been passively replicated can re-fire multiple times. To model this, when origin  $\text{Ori}_{i,j}$  fires (transition  $\text{PreR} \rightarrow \text{RB}$ ) or is passively replicated (transition  $\text{PreR} \rightarrow \text{PassR}$ ), it generates two origins on two resulting strands (denoted as  $\text{Ori}_{i,2j}$  and  $\text{Ori}_{i,2j+1}$ ) which automatically fall back into the  $\text{PreR}$  state and can thus fire again.

The guard conditions together with the continuous state dynamics allow us to compute transition times for each origin. Let  $\text{Ori}_{l,j}$ ,  $1 \leq l \leq (i-1)$  denote the closest origin from the left of  $\text{Ori}_{i,j}$  on the same strand, for which  $q_{l,j} \in \{\text{RB}, \text{RR}\}$ . Likewise, let  $\text{Ori}_{r,j}$ ,  $(i+1) \leq r \leq n$  denote the closest origin from the right of  $\text{Ori}_{i,j}$  on the same strand, for which  $q_{r,j} \in \{\text{RB}, \text{RL}\}$  (see Fig.S2). In other words:

$$\begin{aligned} \text{Ori}_{l,j} &= \max\{l \leq i \mid q_{l,j} \in \{\text{RB}, \text{RR}\}\} \\ \text{Ori}_{r,j} &= \min\{r \geq (i+1) \mid q_{r,j} \in \{\text{RB}, \text{RL}\}\} \end{aligned} \quad (\text{S4})$$

Now let us define the following transitions between the discrete states of the model:

1.  **$\text{PreR} \rightarrow \text{RB}$** . Occurs when an origin fires and results in the production of 2 offsprings on 2 resulting strands.
2.  **$\text{RB} \rightarrow \text{RR}$** . Occurs when the left fork of  $\text{Ori}_{i,j}$  merges with the right fork of  $\text{Ori}_{l,j}$ , resulting in  $\text{Ori}_{i,j}$  replicating only to the right. For the merging to occur, the following guard condition must be satisfied:

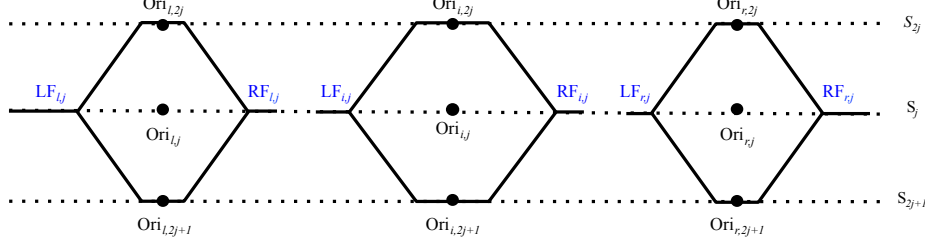


Figure S2: **Indices of origin and fork locations of left and right incoming origins.** Black dots represent different origins, solid lines represent synthesized DNA and dashed horizontal lines different strands. Origin indices are given in black and fork indices in blue.

$$(RF_{l,j} \geq LF_{i,j}) \wedge (RF_{i,j} < LF_{r,j})$$

Then, the time of transition, denoted as  $TSL$  (Time to Stop from the Left), is computed as follows:

$$TSL_{i,j} = \frac{L_i - L_l + TF_{i,j} * v + TF_{l,j} * v}{2v}$$

3. **RB**  $\rightarrow$  **RL**. Similarly with the above, occurs when the right fork of  $Ori_{i,j}$  merges with the left fork of  $Ori_{r,j}$ , resulting in  $Ori_{i,j}$  replicating only to the left. The following guard condition must be satisfied:

$$(LF_{r,j} \leq RF_{i,j}) \wedge (LF_{i,j} > RF_{l,j})$$

The time of transition  $TSR$  (Time to Stop from the Right) is defined as follows:

$$TSR_{i,j} = \frac{L_r - L_i + TF_{i,j} * v + TF_{r,j} * v}{2v}$$

4. **RR**  $\rightarrow$  **PostR**: Occurs when  $Ori_{i,j}$  has already merged from the left and subsequently merges from the right. The guard is defined as follows:

$$LF_{r,j} \leq RF_{i,j}$$

The time of transition  $TSR$  is defined as above.

5. **RL**  $\rightarrow$  **PostR**: Occurs when  $\text{Ori}_{i,j}$  has already merged from the left and subsequently merges from the right. The guard is defined as follows:

$$RF_{l,j} \leq LF_{i,j}$$

The time of transition  $TSL$  is defined again as above.

6. **RB**  $\rightarrow$  **PostR**: Occurs when  $\text{Ori}_{i,j}$  merges simultaneously from the left and the right. The guard is defined as follows:

$$(RF_{l,j} \geq LF_{i,j}) \wedge (LF_{r,j} \leq RF_{i,j})$$

Since the merging events are simultaneous, the time of transition is  $TSL = TSR$ , computed again as above.

7. **PreR**  $\rightarrow$  **PassR**: Occurs when  $\text{Ori}_{i,j}$  has not fired yet and is passively replicated by any incoming origin from its left or right. Results in the production of two offspring origins, denoted as  $\text{Ori}_{i,2j}, \text{Ori}_{i,2j+1}$ . For the transition to occur, the following guard statement must be satisfied:

$$(RF_{l,j} \geq L_i) \vee (LF_{r,j} \leq L_i)$$

Then the time of transition (Time of Passive Replication - TPR) is computed as follows:

$$TPR_{i,j} = \begin{cases} \frac{L_i - L_l}{v} + TF_{l,j} & \text{if } RF_{l,j} \geq L_i \\ \frac{L_r - L_i}{v} + TF_{r,j} & \text{if } LF_{r,j} \leq L_i \end{cases}$$