Stability of recombination hotspots and mutation load

Thibault Joseph

Le Cross-over

Julien Latrille

Team meeting 11/02/2022

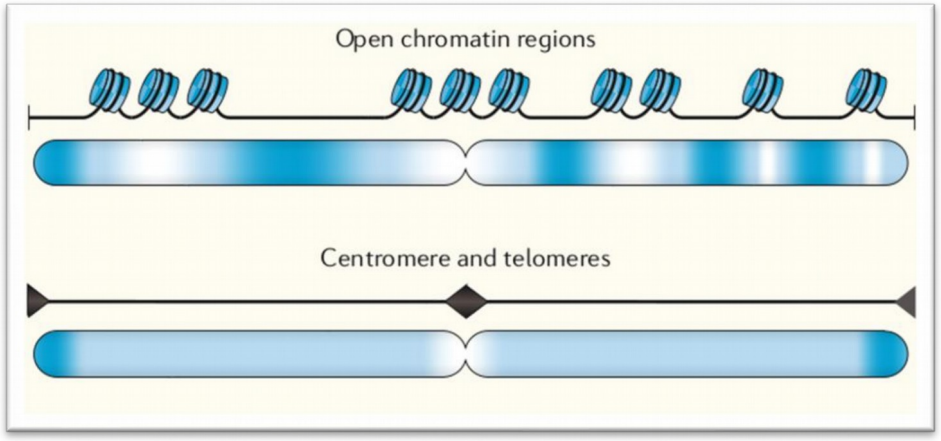
LBBE
BIOMETRIE ET BIOLOGIE EVOLUTIVE

cnrs

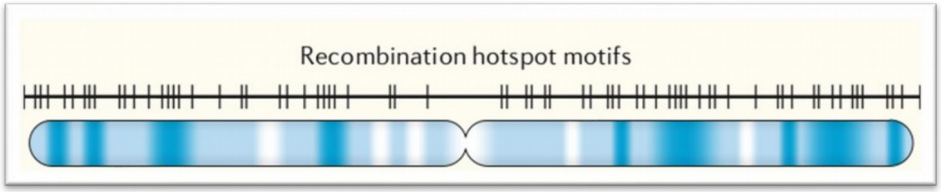
Université Claude Bernard Lyon 1

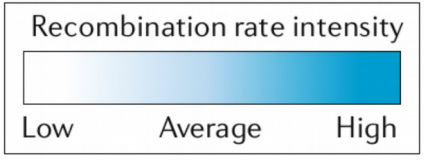
Recombination rate varies along the genome at different scales

At the Mb scale



At the kb scale





Some species have recombination hotspots, others not.

What are the advantages and disadvantages of having recombination hotspots ?

Why are hotspots a good thing ?

Increase fertility

Why are hotspots a good thing ?

Increase fertility

Facilitate homolgy search by limiting the number of sites in a large genome.

Decrease the time of gametogenesis and enhance fertility.

1st order selection on modifyer of recombination heterogeneity

Why are hotspots a good thing ?

Increase fertility

Facilitate homolgy search by limiting the number of sites in a large genome.

Decrease the time of gametogenesis and enhance fertility.

1st order selection on modifyer of recombination heterogeneity

Only evidence in rodents (mice and rats with PRDM9 which have different roles)

Why are hotspots a good thing ?

Increase fertility

Facilitate homolgy search by limiting the number of sites in a large genome.

Decrease the time of gametogenesis and enhance fertility.

1st order selection on modifyer of recombination heterogeneity

Only evidence in rodents (mice and rats with PRDM9 which have different roles)

Enhance selection efficiency

Why are hotspots a good thing ?

Increase fertility

Facilitate homolgy search by limiting the number of sites in a large genome.

Decrease the time of gametogenesis and enhance fertility.

1st order selection on modifyer of recombination heterogeneity

Only evidence in rodents (mice and rats with PRDM9 which have different roles)

Enhance selection efficiency

Concentrate recombination near gene promoter could increase the efficacy of selection in region where mutations display the most extremes fitness effects.

Why are hotspots a good thing ?

Increase fertility

Facilitate homolgy search by limiting the number of sites in a large genome.

Decrease the time of gametogenesis and enhance fertility.

1st order selection on modifyer of recombination heterogeneity

Only evidence in rodents (mice and rats with PRDM9 which have different roles)

Enhance selection efficiency

Concentrate recombination near gene promoter could increase the efficacy of selection in region where mutations display the most extremes fitness effects.

Recombination in hotspots is far more elevated than what would be needed to break genetic linkage

Why are hotspots a bad thing ?

They call it the achilles heels of genomes…

Why are hotspots a bad thing ?

They call it the achilles heels of genomes…

The dark side of recombination…

Why are hotspots a bad thing ?

They call it the achilles heels of genomes…

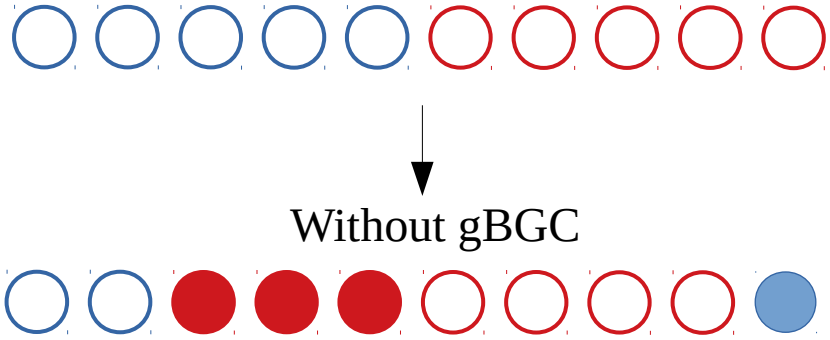
The dark side of recombination…

gbgc

Advantage and disadvantage of gBGC

Advantage and disadvantage of gBGC

If there is a mutation bias towards AT



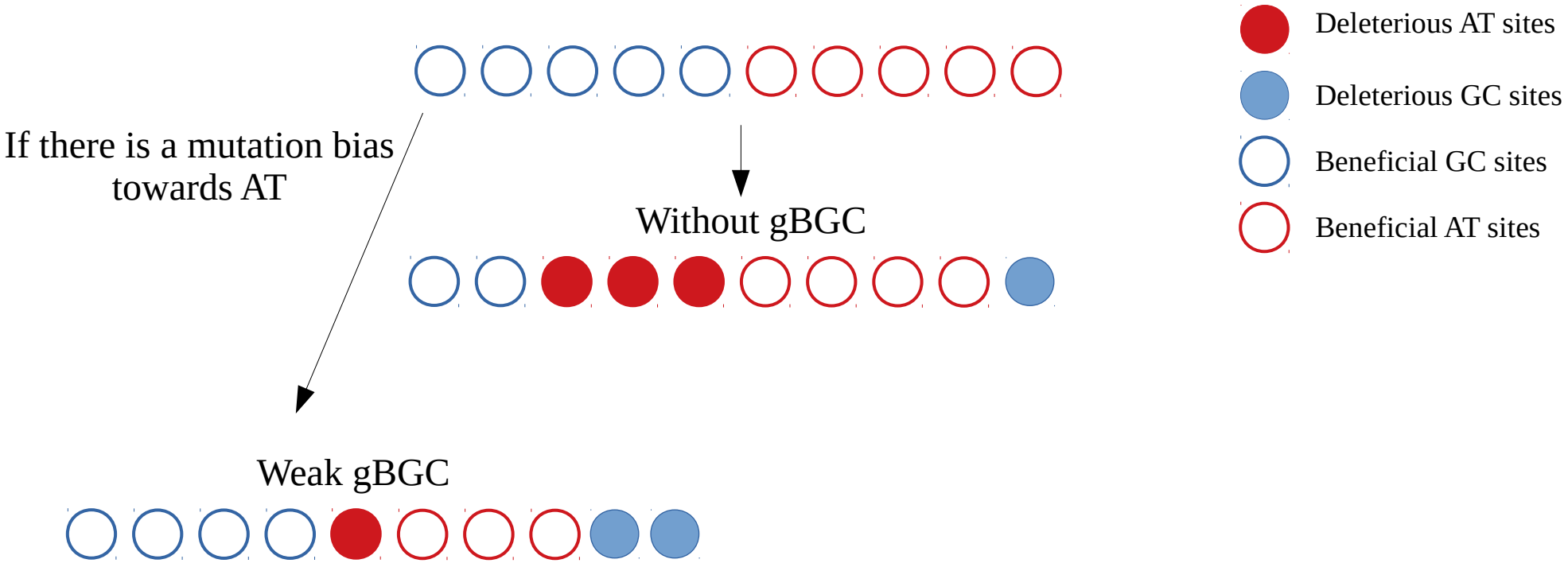
Deleterious AT sites

Deleterious GC sites

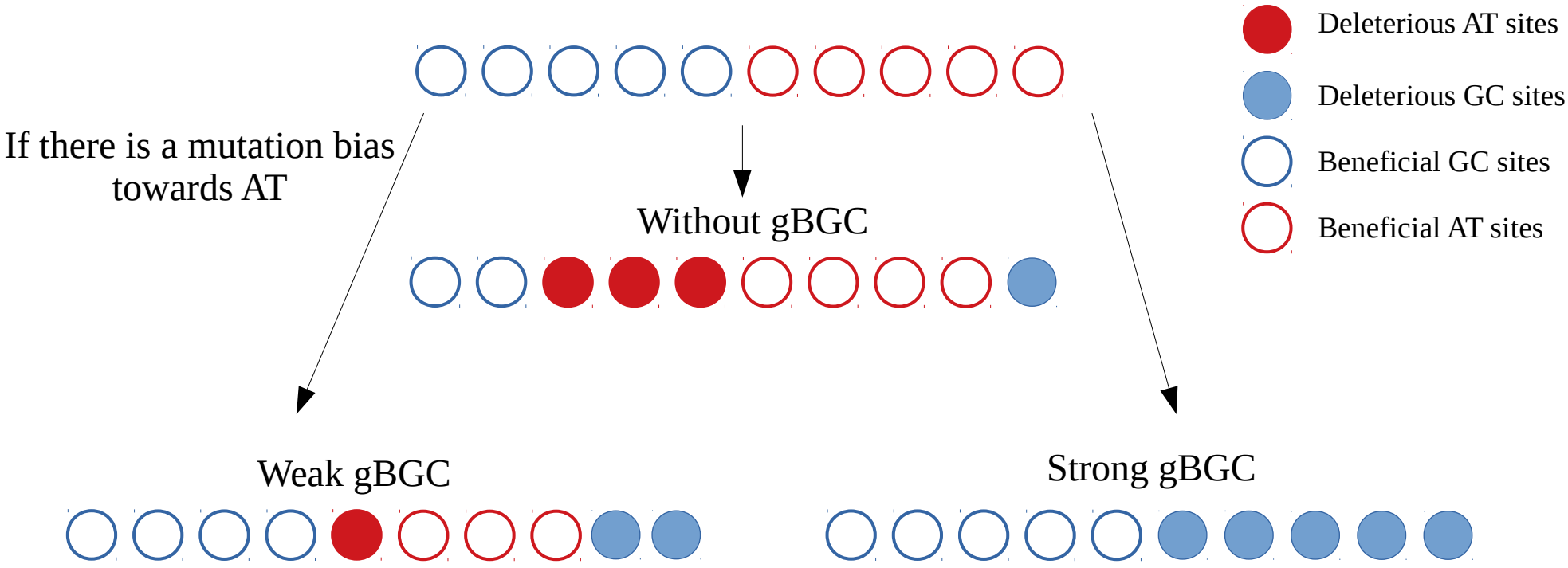
Beneficial GC sites

Beneficial AT sites

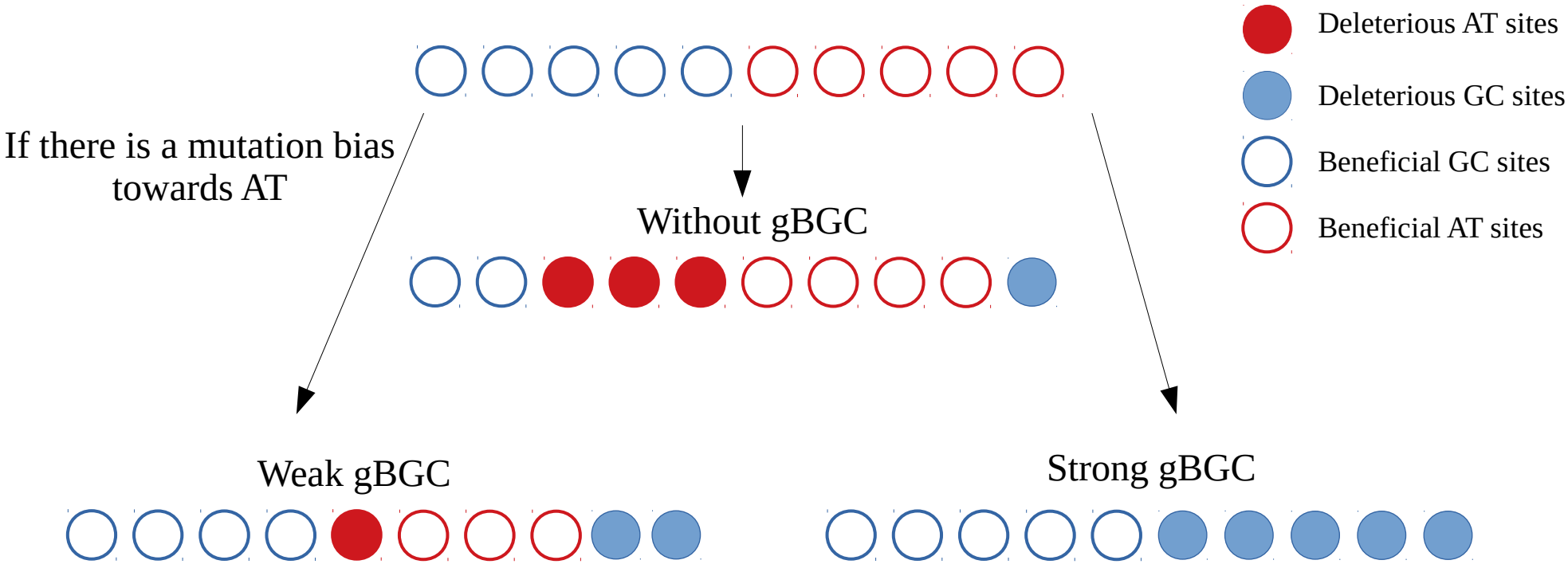
Advantage and disadvantage of gBGC



Advantage and disadvantage of gBGC



Advantage and disadvantage of gBGC



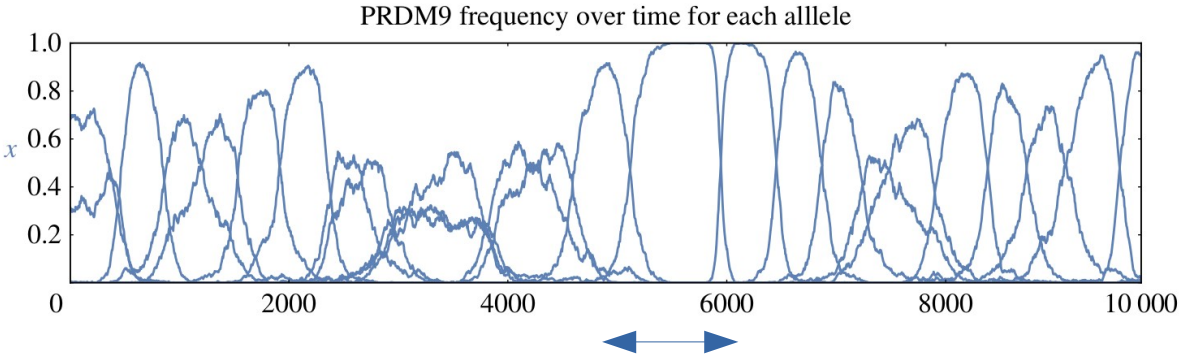
The impact of gBGC on the mutation load has been broadly studied theoretically in Nagylaki 1983, Glemin et al. 2003, Glemin 2010.

Problematic

All previous models assume a constant intensity of gBGC across time.

« I also assume that the gBGC intensity is constant. However, strong gBGC events are thought to be associated with short-lived recombination hotspots, at least in humans (McVean *et al.* 2004; Myers *et al.* 2005). I thus implicitly assume that gBGC/selection dynamics are shorter than the recombination hotspot lifespan. »

S. Glémin 2010



Latrille et al. 2017

Problematic

All previous models assume a constant intensity of gBGC across time.

« I also assume that the gBGC intensity is constant. However, strong gBGC events are thought to be associated with short-lived recombination hotspots, at least in humans (McVean *et al.* 2004; Myers *et al.* 2005). I thus implicitly assume that gBGC/selection dynamics are shorter than the recombination hotspot lifespan. »

S. Glémin 2010

For which value of b, s and hotspot lifespan does this assumption hold ?

Problematic

All previous models assume a constant intensity of gBGC across time.

« I also assume that the gBGC intensity is constant. However, strong gBGC events are thought to be associated with short-lived recombination hotspots, at least in humans (McVean *et al.* 2004; Myers *et al.* 2005). I thus implicitly assume that gBGC/selection dynamics are shorter than the recombination hotspot lifespan. »

S. Glémin 2010

For which value of b, s and hotspot lifespan does this assumption hold ?

Can short-lived hotspot reduce the mutation load ?

Deterministic model under gBGC and selection

Let x_t be the frequency of a strong (GC) allele at time t.

- We consider deleterious mutations from weak (AT) to strong (GC), with selection coefficient s and dominance h.
- We consider that weak alleles are converted to strong allele in heterozygous due to gBGC with conversion rate b.

Genotype	**Homozygous SS**	**Heterozygous WS**	**Homozygous WW**
Frequency	x_t^2	$2x_t(1-x_t)$	$(1-x_t)^2$
Fitness	$1-s$	$1-hs$	1

Nagylaki (1983), Glémin (2010)

Deterministic model under gBGC and selection

Let x_t be the frequency of a strong (GC) allele at time t.

- We consider deleterious mutations from weak (AT) to strong (GC), with selection coefficient s and dominance h.
- We consider that weak alleles are converted to strong allele in heterozygous due to gBGC with conversion rate b.

Genotype	**Homozygous *SS***	**Heterozygous *WS***	**Homozygous *WW***
Frequency	x_t^2	$2x_t(1-x_t)$	$(1-x_t)^2$
Fitness	$1-s$	$1-hs$	1

The frequency of a strong allele at the next generation (x_{t+1}) is thus:

$$x_{t+1} = \frac{(1-s)x_t^2 + (1+b)(1-hs)x_t(1-x_t)}{\bar{\omega}}.$$

And the mean fitness ($\bar{\omega}$) is:

$$\begin{aligned}\bar{\omega} &= (1-s)x_t^2 + (1-hs)2x_t(1-x_t) + (1-x_t)^2, \\ &= 1 - sx_t^2 - 2hsx_t(1-x_t).\end{aligned}$$

- **What is the steady state of strong alleles? Fixed or purged away?**

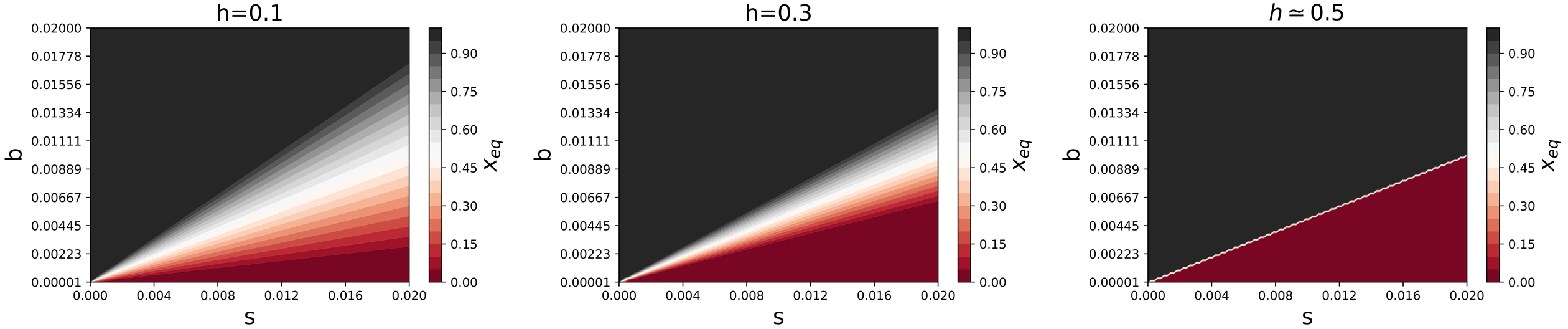
Nagylaki (1983), Glémin (2010)

Equilibrium for the deterministic model under gBGC and selection

At equilibrium, the frequency of a strong allele is:

$$x_{\text{eq}} = \frac{(1-s)x_{\text{eq}}^2 + (1+b)(1-hs)x_{\text{eq}}(1-x_{\text{eq}})}{1 - sx_{\text{eq}}^2 - 2hsx_{\text{eq}}(1-x_{\text{eq}})},$$

$$\Rightarrow x_{\text{eq}} = \frac{b-hs}{s-2hs} \text{ if } h < 1/2.$$



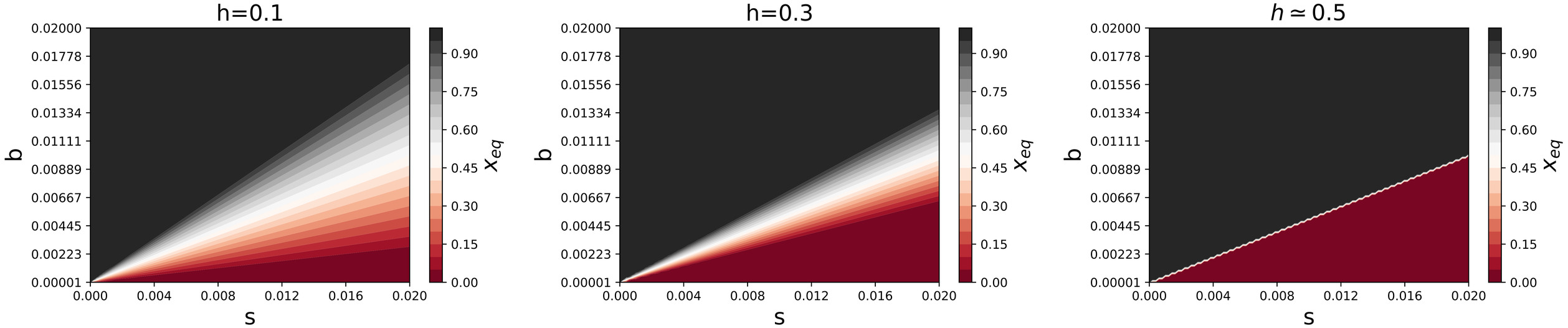
Glémin (2010)

Equilibrium for the deterministic model under gBGC and selection

At equilibrium, the frequency of a strong allele is:

$$x_{\text{eq}} = \frac{(1-s)x_{\text{eq}}^2 + (1+b)(1-hs)x_{\text{eq}}(1-x_{\text{eq}})}{1 - sx_{\text{eq}}^2 - 2hsx_{\text{eq}}(1-x_{\text{eq}})},$$

$$\Rightarrow x_{\text{eq}} = \frac{b-hs}{s-2hs} \text{ if } h < 1/2.$$



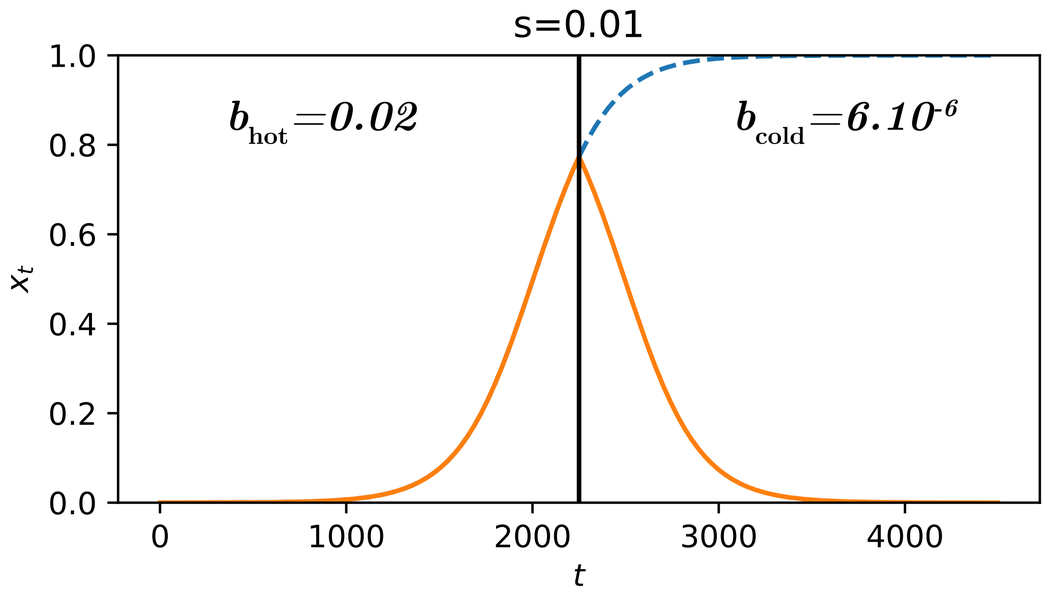
- **If b (gBGC) is strong compared to s (selection), deleterious mutations reach fixation.**
- **If b (gBGC) is weak compared to s (selection), deleterious mutations are purged away.**
- **If b and s are comparable, the mutation is maintained in the population.**
- **But what if b is not constant across time (from b_{hot} to b_{cold})?**

Glémin (2010)

Unstable hotspots in the deterministic model

The hotspots are unstable, with lifespan τ:

$$\begin{cases} t < \tau & \Rightarrow b = b_{\text{hot}} = 0.02, \\ t > \tau & \Rightarrow b = b_{\text{cold}} = 6 \times 10^{-6}. \end{cases}$$



- **Under constant gBGC, the GC allele should have reached fixation (blue line).**

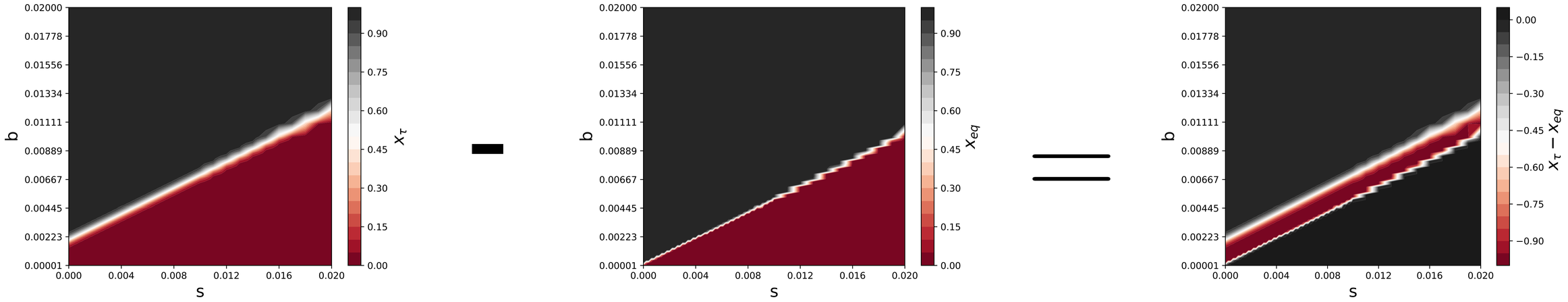
- **But at the end of the hotspot's life, selection overwhelms gBGC and the mutation is purged away (orange line).**

- **For which value of b_{hot}, s and τ does this kind of situation occur ?**

Unstable hotspots in the deterministic model

- x_τ is the frequency of the GC deleterious allele at the end of the hotspot lifespan.
- x_{eq} is the frequency that the allele should have reached if the hotspot were stable.
- The difference between x_{eq} and x_τ represents the benefit of having a short-lived hotspot.

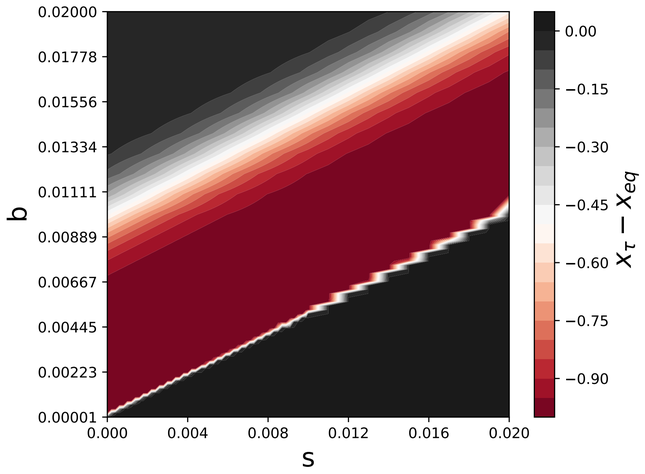
$\tau = 0.5 N_e = 5000$
$h = 0.5$



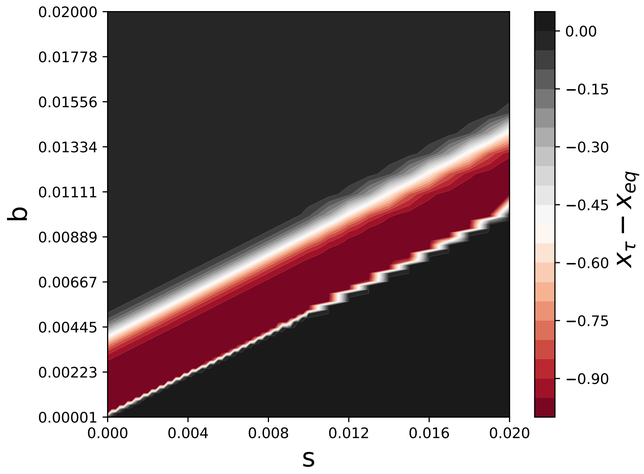
- Having unstable hotspot is only beneficial for a narrow range of values for b_{hot} and s.

Unstable hotspots in the deterministic model

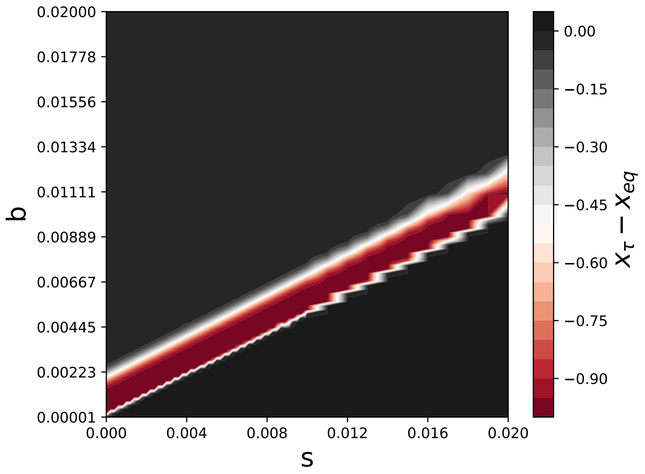
$h=0.5$



$\tau=0.1N_e=1000$



$\tau=0.25N_e=2500$



$\tau=0.5N_e=5000$

- **Shorter lifespan of hotspots leads to deleterious mutations more efficiently purified away.**
  - **Thus far, we assumed only gBGC and selection but no drift.**
    - **Under drift, is this result still holding?**

Stochastic model under gBGC and selection.

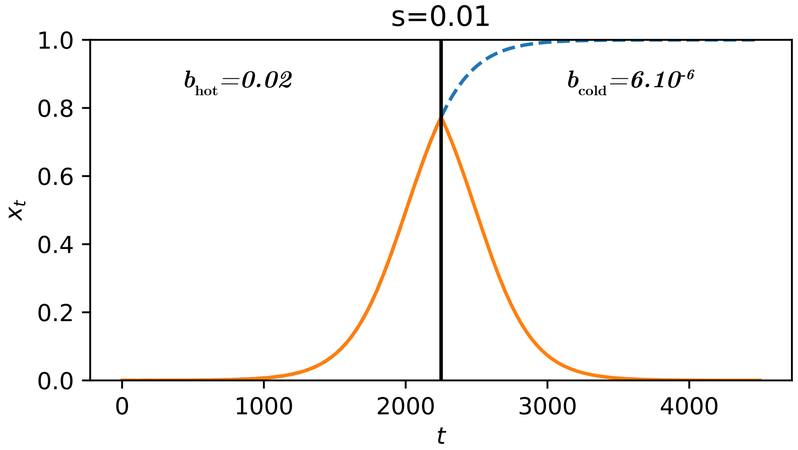
Under a stochastic model, the number of strong alleles is drawn from a binomial distribution:

$$x_{t+1} \sim \frac{1}{2N_{\mathrm{e}}}\mathcal{B}\left(2N_{\mathrm{e}}, p\right),$$

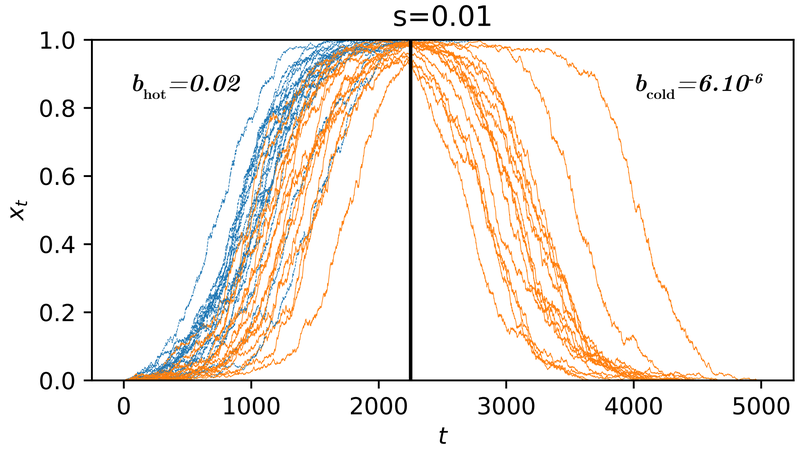
$$p = \frac{(1-s)x_t^2 + (1+b)(1-hs)x_t(1-x_t)}{\bar{\omega}}.$$

h=0.5

Deterministic



Stochastic



The probability of fixation for a GC allele is $\mathbb{P}_{\mathrm{fix}}$, and the fixation load L is defined as:

$$L = s\mathbb{P}_{\mathrm{fix}}$$

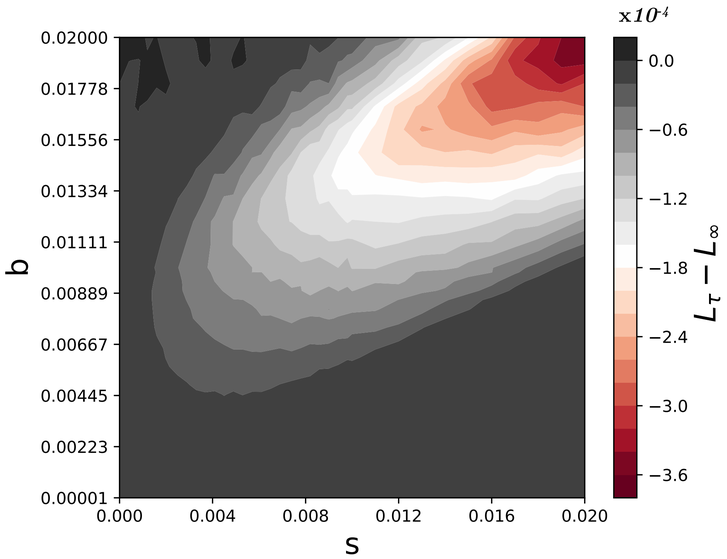
Fixation load

The probability of fixation for a GC allele is $\mathbb{P}_{\text{fix}}$, and the fixation load L is defined as:

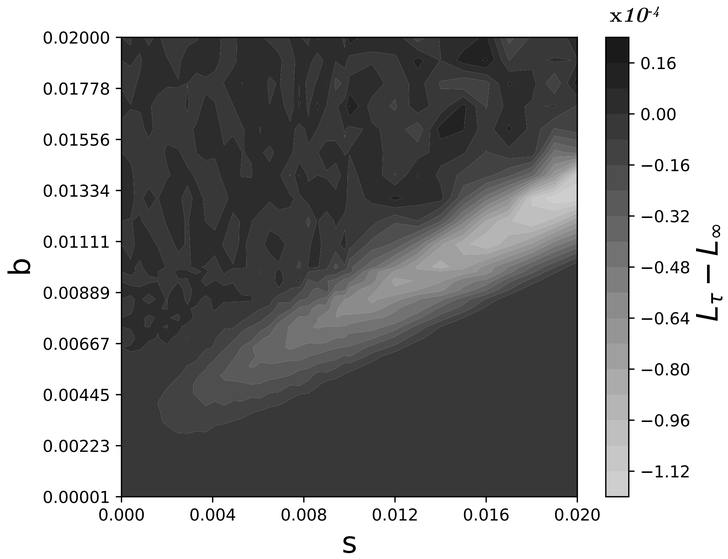
$$L = s\mathbb{P}_{\text{fix}}$$

$$\begin{cases} L_\tau \text{ is the load for an unstable hotspot (lifespan } t = \tau), \\ L_\infty \text{ is the load for a constant hotspot.} \end{cases}$$

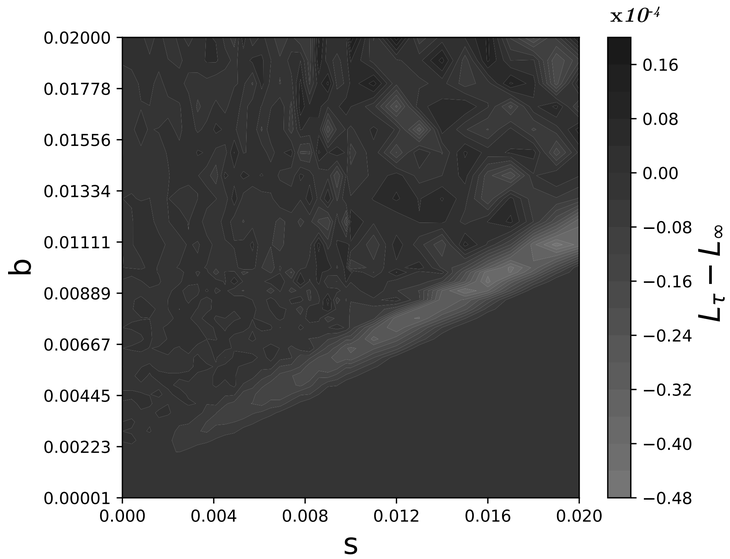
h=0.5



τ=0.1N_e=1000



τ=0.25N_e=2500



τ=0.5N_e=5000

- **If hotspots are lasting more than $0.5N_e$, we barely see any advantage of unstable hotspots.**

Biological relevance : PRDM9-dependent hotspots

Our model mimics a succession regime when we have more or less one PRDM9 allele at a time.

Biological relevance : PRDM9-dependent hotspots

Our model mimics a succession regime when we have more or less one PRDM9 allele at a time.

When PRDM9 is highly polymorphic, the conversion rate from W to S becomes b times the PRDM9 allele frequency, which reduces the load at that site and can bring significant advantage.

Biological relevance : PRDM9-dependent hotspots

Our model mimics a succession regime when we have more or less one PRDM9 allele at a time.

When PRDM9 is highly polymorphic, the conversion rate from W to S becomes b times the PRDM9 allele frequency, which reduces the load at that site and can bring significant advantage.

It is difficult to estimate the turnover time of PRDM9 alleles.

Biological relevance : PRDM9-dependent hotspots

Our model mimics a succession regime when we have more or less one PRDM9 allele at a time.

When PRDM9 is highly polymorphic, the conversion rate from W to S becomes b times the PRDM9 allele frequency, which reduces the load at that site and can bring significant advantage.

It is difficult to estimate the turnover time of PRDM9 alleles.

For mice, Latrille et al. 2017 predicted it between 0.6xNe and 1.6xNe (in a polymorphic regime under several assumptions)

Biological relevance : PRDM9-dependent hotspots

Our model mimics a succession regime when we have more or less one PRDM9 allele at a time.

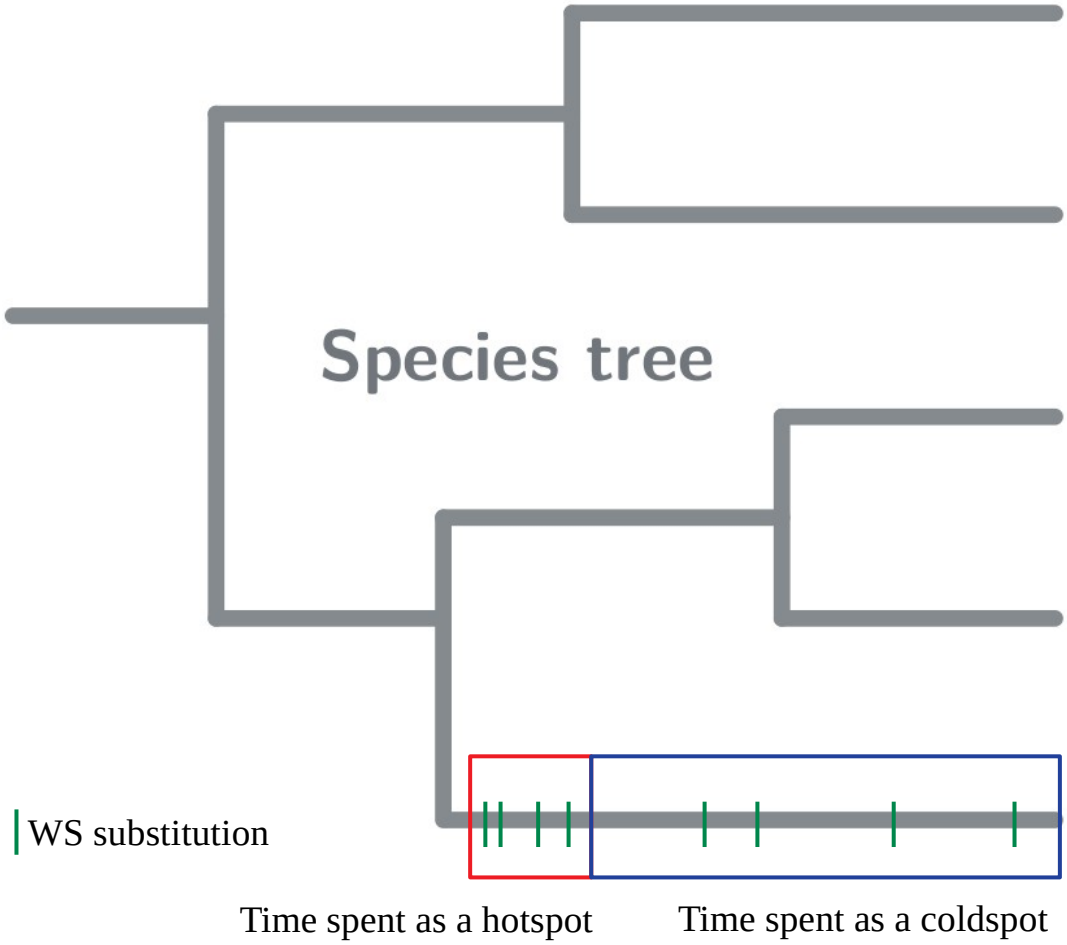
When PRDM9 is highly polymorphic, the conversion rate from W to S becomes b times the PRDM9 allele frequency, which reduces the load at that site and can bring significant advantage.

It is difficult to estimate the turnover time of PRDM9 alleles.

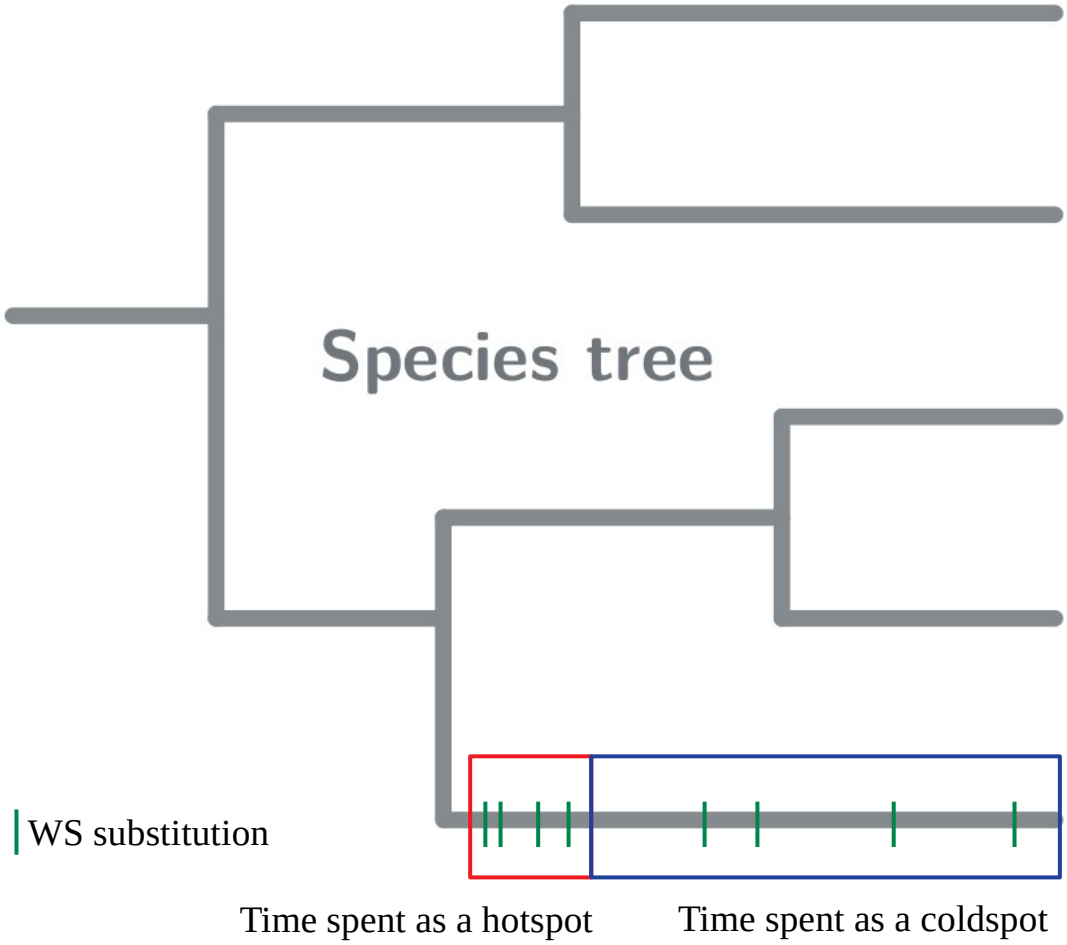
For mice, Latrille et al. 2017 predicted it between 0.6xNe and 1.6xNe (in a polymorphic regime under several assumptions)

If those values are the same in humans, the gBGC/selection dynamic should not be affected that much by the hotspots lifespan.

Consequences on estimations of B=4Neb

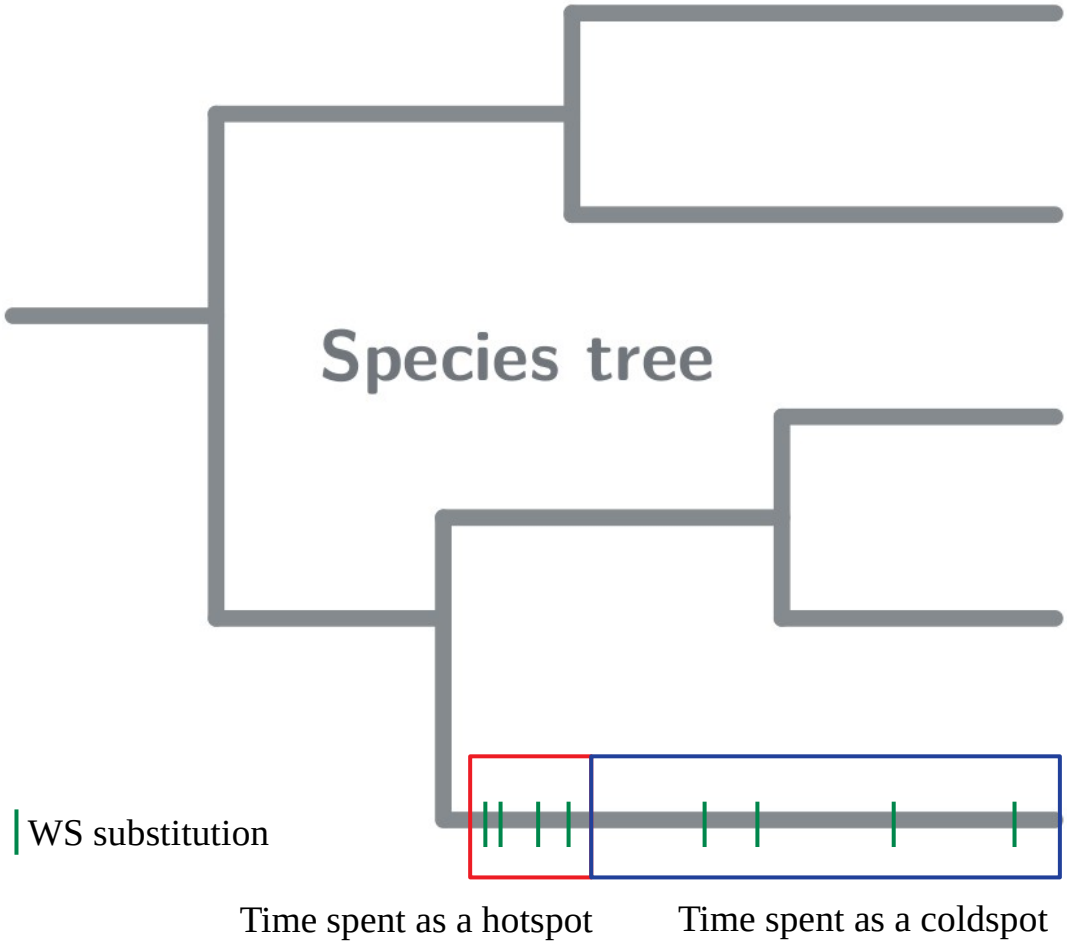


Consequences on estimations of B=4Neb



If the probability of fixation is not affected by the lifespan of the hotspot, the number of substitutions will only depend on the time spent as a hotspot and its strength.

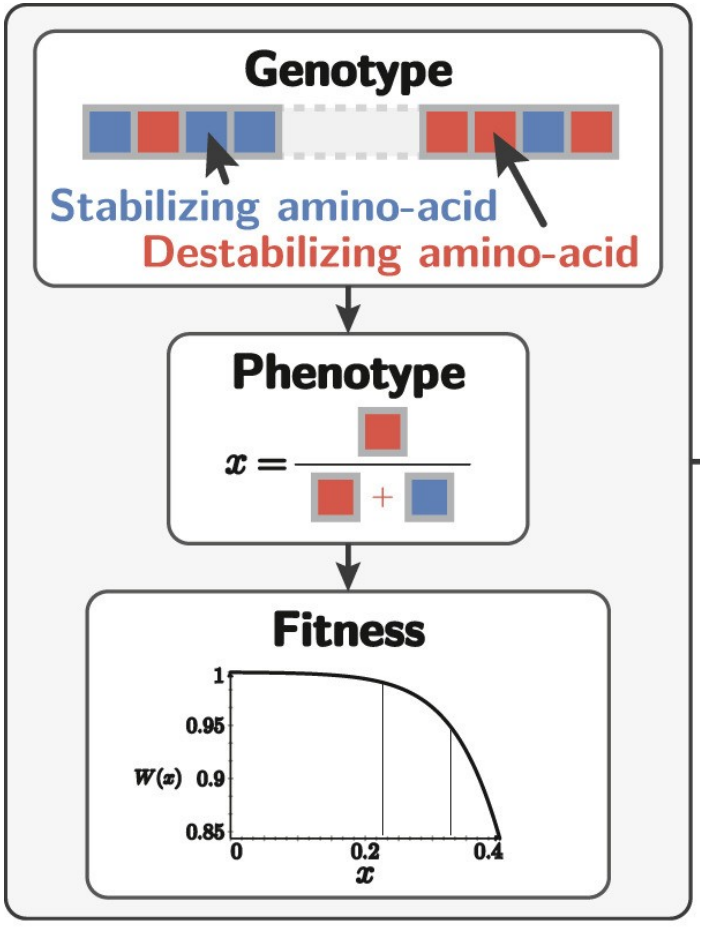
Consequences on estimations of B=4Neb



If the probability of fixation is not affected by the lifespan of the hotspot, the number of substitutions will only depend on the time spent as a hotspot and its strength.

Our results suggest that the mean B integrated over time estimated from neutral substitutions is valid even for PRDM9-induced hotspots.

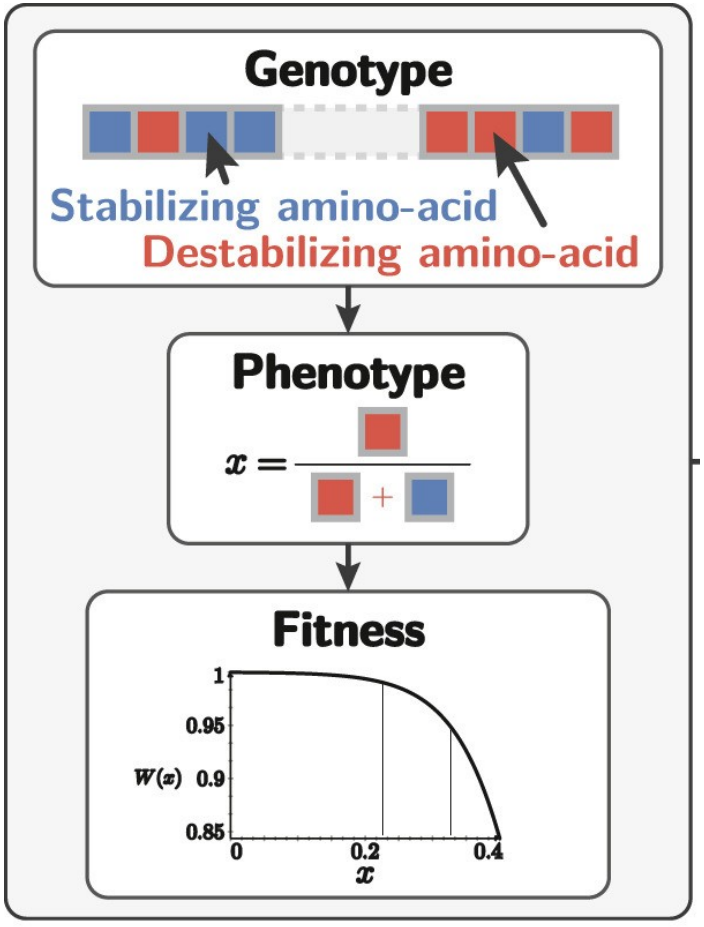
Accounting for fitness landscape and epistasis



Individual genotype to fitness

Latrille and Lartillot 2021

Accounting for fitness landscape and epistasis

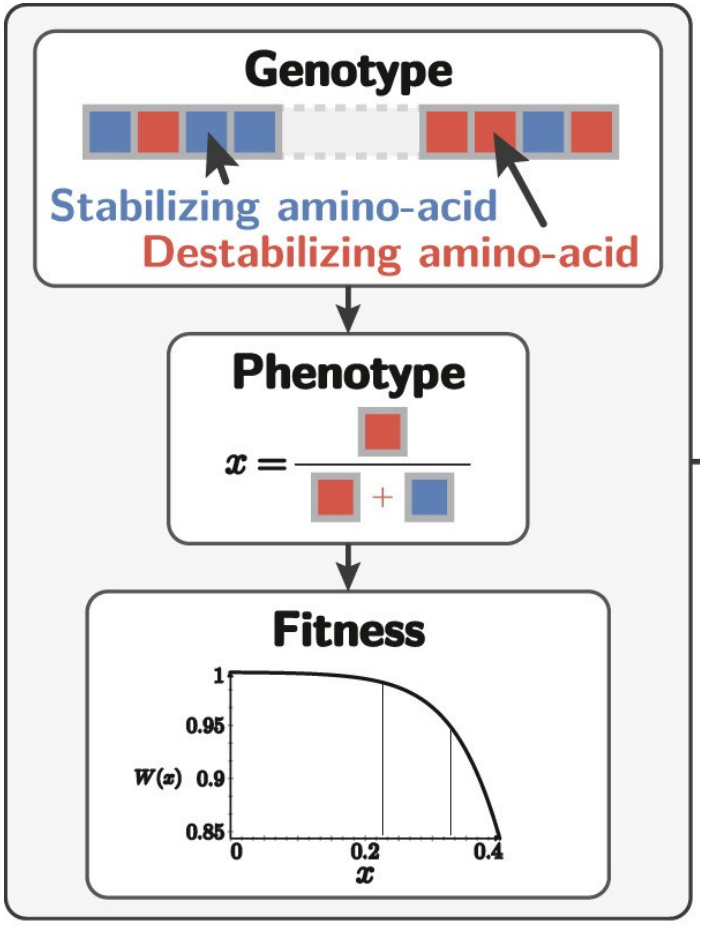


Individual genotype to fitness

Latrille and Lartillot 2021

If epistasis is more likely to occur between proximal sites, for the same amount of deleterious mutations fixed in the genome, clustered deleterious substitutions can have more impact on fitness.

Accounting for fitness landscape and epistasis



Individual genotype to fitness

Latrille and Lartillot 2021

If epistasis is more likely to occur between proximal sites, for the same amount of deleterious mutations fixed in the genome, clustered deleterious substitutions can have more impact on fitness.

Under this condition, having unstable hotspots can limit the deleterious effects of gBGC on fitness.

Thank you for your attention !

Questions ? Suggestions ?