Evolution of the Ancestral Recombination Graph along the genome in case of selective sweep

Stéphanie LEOCARD

LATP, Université de Provence, Marseille, France.

March 27, 2009

Plan

1. Introduction
   - Selective sweep with strong selective advantage
   - Recombinations
2. Evolution of the coalescent tree along the genome
   - Description of the coalescent tree
   - Evolution of the tree
3. Evolution of the ARG along the genome
   - Evolution under neutrality
   - Evolution in case of selective sweep
   - Back to the process of coalescent trees

Plan

1. Introduction
   - Selective sweep with strong selective advantage
   - Recombinations
2. Evolution of the coalescent tree along the genome
   - Description of the coalescent tree
   - Evolution of the tree
3. Evolution of the ARG along the genome
   - Evolution under neutrality
   - Evolution in case of selective sweep
   - Back to the process of coalescent trees

Plan

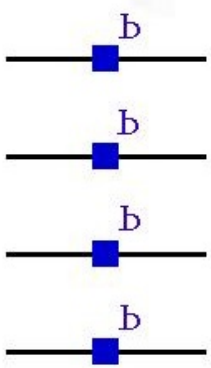
1. Introduction
   - Selective sweep with strong selective advantage
   - Recombinations
2. Evolution of the coalescent tree along the genome
   - Description of the coalescent tree
   - Evolution of the tree
3. Evolution of the ARG along the genome
   - Evolution under neutrality
   - Evolution in case of selective sweep
   - Back to the process of coalescent trees

Haploid population of infinite size.
Genome considered as a single chromosome, identified to $\mathbb{R}$.

Haploid population of infinite size.
Genome considered as a single chromosome, identified to $\mathbb{R}$.

We assume that the gene at locus 0 is under selective sweep :

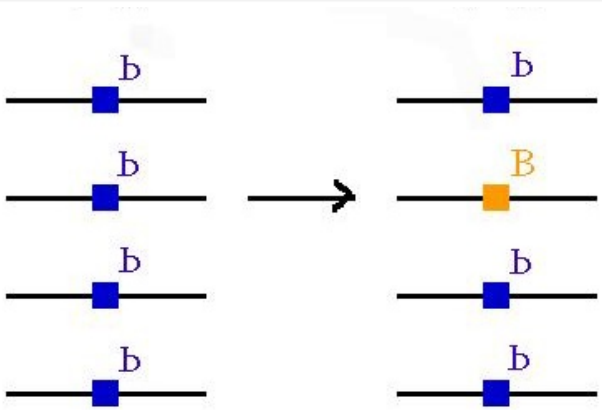
- Before the sweep, everybody carries the wild-type allele b at locus 0.



Haploid population of infinite size.
Genome considered as a single chromosome, identified to $\mathbb{R}$.

We assume that the gene at locus 0 is under selective sweep :

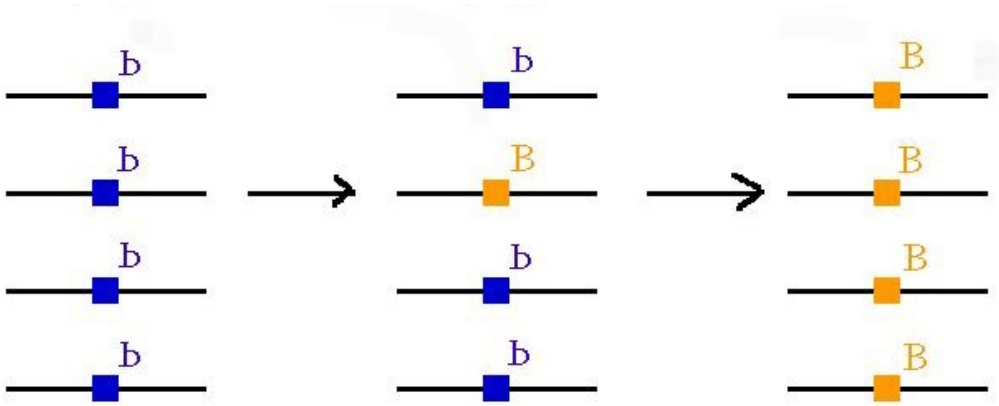
- Before the sweep, everybody carries the wild-type allele b at locus 0.
- An advantageous mutation from b to B happens in a SINGLE individual J. Selective advantage of B over b: α.



Haploid population of infinite size.
Genome considered as a single chromosome, identified to $\mathbb{R}$.

We assume that the gene at locus 0 is under selective sweep :

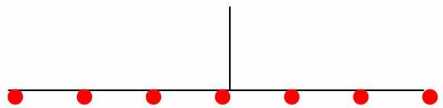
- Before the sweep, everybody carries the wild-type allele b at locus 0.
- An advantageous mutation from b to B happens in a SINGLE individual J. Selective advantage of B over b: α.
- B quickly spreads in the population until fixation.



Assume $\alpha \to +\infty$.
The duration of the selective sweep tends to 0 as the selective advantage α goes to $+\infty$, so the selective sweep is instantaneous.

Assume $\alpha \to +\infty$.
The duration of the selective sweep tends to 0 as the selective advantage α goes to $+\infty$, so the selective sweep is instantaneous.

Sample of n individuals at the end of the selective sweep.
Genealogy at the site under selection:

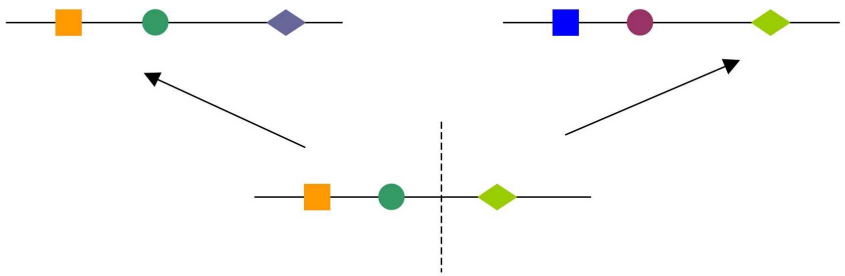


Plan

1. Introduction
   - Selective sweep with strong selective advantage
   - Recombinations
2. Evolution of the coalescent tree along the genome
   - Description of the coalescent tree
   - Evolution of the tree
3. Evolution of the ARG along the genome
   - Evolution under neutrality
   - Evolution in case of selective sweep
   - Back to the process of coalescent trees

Definition

A recombination happens when a portion of the genome is inherited from an individual and the other portion by another individual.



Consider a neutral site at position $x > 0$.

- If there is no recombination on $[0, x]$ in the sample, same genealogy as the site under selection : comb. (hitchhiking)
- If recombinations occur on $[0, x]$, no comb anymore.

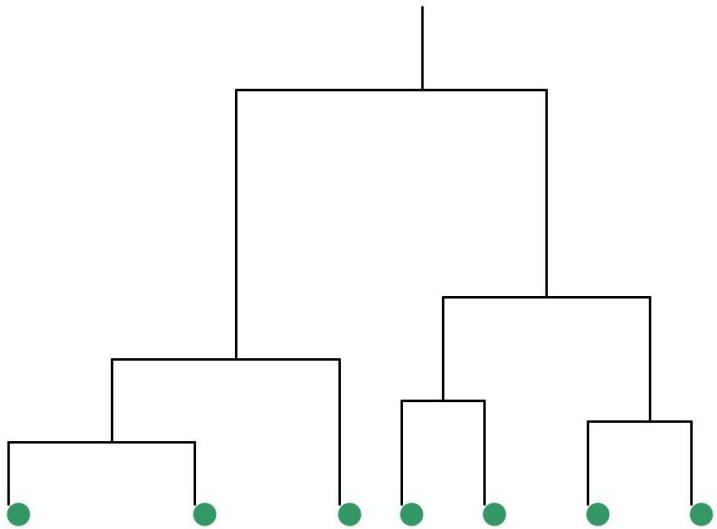
Far from the site under selection (x large):

Many recombinations on $[0, x]$
$\Rightarrow$ No hitchhiking anymore
$\Rightarrow$ Evolution under neutrality.

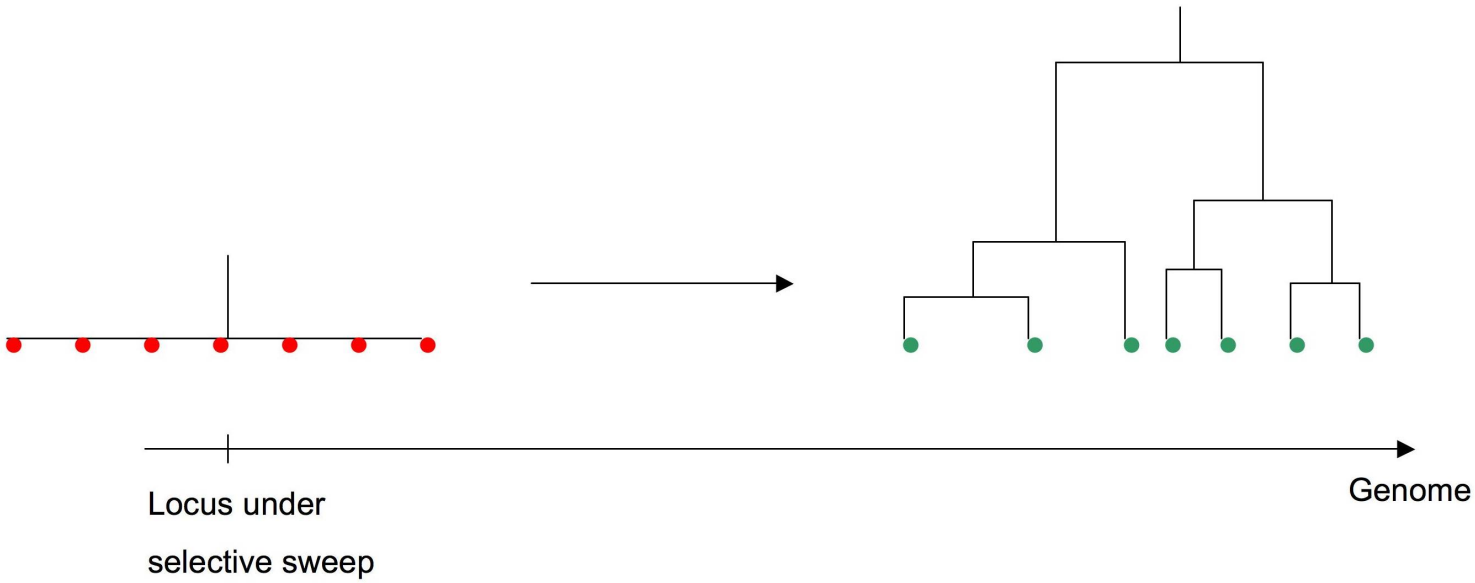
Far from the site under selection (x large):

Many recombinations on $[0, x]$
$\Rightarrow$ No hitchhiking anymore
$\Rightarrow$ Evolution under neutrality.

Genealogy of the n-sample: Kingman n-coalescent.
Coalescence rate when k lineages: $\binom{k}{2}$.



Objective: Study of the coalescent tree as the distance from the selected site increases.
Restriction to $[0, +\infty)$. Description of the evolution from the comb to the Kingman coalescent.



Plan

1. Introduction
   - Selective sweep with strong selective advantage
   - Recombinations
2. Evolution of the coalescent tree along the genome
   - Description of the coalescent tree
   - Evolution of the tree
3. Evolution of the ARG along the genome
   - Evolution under neutrality
   - Evolution in case of selective sweep
   - Back to the process of coalescent trees

Plan

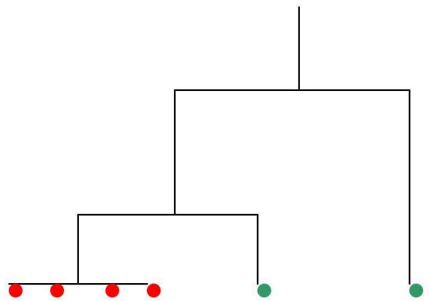
1. Introduction
   - Selective sweep with strong selective advantage
   - Recombinations
2. Evolution of the coalescent tree along the genome
   - Description of the coalescent tree
   - Evolution of the tree
3. Evolution of the ARG along the genome
   - Evolution under neutrality
   - Evolution in case of selective sweep
   - Back to the process of coalescent trees

Let $x > 0$ be a position on the genome.
$R(x) =$ number of alleles at position x in the n-sample, that are inherited from J.

Let $x > 0$ be a position on the genome.
$R(x) =$ number of alleles at position x in the n-sample, that are inherited from J.

Proposition

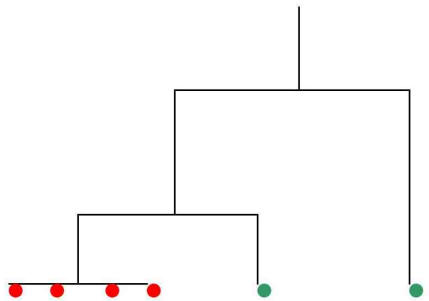
The coalescent tree at locus x is a Kingman $(n - R(x) + 1)$-coalescent, where one leaf is a comb with $R(x)$ teeth. The comb gathers the alleles inherited from J.



Let $x > 0$ be a position on the genome.
$R(x) =$ number of alleles at position x in the n-sample, that are inherited from J.

Proposition

The coalescent tree at locus x is a Kingman $(n - R(x) + 1)$-coalescent, where one leaf is a comb with $R(x)$ teeth. The comb gathers the alleles inherited from J.



Question: Evolution of the process $x \in \mathbb{R}_+ \to R(x) \in \{0, ..., n\}$?

Theorem

The process $x \in \mathbb{R}_+ \longrightarrow R(x) \in \{0, ..., n\}$ *has the following properties:*

1. $R(0) = n$,
2. *R is a non-homogeneous Markov process whose jump rates are given as follows:*

$$Q_{k,\ell}(x) = \begin{cases} (1 - \frac{1}{2}\exp(-\gamma x))k \times 2\gamma & \text{if } \ell = k-1, \\ \frac{1}{2}\exp(-\gamma x)(n-k) \times 2\gamma & \text{if } \ell = k+1, \\ 0 & \text{if } \ell \notin \{k-1, k, k+1\}. \end{cases}$$

Theorem

The process $x \in \mathbb{R}_+ \rightarrow R(x) \in \{0, ..., n\}$ *has the following properties:*

1. $R(0) = n$,
2. *R is a non-homogeneous Markov process whose jump rates are given as follows:*

$$Q_{k,\ell}(x) = \begin{cases} (1 - \frac{1}{2}\exp(-\gamma x))k \times 2\gamma & if\ \ell = k-1, \\ \frac{1}{2}\exp(-\gamma x)(n-k) \times 2\gamma & if\ \ell = k+1, \\ 0 & if\ \ell \notin \{k-1, k, k+1\}. \end{cases}$$

Idea for $Q_{k,k+1}(x)$

Proposition

$\exists x(\omega) > 0; R(x) = 0\ \forall x \geq x(\omega)$ *a.s.*

Plan

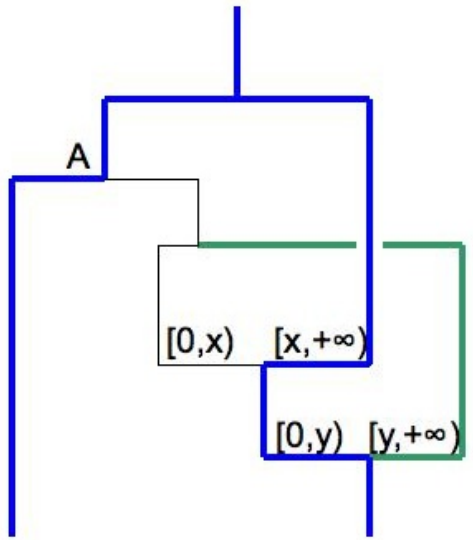
1. Introduction
   - Selective sweep with strong selective advantage
   - Recombinations
2. Evolution of the coalescent tree along the genome
   - Description of the coalescent tree
   - Evolution of the tree
3. Evolution of the ARG along the genome
   - Evolution under neutrality
   - Evolution in case of selective sweep
   - Back to the process of coalescent trees

Wiuf and Hein, 1999: The process of the coalescent trees along the genome is not Markovian.
We add all the "past": Ancestral Recombination Graph (ARG).

Definition

We denote by $ARG(u)$ the graph that sums up the genealogy of the sample, implied by coalescence and recombination events on the portion $[0, u]$.

$0 < x < y$



Plan

1. Introduction
   - Selective sweep with strong selective advantage
   - Recombinations
2. Evolution of the coalescent tree along the genome
   - Description of the coalescent tree
   - Evolution of the tree
3. Evolution of the ARG along the genome
   - Evolution under neutrality
   - Evolution in case of selective sweep
   - Back to the process of coalescent trees

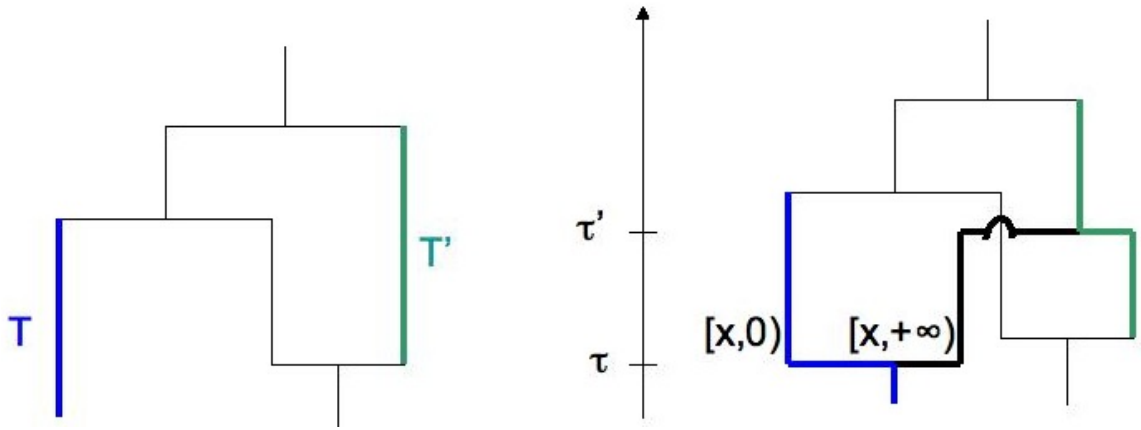
The ARG is modified:

- when a recombination impacts the ARG during the neutral period that predates the selective sweep (rate $\lambda\times$ total length of the ARG)
- when a recombination occurs during the sweep (rate 2γ)

Plan

1. Introduction
   - Selective sweep with strong selective advantage
   - Recombinations
2. Evolution of the coalescent tree along the genome
   - Description of the coalescent tree
   - Evolution of the tree
3. Evolution of the ARG along the genome
   - Evolution under neutrality
   - Evolution in case of selective sweep
   - Back to the process of coalescent trees

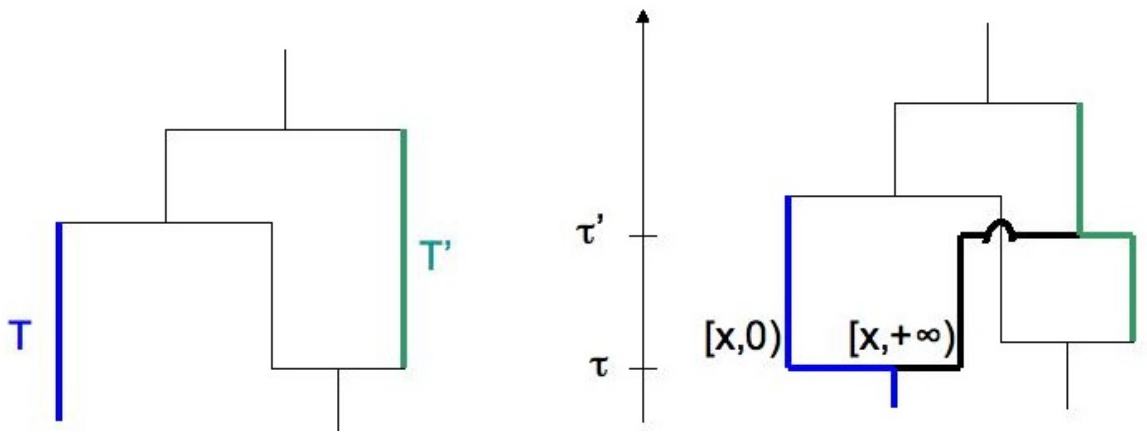
τ: recombination time;
T: branch of the ARG where the recombination happens;
τ': coalescence time of the recombinant lineage;
T': branch of the ARG that coalesces with the recombinant lineage;



$A_t(x)$: number of lineages in ARG(x) at time $t \geq 0$.
$L(x)$: total length of ARG(x) (convention: $L(0) = 1$)
$H(x)$: height of ARG(x) (convention: $H(0) = 1$)

Theorem

1. *The positions of the recombinations are the jump times of a Poisson process with intensity* $\lambda \times L(x)$.
2. $\mathbb{P}_{(\tau,\tau',T,T')}(dt, dt', \mathcal{T}, \mathcal{T}') = \dfrac{\exp\left(-\int_t^{t'} A_s(x^-)ds\right)}{\int_0^{H(x^-)} A_w(x^-)dw} 1_{t<t'} dt' dt.$

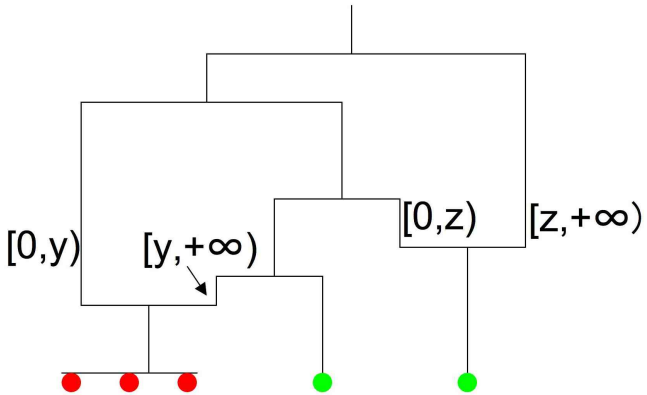


Plan

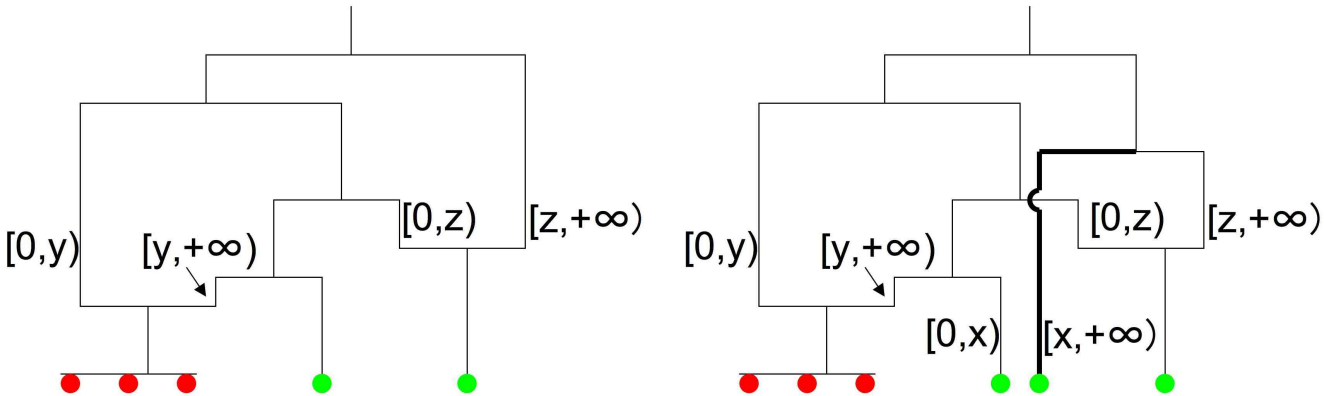
1. Introduction
   - Selective sweep with strong selective advantage
   - Recombinations
2. Evolution of the coalescent tree along the genome
   - Description of the coalescent tree
   - Evolution of the tree
3. Evolution of the ARG along the genome
   - Evolution under neutrality
   - Evolution in case of selective sweep
   - Back to the process of coalescent trees

The ARG is now composed of the "neutral $ARG(x)$", where one of the leaves is a comb $\mathcal{P}(x)$.

Suppose that a recombination happens at position $x > 0$ during the selective sweep and that $ARG(x^-)$ is given. Then the recombination is of one of the four types:

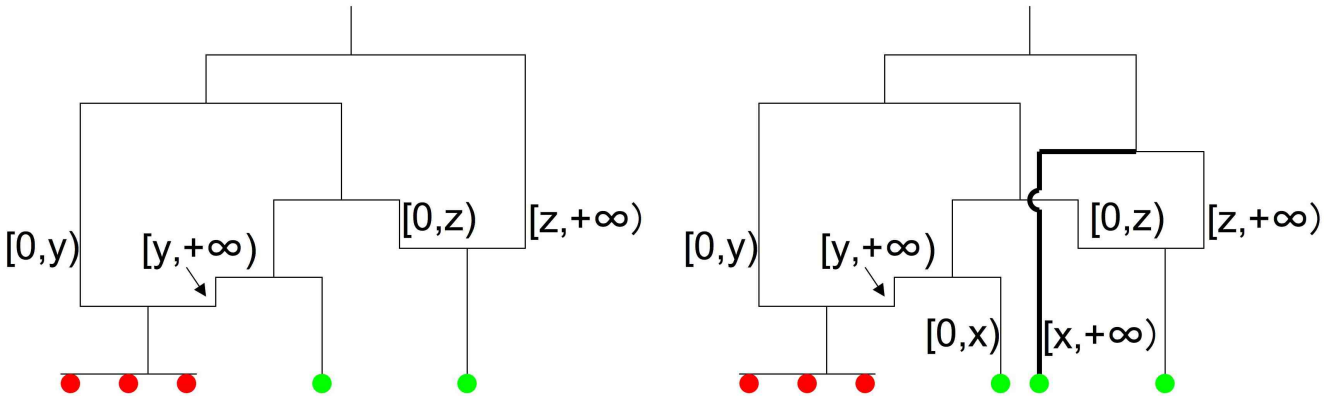


$C_1(x)$:



$$\mathbb{P}_{(\tau',T,T')}(dt',\mathcal{T},\mathcal{T}') = \frac{\exp\left(-\int_0^{t'} A_s(x^-)ds\right)}{A_{0+}(x^-)-1} 1_{t'\in\mathcal{T}'} d\tau'.$$

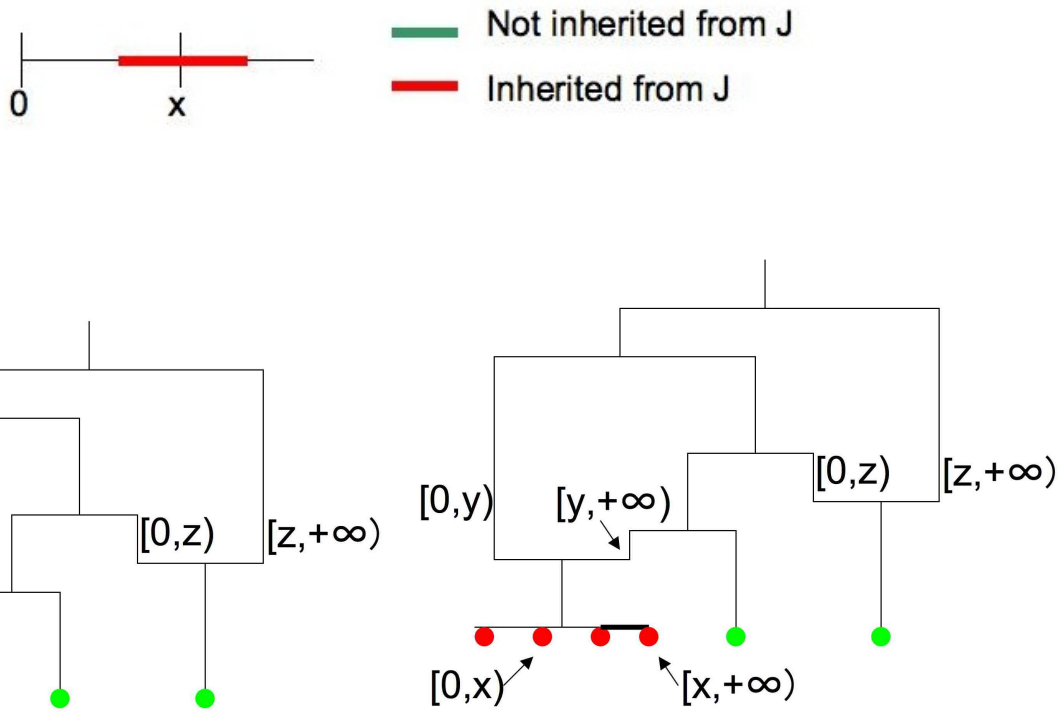
$C_1(x)$:



$$\mathbb{P}_{(\tau',T,T')}(dt',\mathcal{T},\mathcal{T}') = \frac{\exp\left(-\int_0^{t'} A_s(x^-)ds\right)}{A_{0+}(x^-)-1} 1_{t'\in\mathcal{T}'} d\tau'.$$

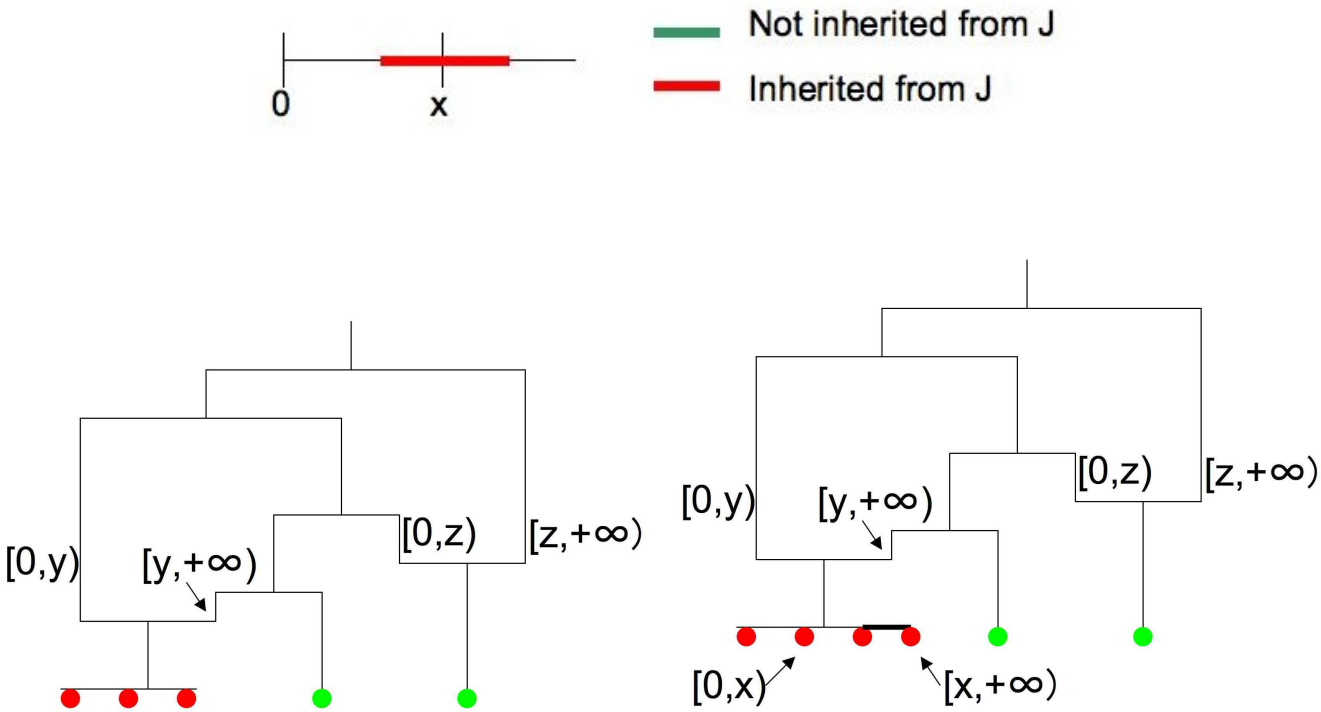
$$\mathbb{P}(C_1(x)|\text{recombination at } x) = \left(1-\frac{R(x^-)}{n}\right)\left(1-\frac{1}{2}\exp(-\gamma x)\right)$$

$C_2(x)$:



The impacted tooth is chosen uniformly on the comb $\mathcal{P}(x^-)$.

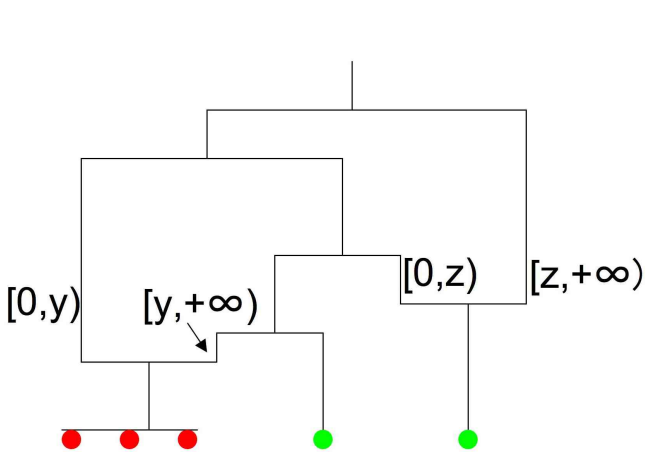
$C_2(x)$:

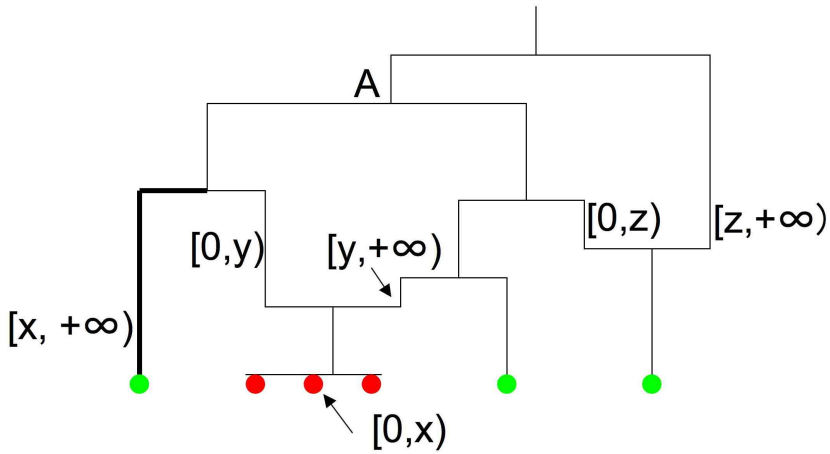


The impacted tooth is chosen uniformly on the comb $\mathcal{P}(x^-)$.

$$\mathbb{P}(C_2(x)|\text{recombination at } x) = \frac{R(x^-)}{n} \times \frac{1}{2}\exp(-\gamma x)$$

$C_3(x)$:

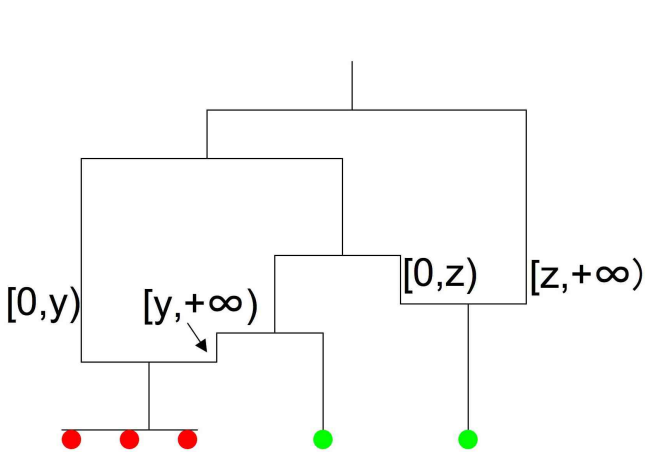


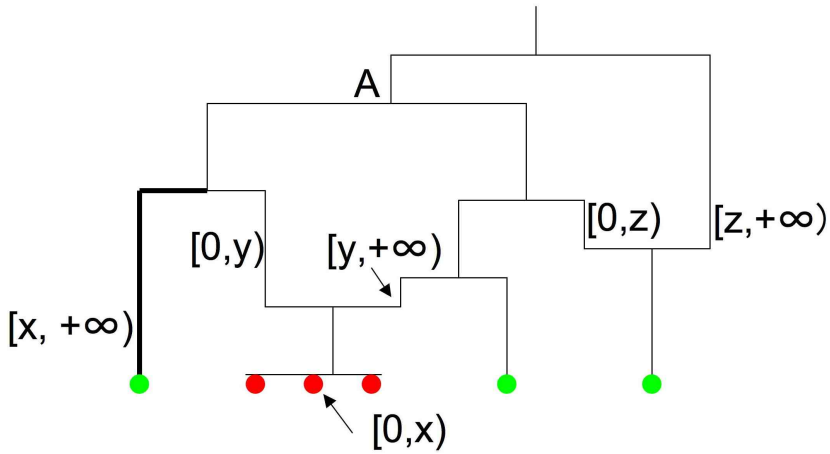


Let I be the impacted tooth of $\mathcal{P}(x^-)$.

$$\mathbb{P}_{(I,\tau',T')}(\mathcal{I}, dt', \mathcal{T}') = \frac{1}{|\mathcal{P}(x^-)|} \exp\left(-\int_0^{t'} A_s(x^-)ds\right) 1_{t' \in \mathcal{T}'} dt'.$$

$C_3(x)$:



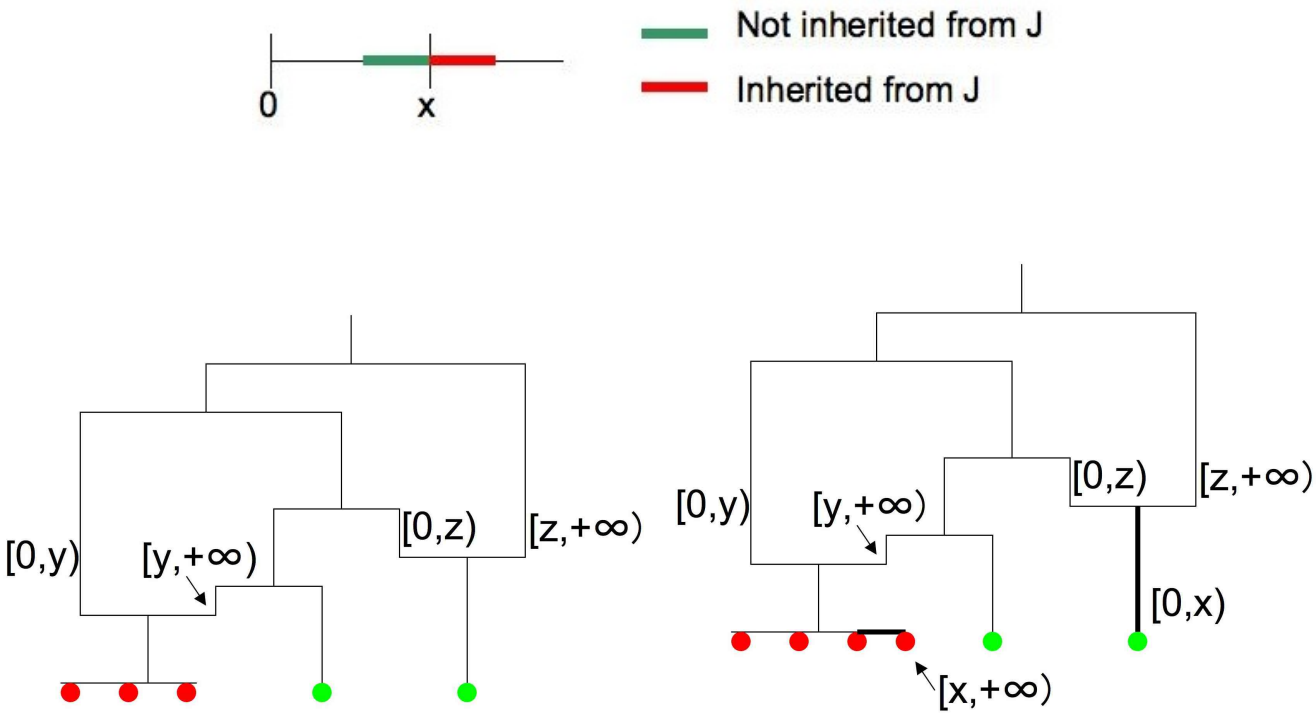


Let I be the impacted tooth of $\mathcal{P}(x^-)$.

$$\mathbb{P}_{(I,\tau',T')}(\mathcal{I}, dt', \mathcal{T}') = \frac{1}{|\mathcal{P}(x^-)|} \exp\left(-\int_0^{t'} A_s(x^-)ds\right) 1_{t' \in \mathcal{T}'} dt'.$$

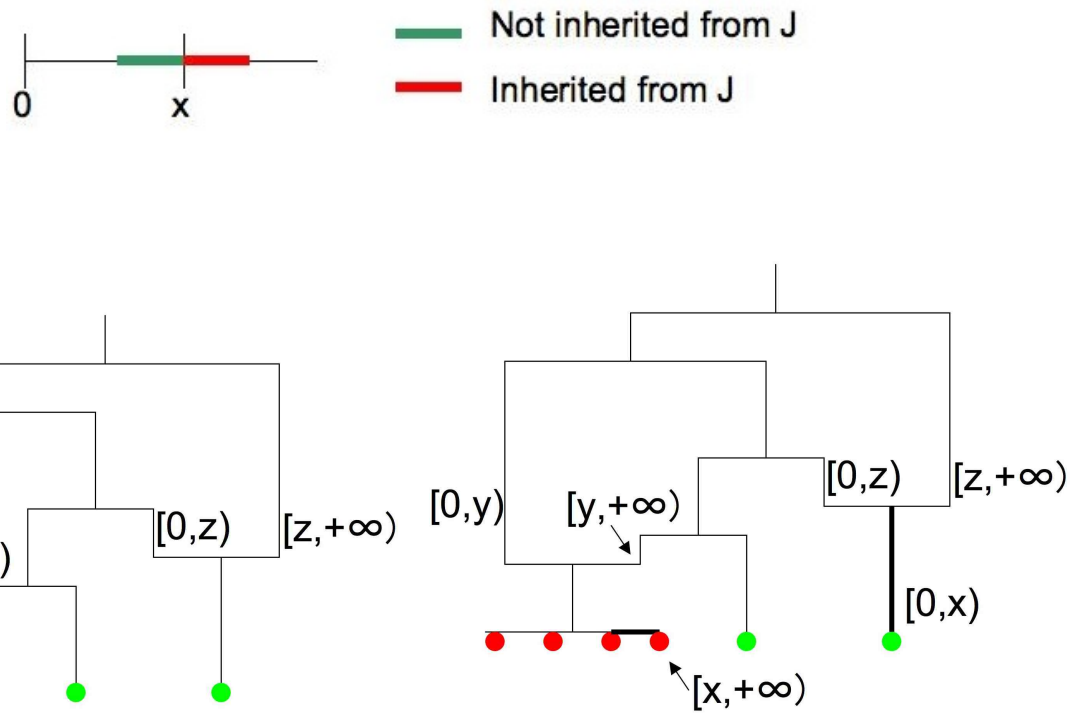
$$\mathbb{P}(C_3(x)|\text{recombination at } x) = \frac{R(x^-)}{n}\left(1 - \frac{\exp(-\gamma x)}{2}\right)$$

$C_4(x)$:



$\mathcal{T}$ is chosen uniformly among the $A_{0^+}(x^-) - 1$ branches attached to one of the leaves of the graph other than the comb.

$C_4(x)$:



$\mathcal{T}$ is chosen uniformly among the $A_{0^+}(x^-) - 1$ branches attached to one of the leaves of the graph other than the comb.

$$\mathbb{P}(C_4(x)|\text{recombination at } x) = \left(1 - \frac{R(x^-)}{n}\right) \frac{1}{2} \exp(-\gamma x)$$

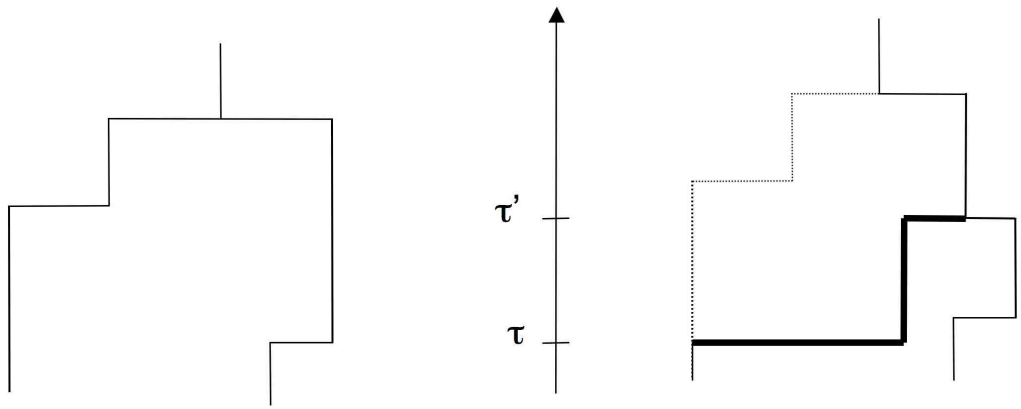
Theorem (Summary)

1. *ARG*(0) *is a comb with n teeth.*
2. *The recombinations positions follow a Poisson process with parameter* $2n\gamma + \lambda L(x)$*, where* $L(x)$ *is the total length of ARG(x).*
3. *At the position x of a jump:*
   - *With probability* $\frac{\lambda L(x^-)}{2n\gamma+\lambda L(x^-)}$*, the recombination occurs during the neutral period. The comb is unchanged and the graph is modified as under neutrality.*
   - *With probability* $\frac{2n\gamma}{2n\gamma+\lambda L(x^-)}$*, the recombination occurs during the selective sweep and the evolution of the comb and the graph is one of the cases* $C_i, 1 \leq i \leq 4$*.*

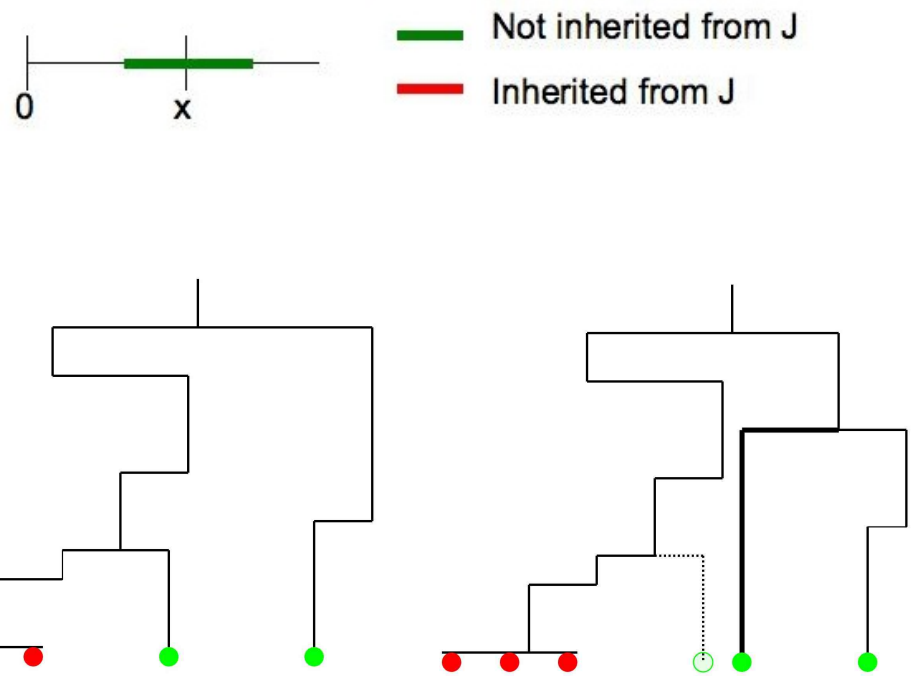
Plan

1. Introduction
   - Selective sweep with strong selective advantage
   - Recombinations
2. Evolution of the coalescent tree along the genome
   - Description of the coalescent tree
   - Evolution of the tree
3. Evolution of the ARG along the genome
   - Evolution under neutrality
   - Evolution in case of selective sweep
   - Back to the process of coalescent trees

If the recombination occurs during the neutral period, a part of a branch is suppressed and a new one is created and coalesces with another branch of the tree.

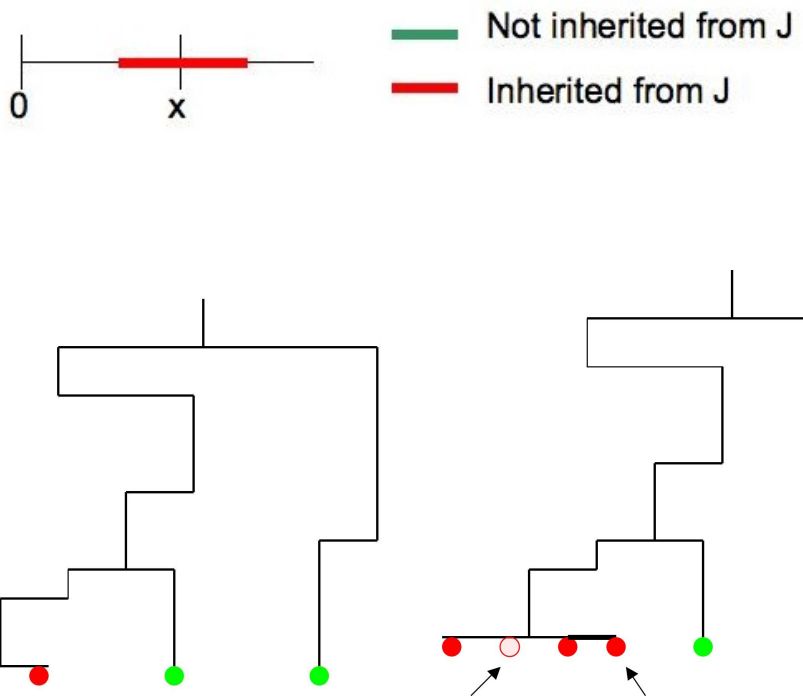


If the recombination occurs during the selective sweep: $C_1(x)$:



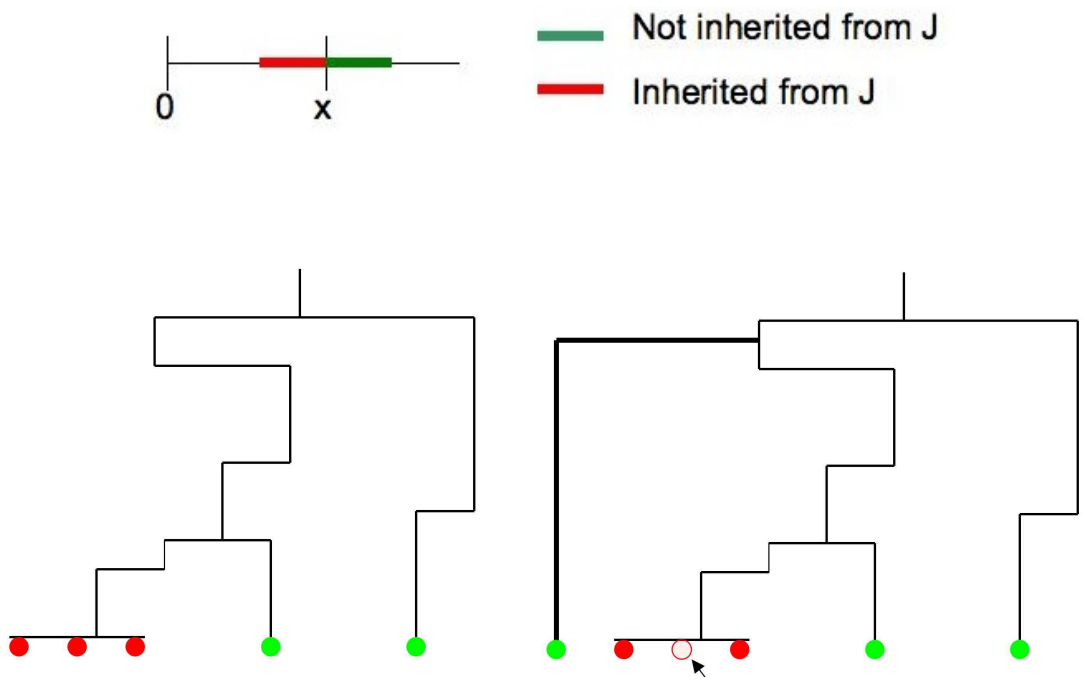
A branch ending by a leaf other than the comb suppressed.
A new one starts at 0. Coalescence rate at time t = number of lineages in the tree at time t.
The branch chosen to coalesce is drawn uniformly among the existing ones.

$C_2(x)$:



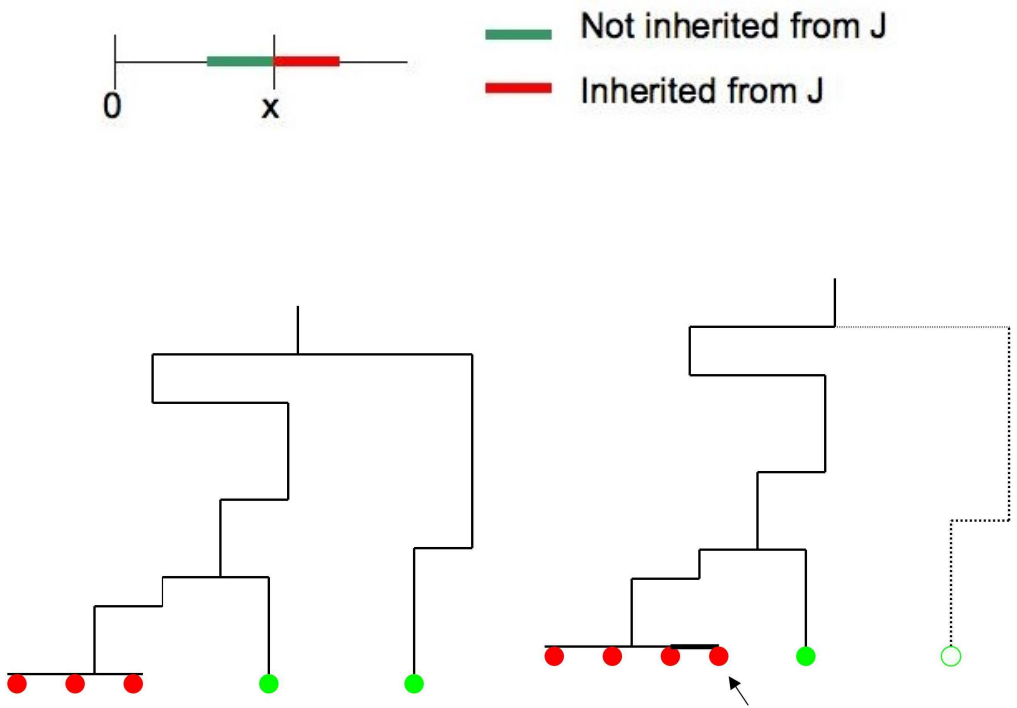
Coalescent tree unchanged.

$C_3(x)$:



One tooth of the comb is suppressed.
A new branch starts at 0. Coalescence rate at time t = number of lineages in the tree at time t.
The branch chosen to coalesce is drawn uniformly among the existing ones.

$C_4(x)$:



A branch ending by a leaf other than the comb is suppressed. One tooth is added to the comb.

$$\nu(x) := \mathbb{P}(C_1(x) \cup C_3(x) \cup C_4(x) | \text{recombination at } x)$$

Theorem (Summary)

1. *Tree(0) is a comb with n teeth.*
2. *The recombinations positions follow a Poisson process with parameter* $2n\gamma\nu(x) + \lambda\tilde{L}(x)$*, where* $\tilde{L}(x)$ *is the total length of tree(x).*
3. *At the position x of a jump:*
   - *With probability* $\frac{\lambda\tilde{L}(x^-)}{2n\gamma\nu(x)+\lambda\tilde{L}(x^-)}$*, the recombination occurs during the neutral period. The comb is unchanged and the tree is modified as under neutrality.*
   - *With probability* $\frac{2n\gamma\nu(x)}{2n\gamma\nu(x)+\lambda\tilde{L}(x^-)}$*, the recombination occurs during the selective sweep and the evolution of the comb and the tree is one of the cases* $C_i, i \in \{1, 3, 4\}$*.*