Modelling DNA sequence evolution with interacting particle systems

M. Falconnet[1] N. Gantert[2] E.Saada[3]

[1]Université d'Évry Val d'Essonne, CNRS – France

[2]Technische Universität München – Germany

[3]CNRS, Paris Descartes – France

Stochastic Models in Ecology, Evolution and Genetics,
Angers, 9-13th December 2013

1. **Nucleotidic substitution processes**
   - The origins: Jukes and Cantor model
   - Entering the field of interacting particle systems
   - Model properties

2. **Extension**
   - Adding ‘cut-and-paste’ mechanism
   - Results
   - How to use the dual process

1. **Nucleotidic substitution processes**
   - The origins: Jukes and Cantor model
   - Entering the field of interacting particle systems
   - Model properties

2. **Extension**
   - Adding 'cut-and-paste' mechanism
   - Results
   - How to use the dual process

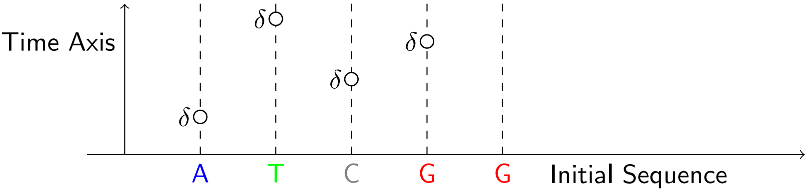
Stochastic nucleotidic substitution models

- A DNA sequence η is an element of $\{\mathrm{A}, \mathrm{T}, \mathrm{C}, \mathrm{G}\}^N$, $N \in \mathbb{N}^*$.
- **Independent** evolution of the sites according to a **Markovian kernel**.

Example: Jukes and Cantor model (1969)

	A	T	C	G
A	$\cdot$	λ	λ	λ
T	λ	$\cdot$	λ	λ
C	λ	λ	$\cdot$	λ
G	λ	λ	λ	$\cdot$

- **Diagonal** entry $-q_{aa}$ is the substitution rate of nucleotide a, here $q_{aa} = -3\lambda$.
- **Non-diagonal** entry q_{ab} is the substitution rate of nucleotide a by b, here $q_{ab} = \lambda$.



δ-marks distributed according to a Poisson process with rate 4λ. The nucleotide is substituted by a with probability $1/4$.

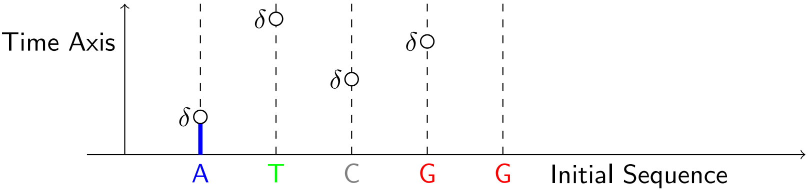
Stochastic nucleotidic substitution models

- A DNA sequence η is an element of $\{A, T, C, G\}^N$, $N \in \mathbb{N}^*$.
- **Independent** evolution of the sites according to a **Markovian kernel**.

Example: Jukes and Cantor model (1969)

	A	T	C	G
A	$\cdot$	λ	λ	λ
T	λ	$\cdot$	λ	λ
C	λ	λ	$\cdot$	λ
G	λ	λ	λ	$\cdot$

- **Diagonal** entry $-q_{aa}$ is the substitution rate of nucleotide a, here $q_{aa} = -3\lambda$.
- **Non-diagonal** entry q_{ab} is the substitution rate of nucleotide a by b, here $q_{ab} = \lambda$.



δ-marks distributed according to a Poisson process with rate 4λ. The nucleotide is substituted by a with probability $1/4$.

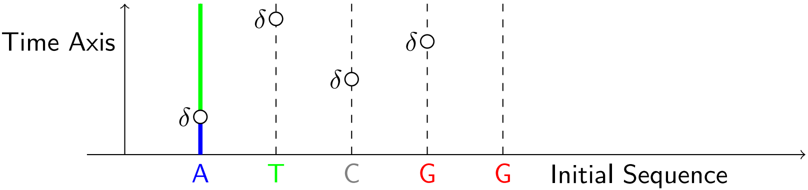
Stochastic nucleotidic substitution models

- A DNA sequence η is an element of $\{A, T, C, G\}^N$, $N \in \mathbb{N}^*$.
- **Independent** evolution of the sites according to a **Markovian kernel**.

Example: Jukes and Cantor model (1969)

	A	T	C	G
A	$\cdot$	λ	λ	λ
T	λ	$\cdot$	λ	λ
C	λ	λ	$\cdot$	λ
G	λ	λ	λ	$\cdot$

- **Diagonal** entry $-q_{aa}$ is the substitution rate of nucleotide a, here $q_{aa} = -3\lambda$.
- **Non-diagonal** entry q_{ab} is the substitution rate of nucleotide a by b, here $q_{ab} = \lambda$.



δ-marks distributed according to a Poisson process with rate 4λ. The nucleotide is substituted by a with probability $1/4$.

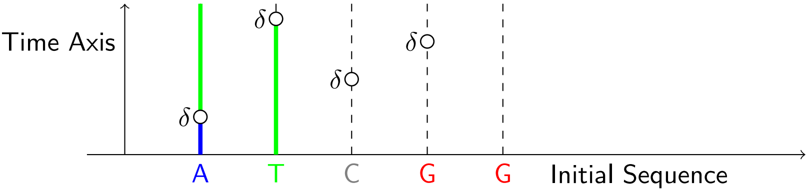
Stochastic nucleotidic substitution models

- A DNA sequence η is an element of $\{A, T, C, G\}^N$, $N \in \mathbb{N}^*$.
- **Independent** evolution of the sites according to a **Markovian kernel**.

Example: Jukes and Cantor model (1969)

	A	T	C	G
A	$\cdot$	λ	λ	λ
T	λ	$\cdot$	λ	λ
C	λ	λ	$\cdot$	λ
G	λ	λ	λ	$\cdot$

- **Diagonal** entry $-q_{aa}$ is the substitution rate of nucleotide a, here $q_{aa} = -3\lambda$.
- **Non-diagonal** entry q_{ab} is the substitution rate of nucleotide a by b, here $q_{ab} = \lambda$.



δ-marks distributed according to a Poisson process with rate 4λ. The nucleotide is substituted by a with probability $1/4$.

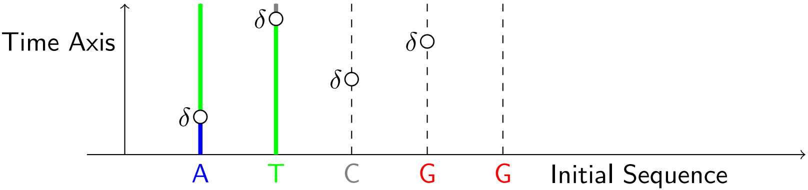
Stochastic nucleotidic substitution models

- A DNA sequence η is an element of $\{A, T, C, G\}^N$, $N \in \mathbb{N}^*$.
- **Independent** evolution of the sites according to a **Markovian kernel**.

Example: Jukes and Cantor model (1969)

	A	T	C	G
A	$\cdot$	λ	λ	λ
T	λ	$\cdot$	λ	λ
C	λ	λ	$\cdot$	λ
G	λ	λ	λ	$\cdot$

- **Diagonal** entry $-q_{aa}$ is the substitution rate of nucleotide a, here $q_{aa} = -3\lambda$.
- **Non-diagonal** entry q_{ab} is the substitution rate of nucleotide a by b, here $q_{ab} = \lambda$.



δ-marks distributed according to a Poisson process with rate 4λ. The nucleotide is substituted by a with probability $1/4$.

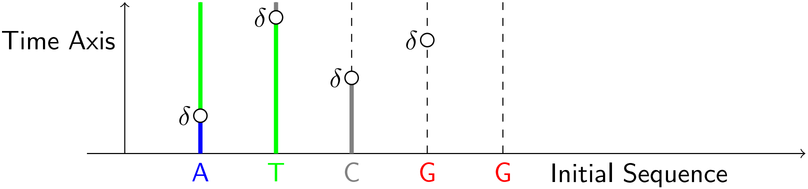
Stochastic nucleotidic substitution models

- A DNA sequence η is an element of $\{A, T, C, G\}^N$, $N \in \mathbb{N}^*$.
- **Independent** evolution of the sites according to a **Markovian kernel**.

Example: Jukes and Cantor model (1969)

	A	T	C	G
A	$\cdot$	λ	λ	λ
T	λ	$\cdot$	λ	λ
C	λ	λ	$\cdot$	λ
G	λ	λ	λ	$\cdot$

- **Diagonal** entry $-q_{aa}$ is the substitution rate of nucleotide a, here $q_{aa} = -3\lambda$.
- **Non-diagonal** entry q_{ab} is the substitution rate of nucleotide a by b, here $q_{ab} = \lambda$.



δ-marks distributed according to a Poisson process with rate 4λ. The nucleotide is substituted by a with probability $1/4$.

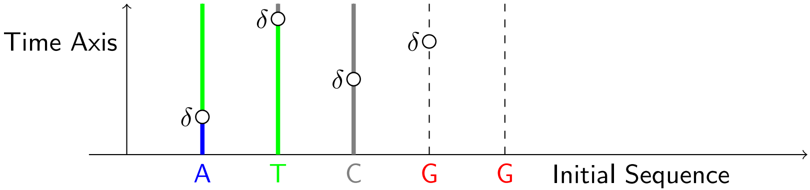
Stochastic nucleotidic substitution models

- A DNA sequence η is an element of $\{\mathrm{A}, \mathrm{T}, \mathrm{C}, \mathrm{G}\}^N$, $N \in \mathbb{N}^*$.
- **Independent** evolution of the sites according to a **Markovian kernel**.

Example: Jukes and Cantor model (1969)

	A	T	C	G
A	$\cdot$	λ	λ	λ
T	λ	$\cdot$	λ	λ
C	λ	λ	$\cdot$	λ
G	λ	λ	λ	$\cdot$

- **Diagonal** entry $-q_{aa}$ is the substitution rate of nucleotide a, here $q_{aa} = -3\lambda$.
- **Non-diagonal** entry q_{ab} is the substitution rate of nucleotide a by b, here $q_{ab} = \lambda$.



δ-marks distributed according to a Poisson process with rate 4λ. The nucleotide is substituted by a with probability $1/4$.

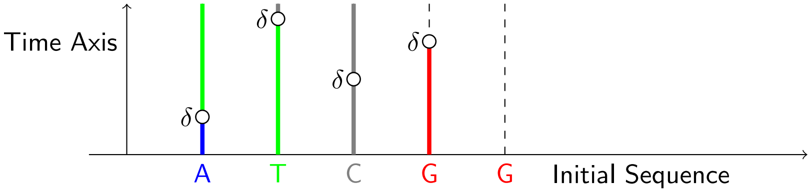
Stochastic nucleotidic substitution models

- A DNA sequence η is an element of $\{A, T, C, G\}^N$, $N \in \mathbb{N}^*$.
- **Independent** evolution of the sites according to a **Markovian kernel**.

Example: Jukes and Cantor model (1969)

	A	T	C	G
A	$\cdot$	λ	λ	λ
T	λ	$\cdot$	λ	λ
C	λ	λ	$\cdot$	λ
G	λ	λ	λ	$\cdot$

- **Diagonal** entry $-q_{aa}$ is the substitution rate of nucleotide a, here $q_{aa} = -3\lambda$.
- **Non-diagonal** entry q_{ab} is the substitution rate of nucleotide a by b, here $q_{ab} = \lambda$.



δ-marks distributed according to a Poisson process with rate 4λ. The nucleotide is substituted by a with probability $1/4$.

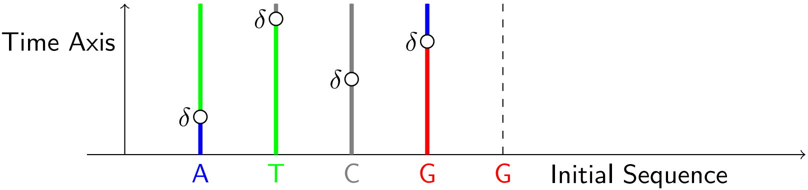
Stochastic nucleotidic substitution models

- A DNA sequence η is an element of $\{\mathsf{A}, \mathsf{T}, \mathsf{C}, \mathsf{G}\}^N$, $N \in \mathbb{N}^*$.
- **Independent** evolution of the sites according to a **Markovian kernel**.

Example: Jukes and Cantor model (1969)

	A	T	C	G
A	$\cdot$	λ	λ	λ
T	λ	$\cdot$	λ	λ
C	λ	λ	$\cdot$	λ
G	λ	λ	λ	$\cdot$

- **Diagonal** entry $-q_{aa}$ is the substitution rate of nucleotide a, here $q_{aa} = -3\lambda$.
- **Non-diagonal** entry q_{ab} is the substitution rate of nucleotide a by b, here $q_{ab} = \lambda$.



δ-marks distributed according to a Poisson process with rate 4λ. The nucleotide is substituted by a with probability $1/4$.

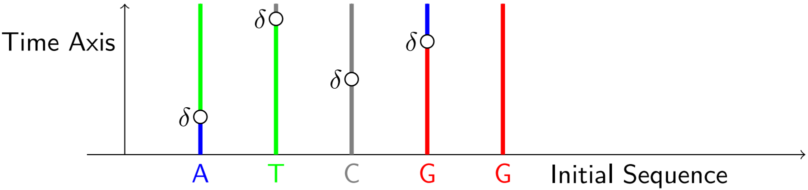
Stochastic nucleotidic substitution models

- A DNA sequence η is an element of $\{A, T, C, G\}^N$, $N \in \mathbb{N}^*$.
- **Independent** evolution of the sites according to a **Markovian kernel**.

Example: Jukes and Cantor model (1969)

	A	T	C	G
A	$\cdot$	λ	λ	λ
T	λ	$\cdot$	λ	λ
C	λ	λ	$\cdot$	λ
G	λ	λ	λ	$\cdot$

- **Diagonal** entry $-q_{aa}$ is the substitution rate of nucleotide a, here $q_{aa} = -3\lambda$.
- **Non-diagonal** entry q_{ab} is the substitution rate of nucleotide a by b, here $q_{ab} = \lambda$.



δ-marks distributed according to a Poisson process with rate 4λ. The nucleotide is substituted by a with probability $1/4$.

Stochastic nucleotidic substitution models

Consequences

- Convergence in distribution at any site

- Convergenge in distribution of the whole sequence to the **product measure**.

Problems

- $(a_1 \dots a_\ell)_{\mathrm{obs}} \neq (a_1)_{\mathrm{obs}} \dots (a_\ell)_{\mathrm{obs}}$.

- The substitution rate $\eta(x) \to a$ may **depend** on $\eta(x-1)$, $\eta(x)$ and $\eta(x+1)$.

Famous example : CpG dinucleotides

- Rate C $\to$ T up to ten times larger when C is involved in a CpG (in fact $\mathrm{C}^\star$pG).

JC+CpG model

Bérard, Gouéré et Piau, *Mathematical Biosciences* (2008)

- A DNA sequence η is now **doubly infinite**, that is, an element of $\{A, T, C, G\}^{\mathbb{Z}}$.

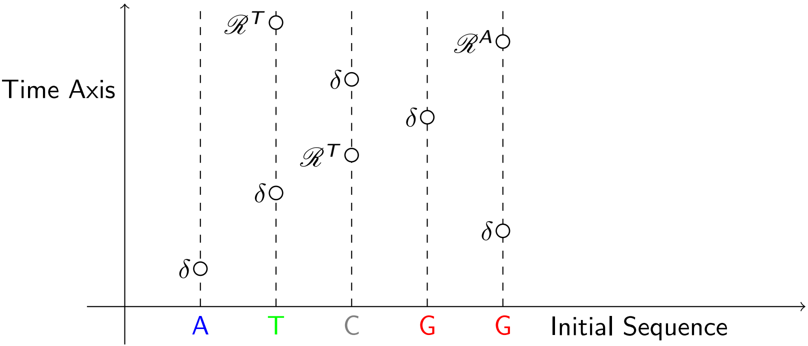
- **Keep** Jukes and Cantor model

	A	T	C	G
A	·	1	1	1
T	1	·	1	1
C	1	1	·	1
G	1	1	1	·

- **Superimpose** "double" substitution mechanism

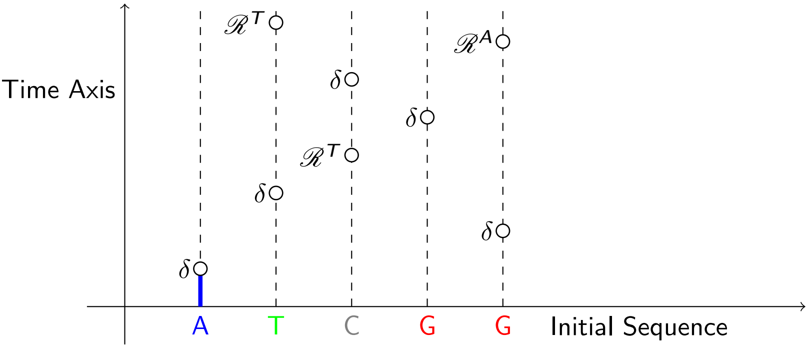
$$\mathrm{CpG} \xrightarrow{r} \mathrm{TpG} \qquad \mathrm{CpG} \xrightarrow{r} \mathrm{CpA}$$

Construction with marked Poisson point processes



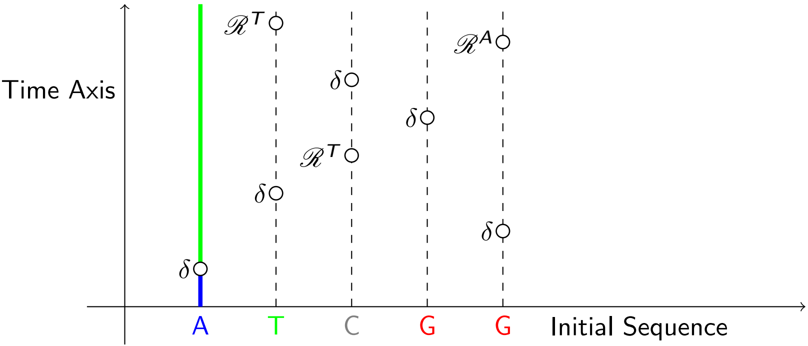
Mark at x	**Rate**	**Action**: $\eta(x)$ moves to a ...
δ	1	... with probability $1/4$.
$\mathscr{R}^a$	r	... if $\eta(x, x+1) =$CG and $a =$T or if $\eta(x-1, x) =$CG and $a =$A.

Construction with marked Poisson point processes



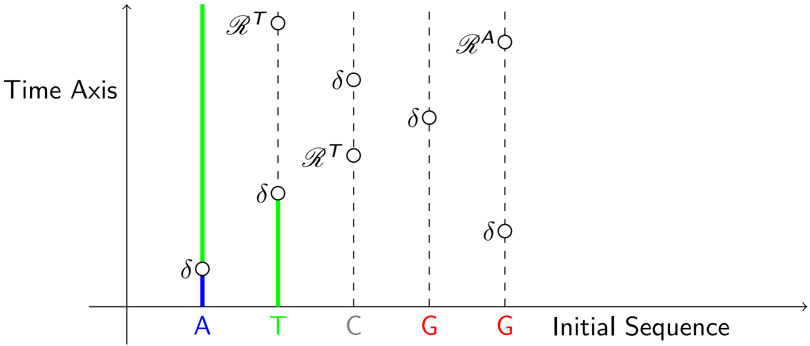
Mark at x	**Rate**	**Action**: $\eta(x)$ moves to a ...
δ	1	... with probability $1/4$.
$\mathscr{R}^a$	r	... if $\eta(x,x+1)=$CG and $a=$T or if $\eta(x-1,x)=$CG and $a=$A.

Construction with marked Poisson point processes



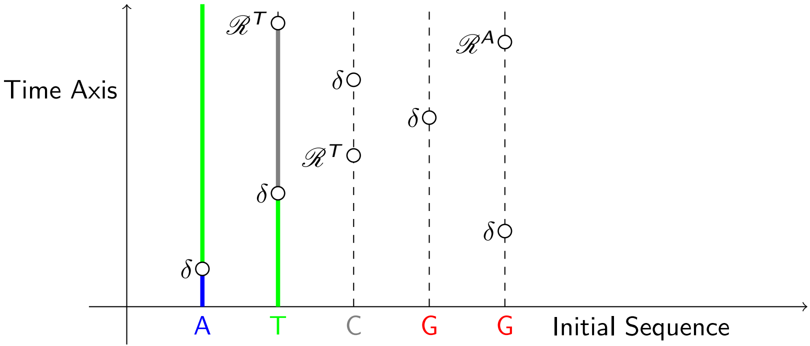
Mark at x	**Rate**	**Action**: $\eta(x)$ moves to a ...
δ	1	... with probability $1/4$.
$\mathscr{R}^a$	r	... if $\eta(x, x+1) =$CG and $a =$T or if $\eta(x-1, x) =$CG and $a =$A.

Construction with marked Poisson point processes



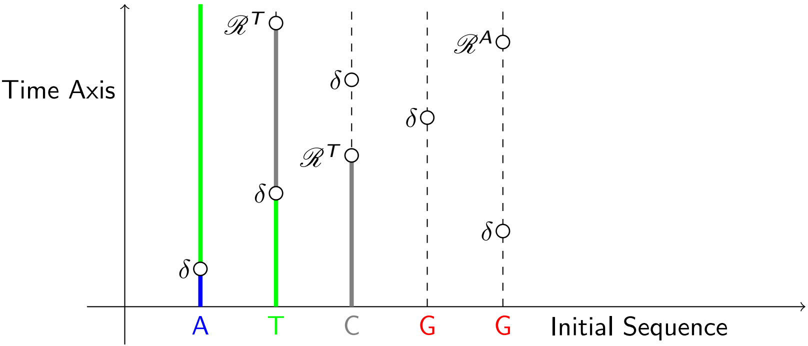
Mark at x	**Rate**	**Action**: $\eta(x)$ moves to a ...
δ	1	... with probability $1/4$.
$\mathscr{R}^a$	r	... if $\eta(x, x+1) =$CG and $a =$T or if $\eta(x-1, x) =$CG and $a =$A.

Construction with marked Poisson point processes



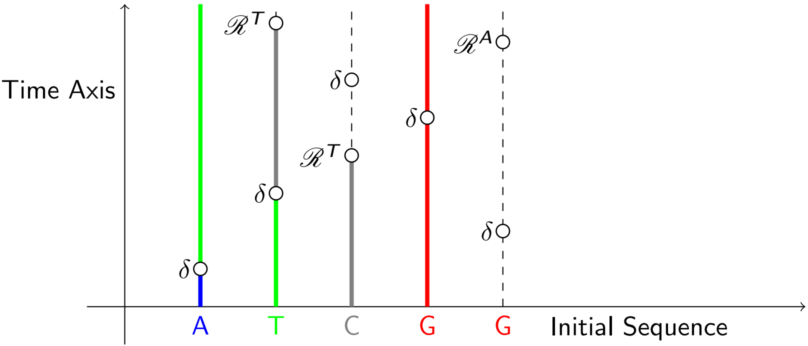
Mark at x	**Rate**	**Action**: $\eta(x)$ moves to a ...
δ	1	... with probability $1/4$.
$\mathscr{R}^a$	r	... if $\eta(x, x+1) =$CG and $a =$T or if $\eta(x-1, x) =$CG and $a =$A.

Construction with marked Poisson point processes



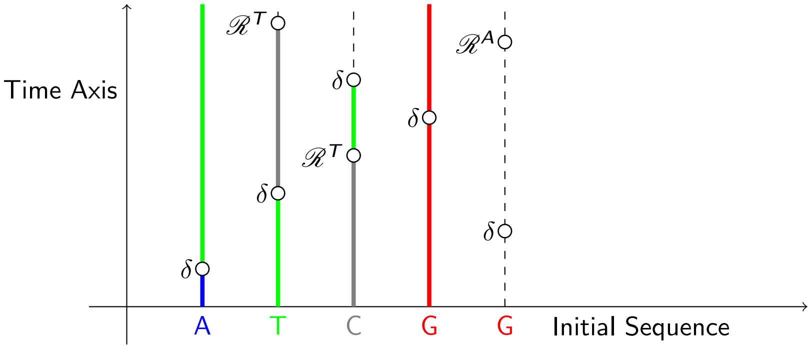
Mark at x	**Rate**	**Action**: $\eta(x)$ moves to a ...
δ	1	... with probability $1/4$.
$\mathscr{R}^a$	r	... if $\eta(x, x+1) =$CG and $a =$T or if $\eta(x-1, x) =$CG and $a =$A.

Construction with marked Poisson point processes



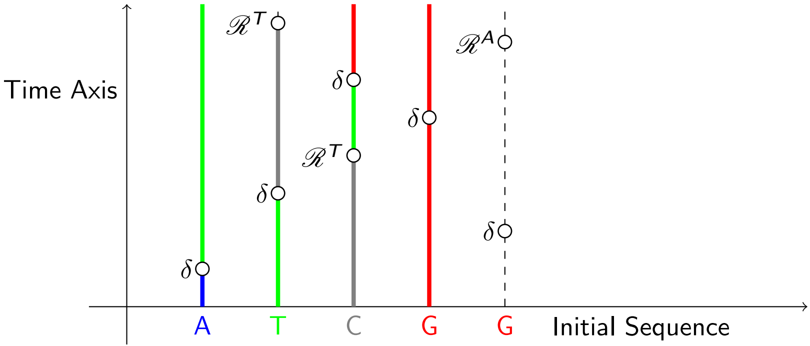
Mark at x	**Rate**	**Action**: $\eta(x)$ moves to a ...
δ	1	... with probability $1/4$.
$\mathscr{R}^a$	r	... if $\eta(x, x+1) =$CG and $a =$T or if $\eta(x-1, x) =$CG and $a =$A.

Construction with marked Poisson point processes



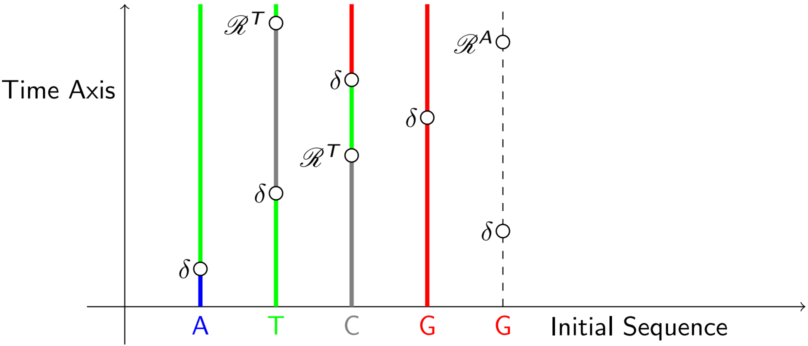
Mark at x	**Rate**	**Action**: $\eta(x)$ moves to a ...
δ	1	... with probability $1/4$.
$\mathscr{R}^a$	r	... if $\eta(x, x+1) =$CG and $a =$T or if $\eta(x-1, x) =$CG and $a =$A.

Construction with marked Poisson point processes



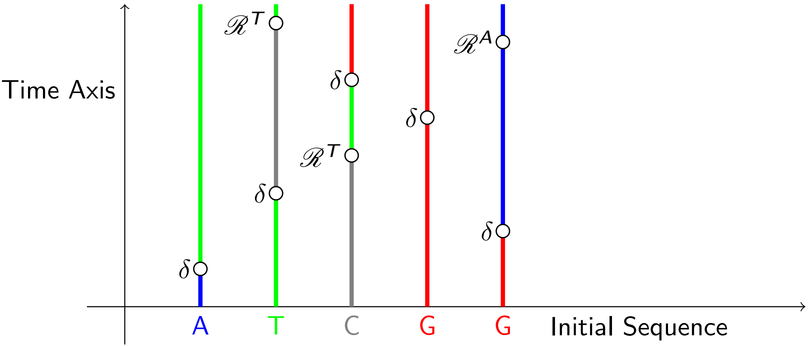
Mark at x	**Rate**	**Action**: $\eta(x)$ moves to a ...
δ	1	... with probability $1/4$.
$\mathscr{R}^a$	r	... if $\eta(x, x+1) =$CG and $a =$T or if $\eta(x-1, x) =$CG and $a =$A.

Construction with marked Poisson point processes



Mark at x	**Rate**	**Action**: $\eta(x)$ moves to a ...
δ	1	... with probability $1/4$.
$\mathscr{R}^a$	r	... if $\eta(x, x+1) =$CG and $a =$T or if $\eta(x-1, x) =$CG and $a =$A.

Construction with marked Poisson point processes



Mark at x	**Rate**	**Action**: $\eta(x)$ moves to a ...
δ	1	... with probability $1/4$.
$\mathscr{R}^a$	r	... if $\eta(x,x+1) =$CG and $a =$T or if $\eta(x-1,x) =$CG and $a =$A.

Properties

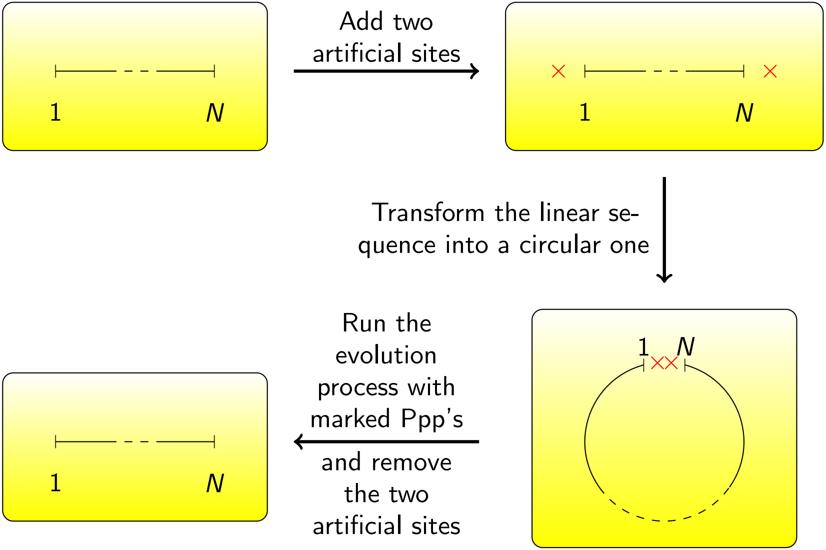
Bérard, Gouéré et Piau, *Mathematical Biosciences* (2008)

- **There exists a unique Markov process** on $\mathscr{A}^{\mathbb{Z}}$ with the transition rates defined before.

- The process is **ergodic**, its unique invariant probability measure π on $\mathscr{A}^{\mathbb{Z}}$ is **translation invariant** and **ergodic** with respect to the translations on $\mathbb{Z}$.

- Any collections $(\eta_x)_{x \in I}$ and $(\eta_y)_{y \in J}$ are **independent** as soon as $\mathrm{dist}(I, J) \geqslant 3$.

Simulate the evolution of a finite DNA sequence



Substitution processes

- Only one coordinate changes in each transition.

- The transition mechanism is specified by a non-negative function c defined on $\mathbb{Z} \times \mathscr{A} \times X$, with $\mathscr{A}$ a finite alphabet and $X = \mathscr{A}^{\mathbb{Z}}$.

- We assume that, for any fixed site x and target a, the function $c(x, a, \cdot)$ depends on $\eta \in X$ only through a finite set $S_x^a \subset \mathbb{Z}$ depending on x and a.

Sufficient conditions for its existence

$$K = \sup\{c(x, a, \eta) : x \in \mathbb{Z}, a \in \mathscr{A}, \eta \in X\} < \infty \tag{1}$$

$$\text{and} \quad s = \sup\{|S_x^a| : x \in \mathbb{Z}, a \in \mathscr{A}\} < \infty. \tag{2}$$

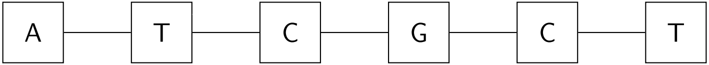
Example: JC+CpG

$$S_x^A = \{x-1, x\} \quad \text{and} \quad c(x, A, \eta) = \begin{cases} 1+r & \text{if } \eta(x-1, x) = CG, \\ 1 & \text{else,} \end{cases}$$

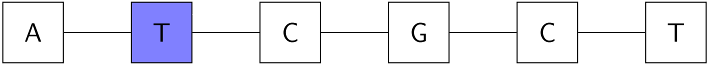
$$S_x^T = \{x, x+1\} \quad \text{and} \quad c(x, T, \eta) = \begin{cases} 1+r & \text{if } \eta(x, x+1) = CG, \\ 1 & \text{else,} \end{cases}$$

$$S_x^C = \emptyset \quad \text{and} \quad c(x, C, \eta) = 1,$$
$$S_x^G = \emptyset \quad \text{and} \quad c(x, G, \eta) = 1.$$

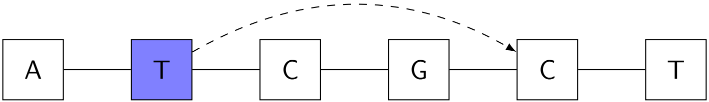
Example of single ‘cut-and-paste’



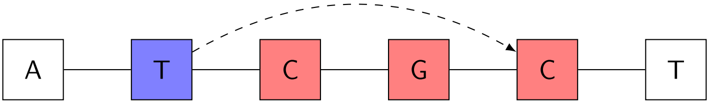
Example of single ‘cut-and-paste’



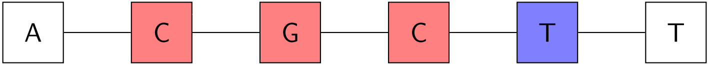
Example of single 'cut-and-paste'



Example of single 'cut-and-paste'

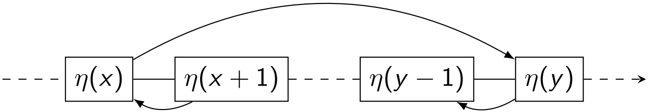


Example of single ‘cut-and-paste’



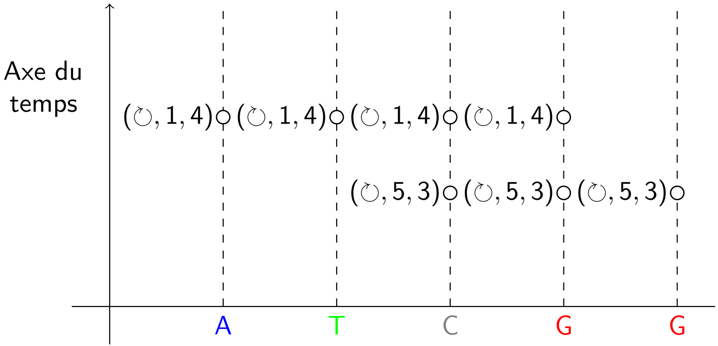
'Cut-and-paste' process

- The transition mechanisms are **circular permutations** of finitely many sites of $\mathbb{Z}$ and are specified by a **transition probability matrix** p on $\mathbb{Z} \times \mathbb{Z}$ and a **cut rate per site** $\rho \geqslant 0$.
- The value $\rho \cdot p(x, y)$ represents the rate at which the coordinate $\eta(x)$ is transfered to site y. We assume that p is translation invariant on $\mathbb{Z}$.



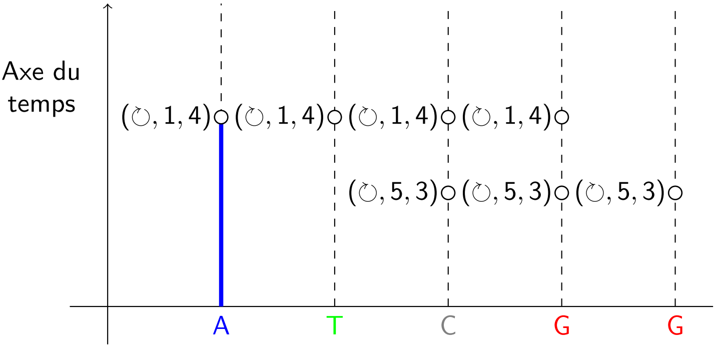
Sufficient condition for its existence: $\sum |x| p(0, x) < \infty$

Construction with marked Poisson point processes



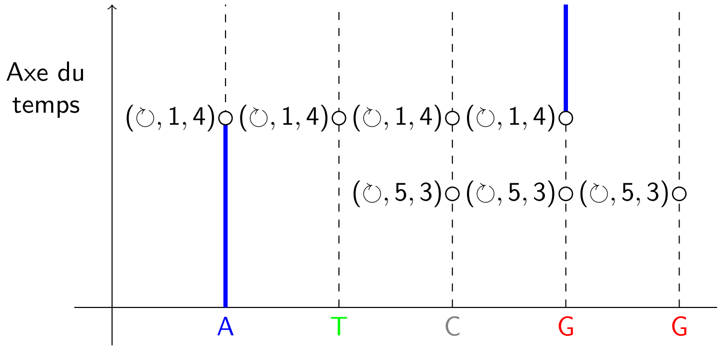
- Mark $(\circlearrowright, x, y)$ distributed at rate $\rho \cdot p(x, y)$. If $x < y$, the contents of sites x, $x + 1$, ..., y are right circularly permuted. If $x > y$, the contents of sites y, $y + 1$, ..., x are left circularly permuted.

Construction with marked Poisson point processes



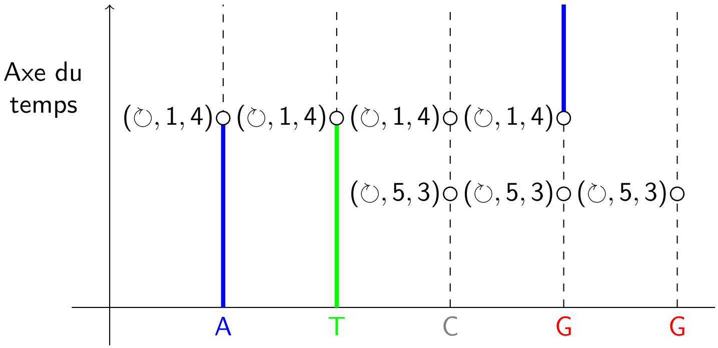
- Mark $(\circlearrowright, x, y)$ distributed at rate $\rho \cdot p(x, y)$. If $x < y$, the contents of sites x, $x+1$, $\dots$, y are right circularly permuted. If $x > y$, the contents of sites y, $y+1$, $\dots$, x are left circularly permuted.

Construction with marked Poisson point processes



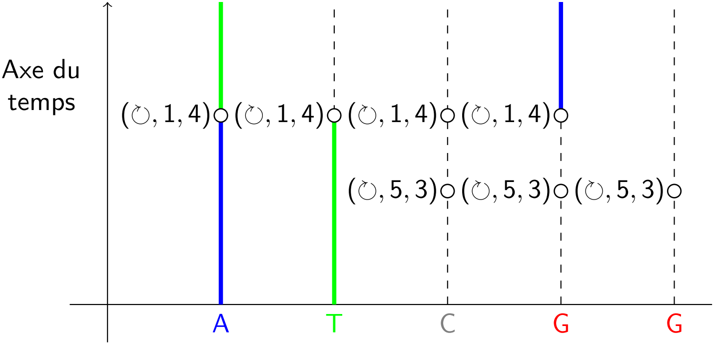
- Mark $(\circlearrowright, x, y)$ distributed at rate $\rho \cdot p(x, y)$. If $x < y$, the contents of sites x, $x + 1$, ..., y are right circularly permuted. If $x > y$, the contents of sites y, $y + 1$, ..., x are left circularly permuted.

Construction with marked Poisson point processes



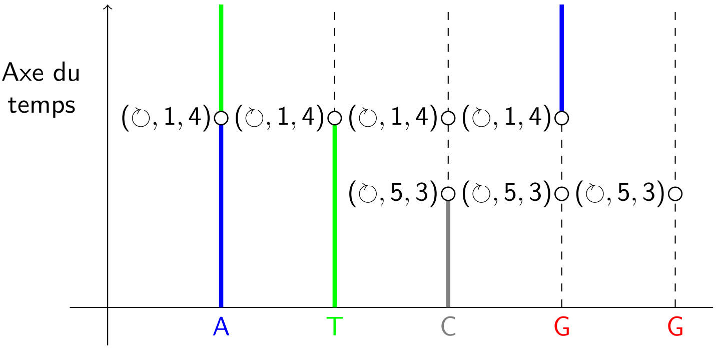
- Mark $(\circlearrowright, x, y)$ distributed at rate $\rho \cdot p(x, y)$. If $x < y$, the contents of sites x, $x + 1$, ..., y are right circularly permuted. If $x > y$, the contents of sites y, $y + 1$, ..., x are left circularly permuted.

Construction with marked Poisson point processes



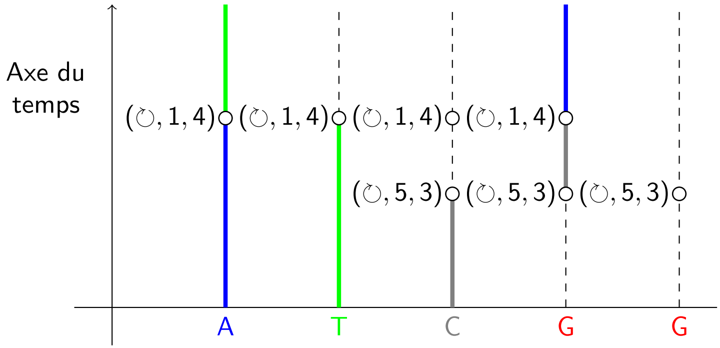
- Mark $(\circlearrowright, x, y)$ distributed at rate $\rho \cdot p(x, y)$. If $x < y$, the contents of sites x, $x+1$, ..., y are right circularly permuted. If $x > y$, the contents of sites y, $y+1$, ..., x are left circularly permuted.

Construction with marked Poisson point processes



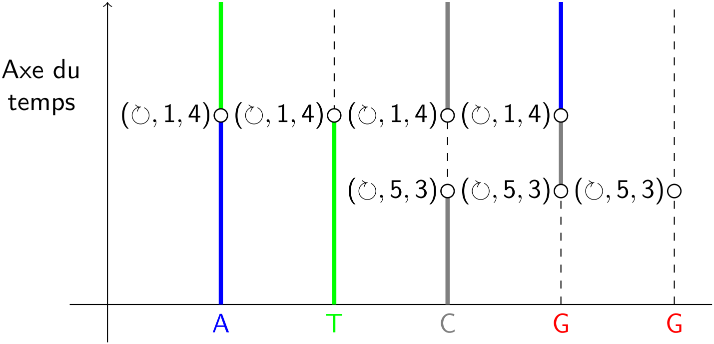
- Mark $(\circlearrowright, x, y)$ distributed at rate $\rho \cdot p(x, y)$. If $x < y$, the contents of sites x, $x + 1$, ..., y are right circularly permuted. If $x > y$, the contents of sites y, $y + 1$, ..., x are left circularly permuted.

Construction with marked Poisson point processes



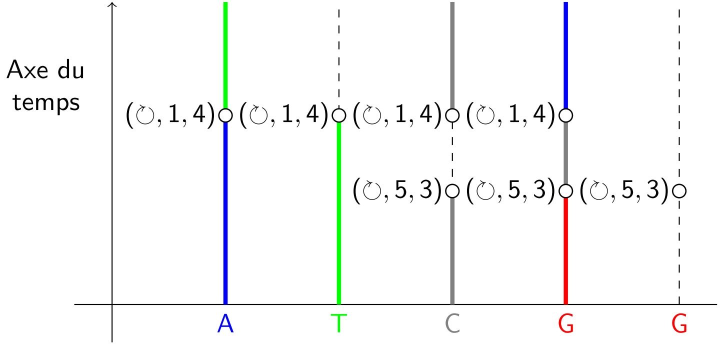
- Mark $(\circlearrowright, x, y)$ distributed at rate $\rho \cdot p(x, y)$. If $x < y$, the contents of sites x, $x+1$, ..., y are right circularly permuted. If $x > y$, the contents of sites y, $y+1$, ..., x are left circularly permuted.

Construction with marked Poisson point processes



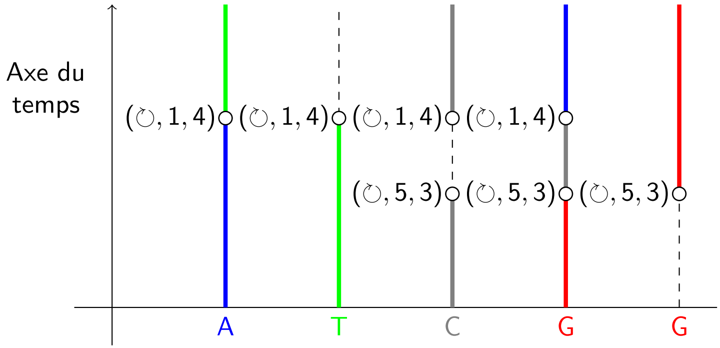
- Mark $(\circlearrowright, x, y)$ distributed at rate $\rho \cdot p(x, y)$. If $x < y$, the contents of sites x, $x + 1$, ..., y are right circularly permuted. If $x > y$, the contents of sites y, $y + 1$, ..., x are left circularly permuted.

Construction with marked Poisson point processes



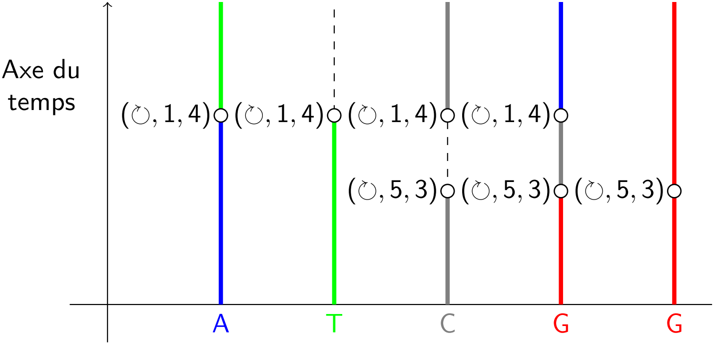
- Mark $(\circlearrowright, x, y)$ distributed at rate $\rho \cdot p(x, y)$. If $x < y$, the contents of sites x, $x+1$, …, y are right circularly permuted. If $x > y$, the contents of sites y, $y+1$, …, x are left circularly permuted.

Construction with marked Poisson point processes



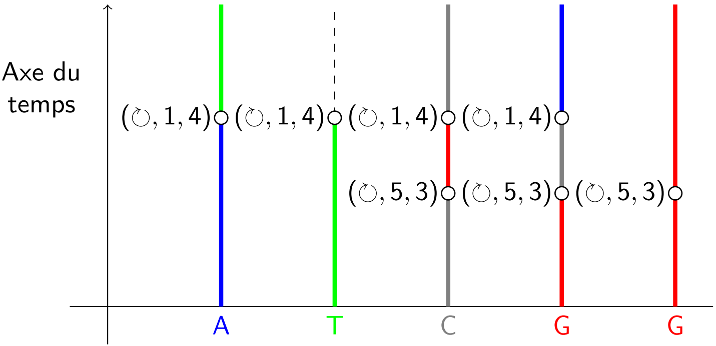
- Mark $(\circlearrowright, x, y)$ distributed at rate $\rho \cdot p(x, y)$. If $x < y$, the contents of sites x, $x+1$, ..., y are right circularly permuted. If $x > y$, the contents of sites y, $y+1$, ..., x are left circularly permuted.

Construction with marked Poisson point processes



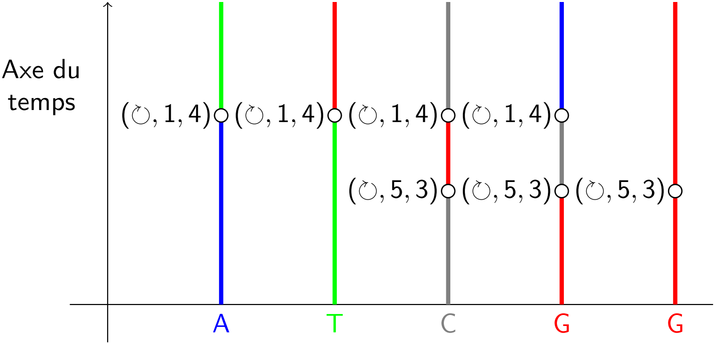
- Mark $(\circlearrowright, x, y)$ distributed at rate $\rho \cdot p(x, y)$. If $x < y$, the contents of sites x, $x+1$, $\ldots$, y are right circularly permuted. If $x > y$, the contents of sites y, $y+1$, $\ldots$, x are left circularly permuted.

Construction with marked Poisson point processes



- Mark $(\circlearrowright, x, y)$ distributed at rate $\rho \cdot p(x, y)$. If $x < y$, the contents of sites x, $x+1$, ..., y are right circularly permuted. If $x > y$, the contents of sites y, $y+1$, ..., x are left circularly permuted.

Construction with marked Poisson point processes



- Mark $(\circlearrowright, x, y)$ distributed at rate $\rho \cdot p(x, y)$. If $x < y$, the contents of sites x, $x + 1$, ..., y are right circularly permuted. If $x > y$, the contents of sites y, $y + 1$, ..., x are left circularly permuted.

Ergodicity for independent evolution models with 'cut-and-paste'

Theorem (Falconnet, Gantert and Saada)

Assume that the substitution rates are **independent** and are ruled by an irreducible Q-matrix. Then the process is **ergodic** and the invariant measure is the product measure on $\mathbb{Z}$.

- Especially, for any usual substitution model (JC69, K80, T92, etc.) and any 'cut-and-paste' mechanism invariant by translation, the dynamic of the process is ergodic.

Ergodicity for substitution processes with ‘cut-and-paste’ mechanism

Theorem (Falconnet, Gantert and Saada)

Define

$$\begin{aligned} m &= \inf\{c(x,a,\eta) : x \in \mathbb{Z}, a \in \mathscr{A}, \eta \in X\}, \\ K &= \sup\{c(x,a,\eta) : x \in \mathbb{Z}, a \in \mathscr{A}, \eta \in X\}, \\ s &= \max_{x\in\mathbb{Z}, a\in\mathscr{A}} |S_x^a|. \end{aligned} \tag{3}$$

Assume that

$$m > 0 \quad \text{and} \quad (s-1)(K-m) < |\mathscr{A}|m.$$

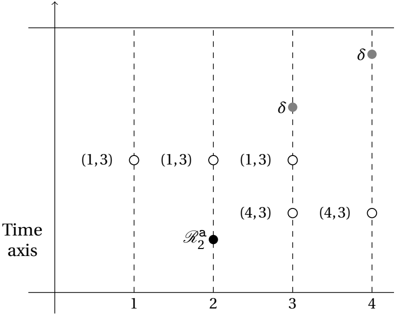
Then, for any $\rho \geqslant 0$, the superimposition of a substitution process and a ‘cut-and-paste’ mechanism is **exponentially ergodic**.

Especially, the JC+CpG+‘cut-and-paste’ model is ergodic as soon as

$$r < 4\lambda.$$

Method : construction of a generalized dual process. Inspired from **Ferrari, Annals of probability (1990)**.

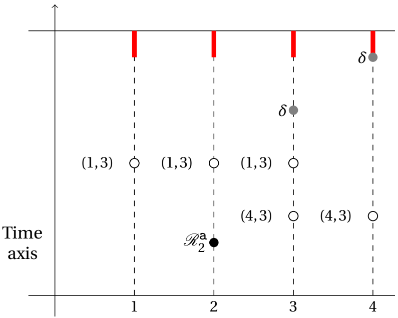
Introduction of a marked branching structure



When we go back in time and the dual process meets

- a δ-mark at site x, the branch **dies**;
- a $\mathscr{R}^a$-mark at site x, the branch dies and **is replaced** by 2 new branches.
- a (x, y)-mark, the involved branches are **circularly permuted**.

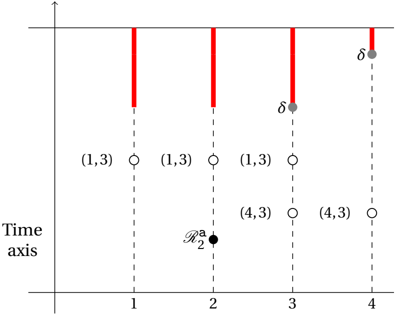
Introduction of a marked branching structure



When we go back in time and the dual process meets

- a δ-mark at site x, the branch **dies**;
- a $\mathscr{R}^a$-mark at site x, the branch dies and **is replaced** by 2 new branches.
- a (x, y)-mark, the involved branches are **circularly permuted**.

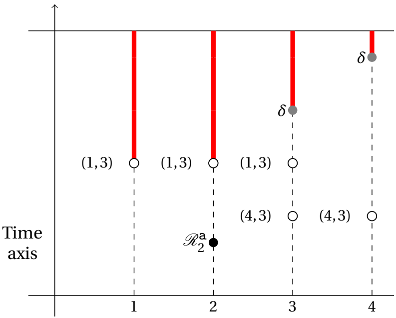
Introduction of a marked branching structure



When we go back in time and the dual process meets

- a δ-mark at site x, the branch **dies**;
- a $\mathscr{R}^a$-mark at site x, the branch dies and **is replaced** by 2 new branches.
- a (x, y)-mark, the involved branches are **circularly permuted**.

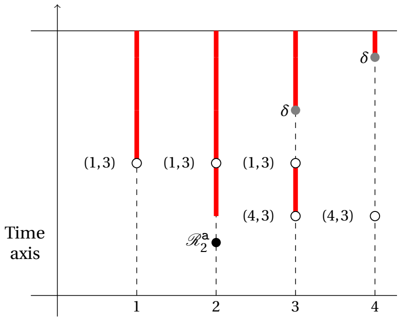
Introduction of a marked branching structure



When we go back in time and the dual process meets

- a δ-mark at site x, the branch **dies**;
- a $\mathscr{R}^{a}$-mark at site x, the branch dies and **is replaced** by 2 new branches.
- a (x, y)-mark, the involved branches are **circularly permuted**.

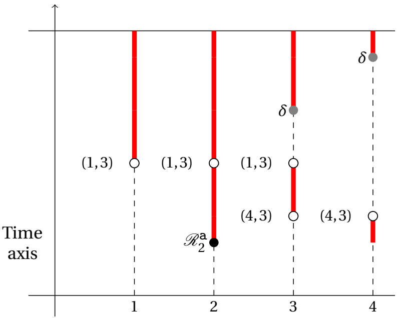
Introduction of a marked branching structure



When we go back in time and the dual process meets

- a δ-mark at site x, the branch **dies**;
- a $\mathscr{R}^a$-mark at site x, the branch dies and **is replaced** by 2 new branches.
- a (x, y)-mark, the involved branches are **circularly permuted**.

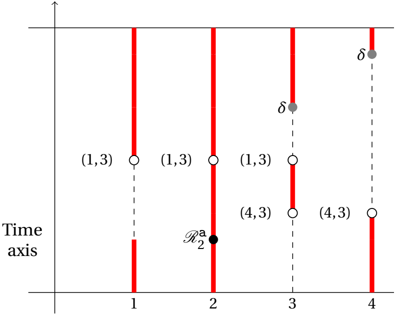
Introduction of a marked branching structure



When we go back in time and the dual process meets

- a δ-mark at site x, the branch **dies**;
- a $\mathscr{R}^a$-mark at site x, the branch dies and **is replaced** by 2 new branches.
- a (x, y)-mark, the involved branches are **circularly permuted**.

Introduction of a marked branching structure



When we go back in time and the dual process meets

- a δ-mark at site x, the branch **dies**;

- a $\mathscr{R}^a$-mark at site x, the branch dies and **is replaced** by 2 new branches.

- a (x, y)-mark, the involved branches are **circularly permuted**.

Central idea

- When we go back in time and the dual process meets a δ-mark at site x, it is not necessary to go further to know the value of $\eta(x)$, because it is determined at that point by an independent random variable.
- Conditions are obtained by coupling with a dying branching process.

Literature

J. Bérard, J.-B. Gouéré, and D. Piau.
Solvable models of neighbor-dependent nucleotide substitution processes.
Mathematical Biosciences, 211:56–88, 2008.

P. A. Ferrari.
Ergodicity for spin systems with stirrings.
The Annals of Probability, 18(4):1523–1538, 1990.

T. H. Jukes and C. R. Cantor.
Mammalian protein metabolism, chapter Evolution of Protein Molecules, pages 21–132.
Academic Press, New York, 1969.

T. M. Liggett.
Interacting Particle System, volume 276 of *Grundlehren der Mathematischen Wissenschaften*.
Springer-Verlag, New York, 1985.