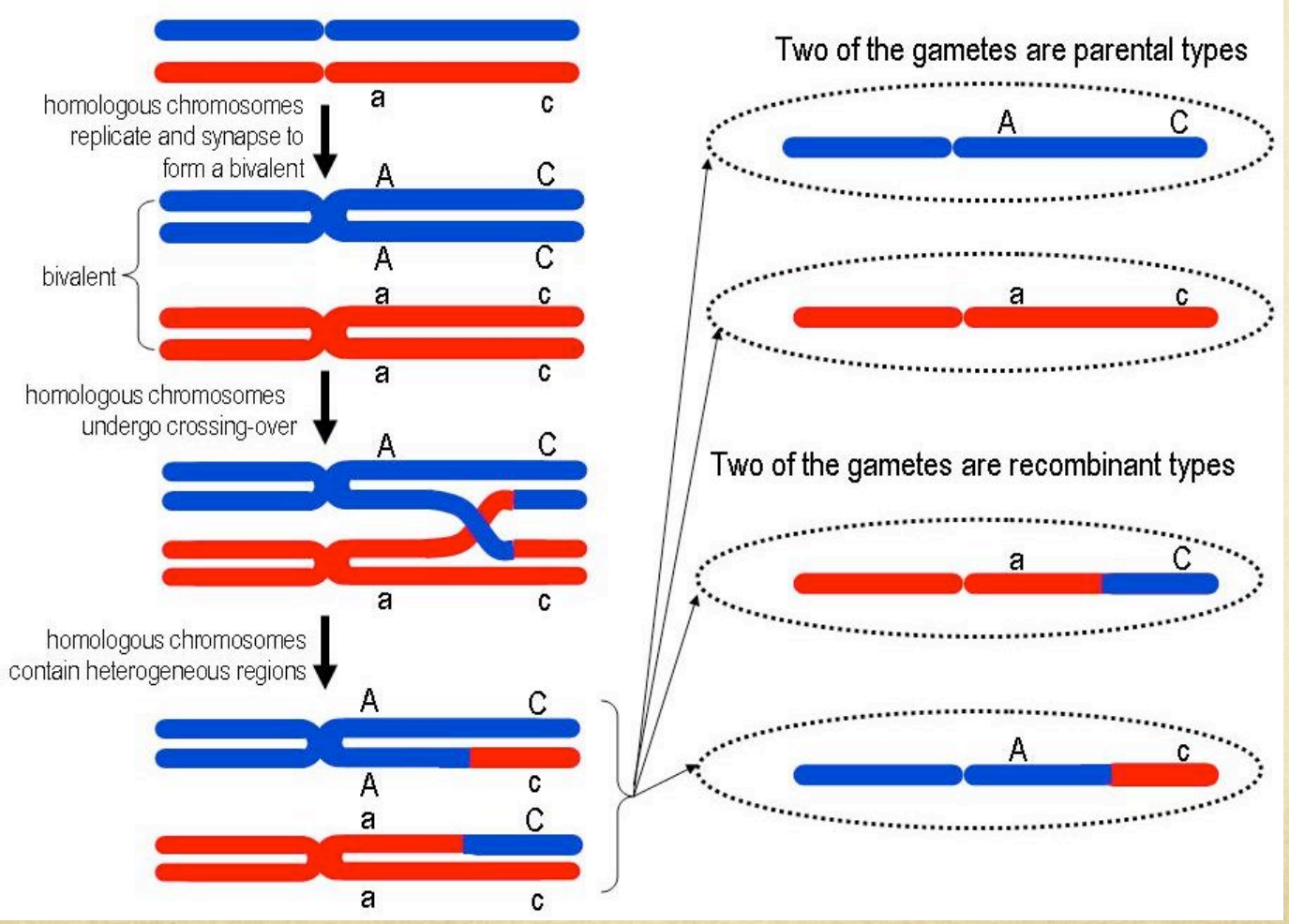
The impact of recombination on genome evolution

Laurent Duret
Laboratoire de Biométrie et Biologie Evolutive, CNRS, Université Lyon 1

Homologous recombination

- What, who, why?
- What:
  - Mechanism of repair of double-strand breaks (DSBs)
  - Mechanism of exchange of genetic material between two similar or identical molecules of DNA

Meiotic recombination

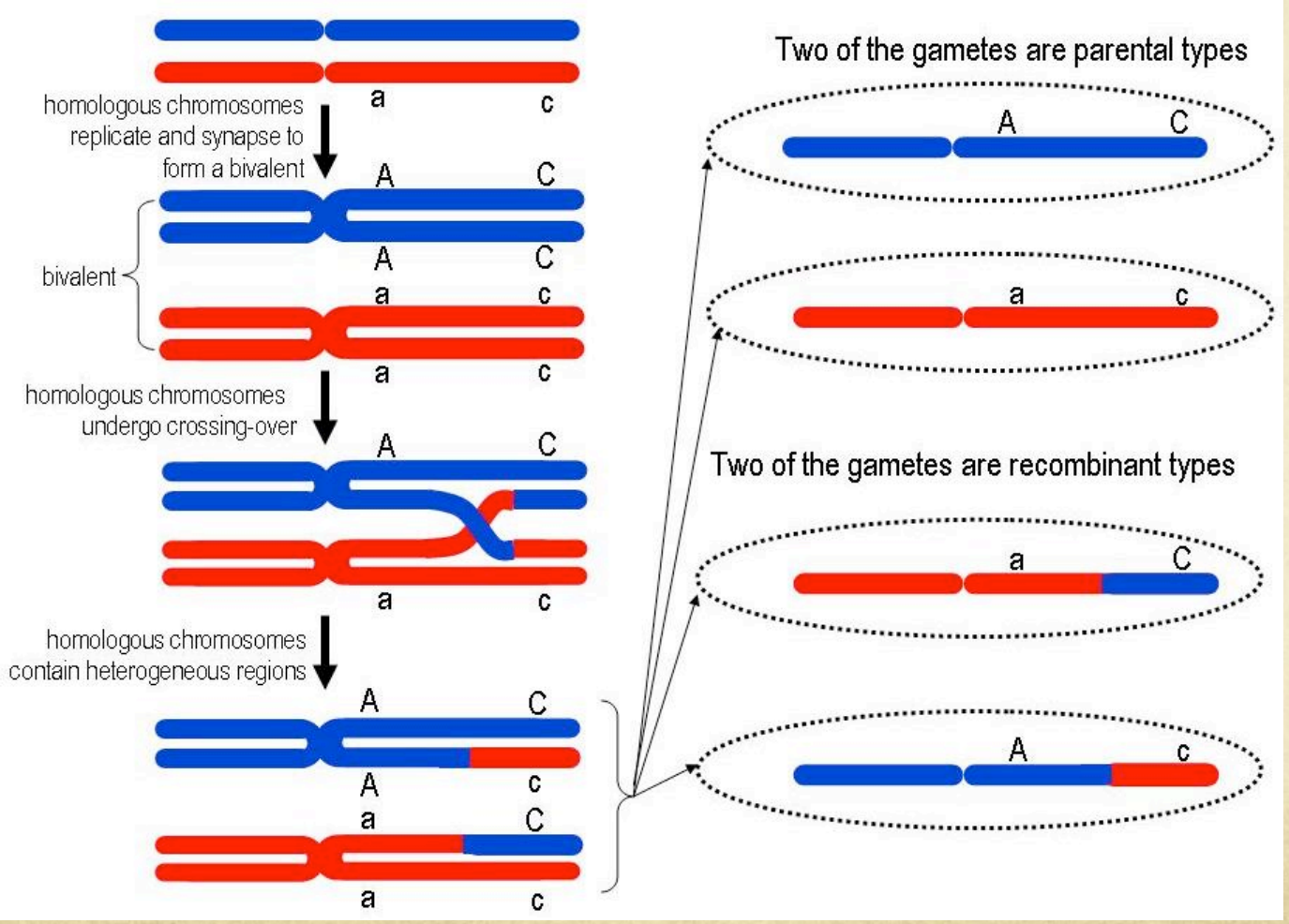


Homologous recombination

- What, who, why?
- What ?
  - Mechanism of repair of double-strand breaks (DSBs)
  - Mechanism of exchange of genetic material between two similar or identical molecules of DNA
- Who ?

  Eukaryotes (meiosis, mitosis), Bacteria, Archaea (horizontal gene transfer)

Why recombine?

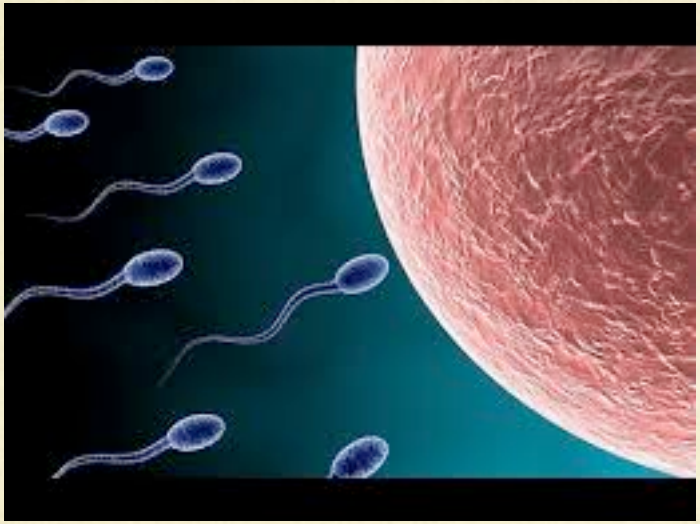


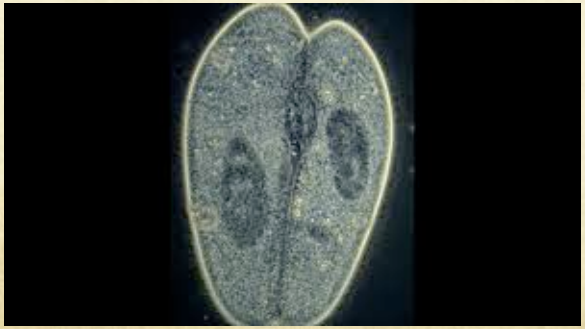
Homologous recombination

- What, who, why?
- What ?
  - Mechanism of repair of double-strand breaks (DSBs)
  - Mechanism of exchange of genetic material between two similar or identical molecules of DNA
- Who ?
  - Eukaryotes (meiosis, mitosis), Bacteria, Archaea (horizontal gene transfer)
- Why ?

  Repair of DSBs
  Genetic diversity (sex, parasexuality)

Why sex?





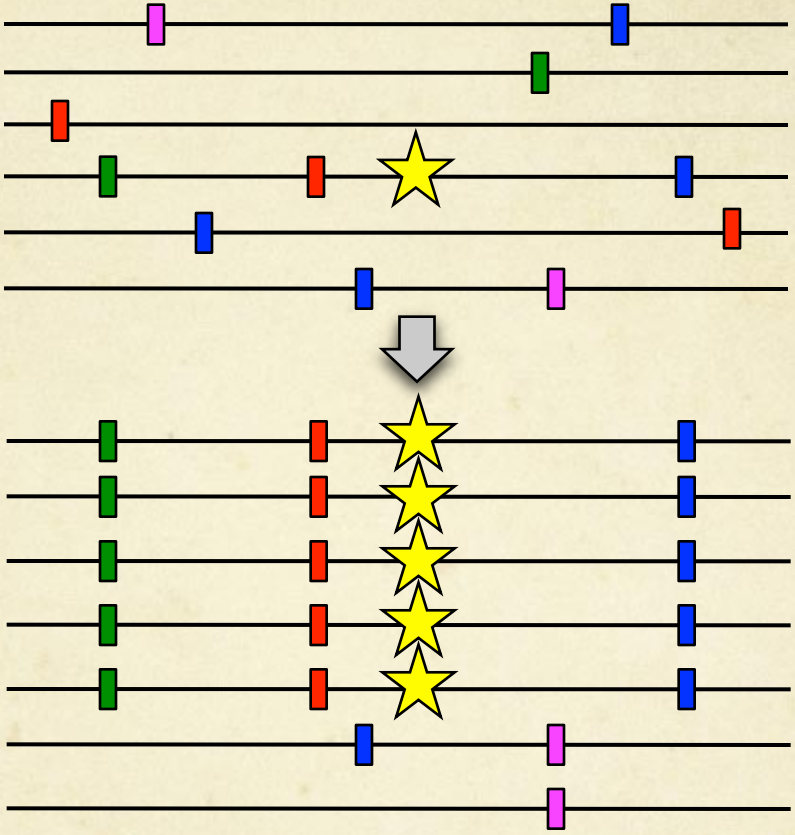
The impact of recombination on the evolution of organisms

- Genetic linkage interferes with selection

Genetic hitch-hiking

Maynard Smith & Haigh (1974)

Genetic diversity in the initial population



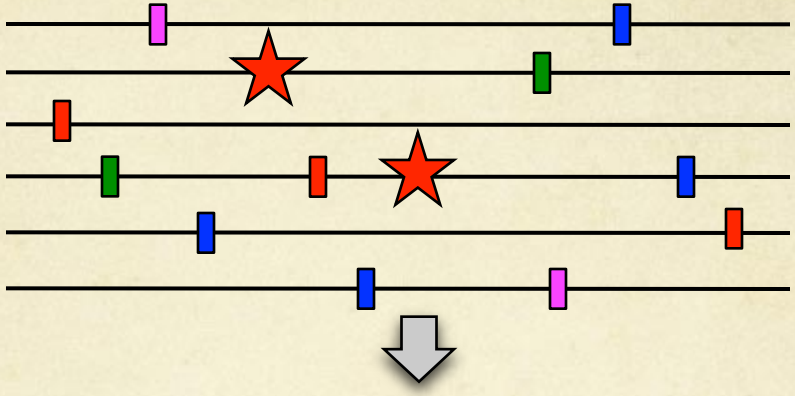
Genetic diversity after the selective event

- Positive selection on a given locus leads to the fixation of alleles at linked sites
- => reduction of polymorphism level at linked sites

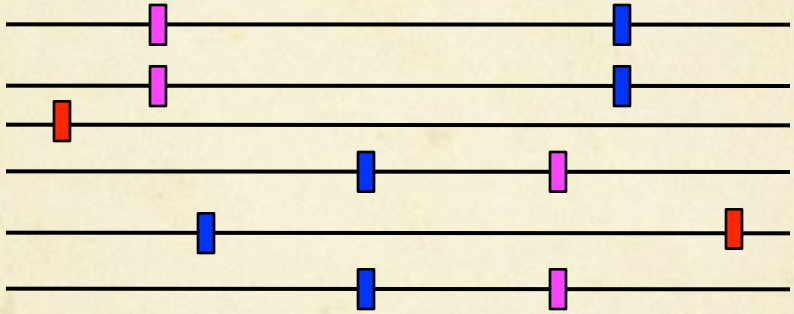
Background selection

Charlesworth et al. (1995)

Genetic diversity in the initial population



Genetic diversity after purifying selection



- Purifying selection leads to a reduction of polymorphism level at linked sites

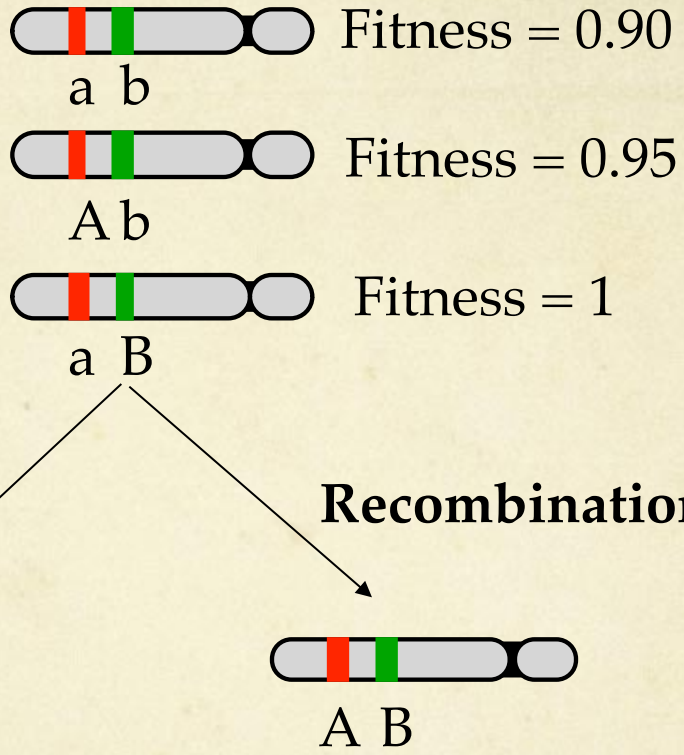
Genetic linkage interferes with selection

- Both positive and purifying selection lead to a reduction of genetic diversity at linked sites
  - => equivalent to a reduction of effective population size (Ne)
  - Lower Ne => more genetic drift => higher probability of fixation of deleterious alleles

Hill-Robertson interference

Linked loci: *a*, *b*

A, *B*: slightly advantageous alleles



Selection in favor of *B* will lead to the fixation of the *aB* haplotype. The advantageous *A* allele will disappear.

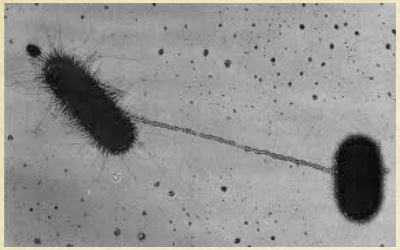
Haplotype *AB* will appear in the population and its fixation will be favored by selection.

Hill & Robertson (1966)

The impact of recombination on the evolution of organisms

- Selection efficiency
  - Selection is less efficient in genomic regions of low recombination rate
  - Lack of recombination leads to the accumulation of deleterious mutations, and decreases the adaptive potential of species
- Degenerate evolution of non-recombining genomes (e.g. Y chromosome)
- Short lifespan of asexual species*

Homologous recombination is universal

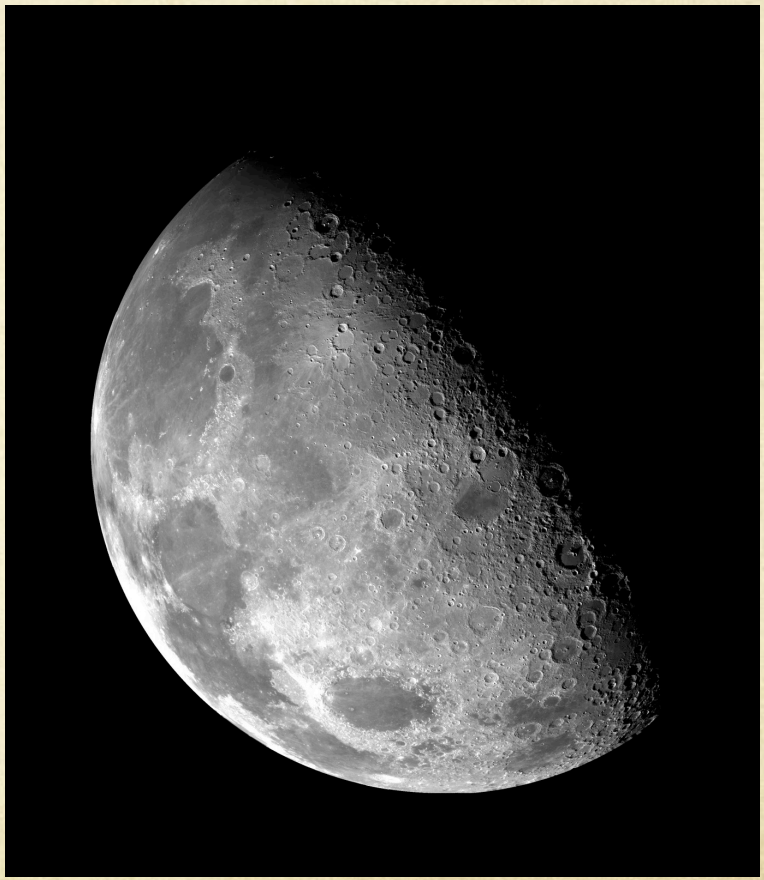


- Homologous recombination is present in most of prokaryotic organisms :
  - Acquisition of external DNA and integration within the genome by homologous recombination (horizontal gene transfer)
  - => gene flux among individuals within a population
  - => parasexuality
  - Non-recombining genomes (e.g. endosymbiotic bacteria) degenerate
- Ancient ameiotic eukaryotes (e.g. bdeloid rotifers): evidence of gene flux by homologous recombination via horizontal gene transfer

The impact of recombination on the evolution of organisms

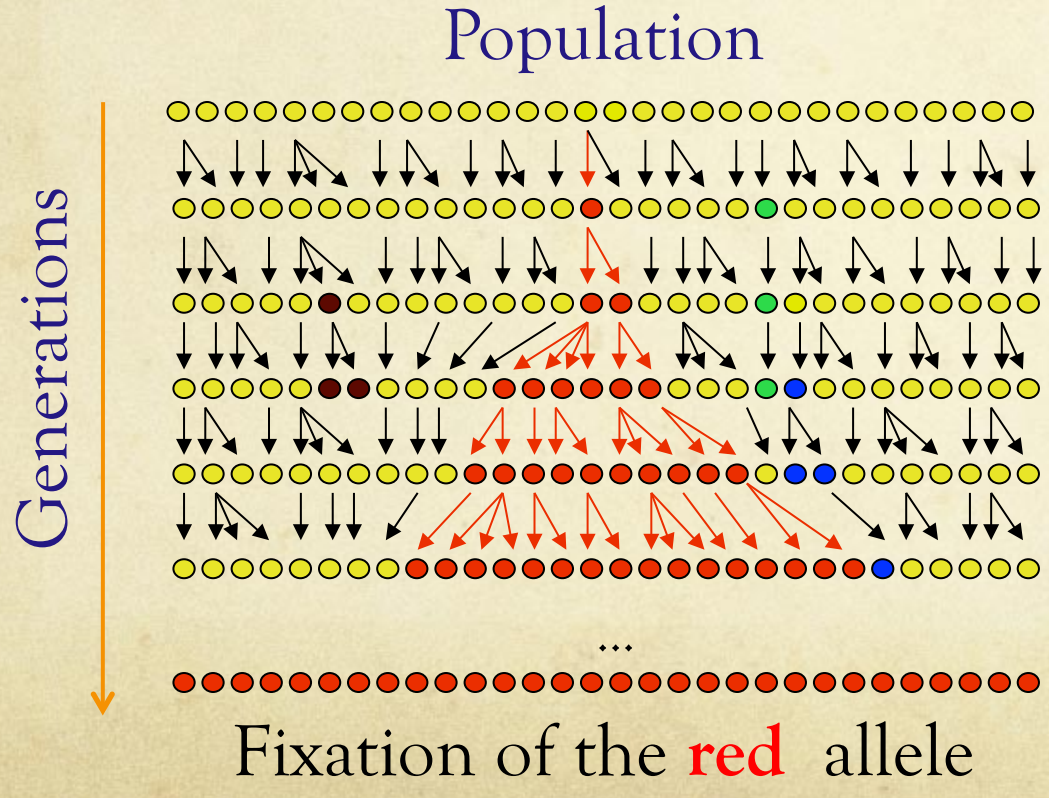
- Genetic linkage interferes with selection
- Recombination decreases linkage, and hence increases the efficacy of selection
- In the long-term, homologous recombination is essential to promote adaptation and limit the risk of species extinction

Biased gene conversion: the dark side of recombination



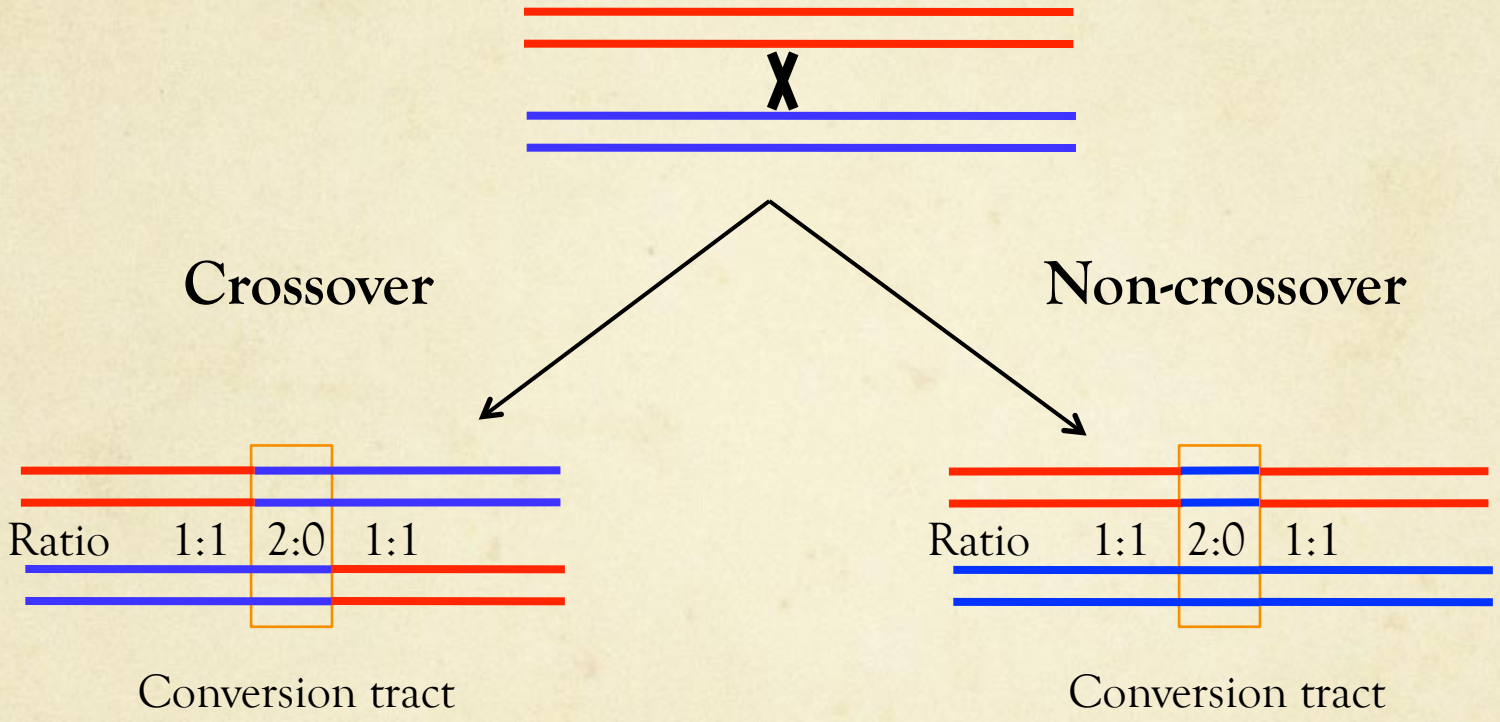
Evolution

- Mutation => new alleles
- Changes of allele frequencies over generations



- ✓ Genetic drift
- ✓ Natural selection
- ✓ Biased Gene Conversion

Biased gene conversion (BGC): the other facet of recombination



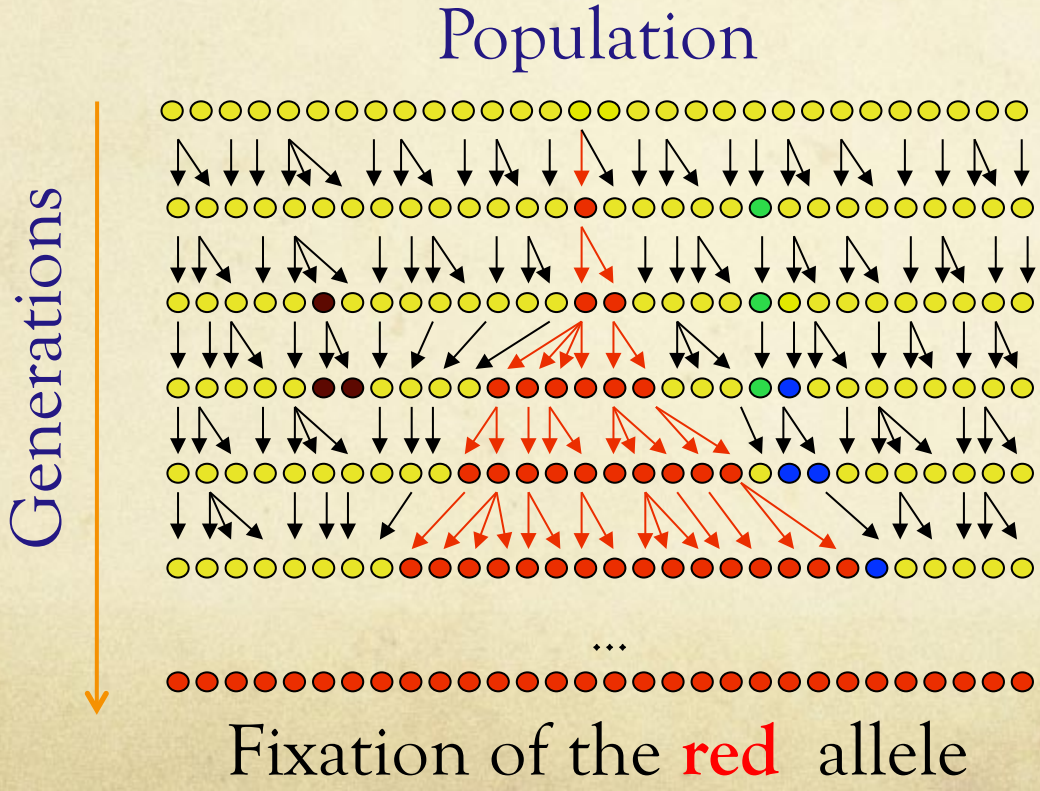
Gene conversion => non-mendelian transmission of alleles

If one allele has a higher probability to be the donor => biased gene conversion (BGC)

BGC increases the frequency of donor alleles in the pool of gametes => increases their probability of fixation in populations

Biased Gene Conversion

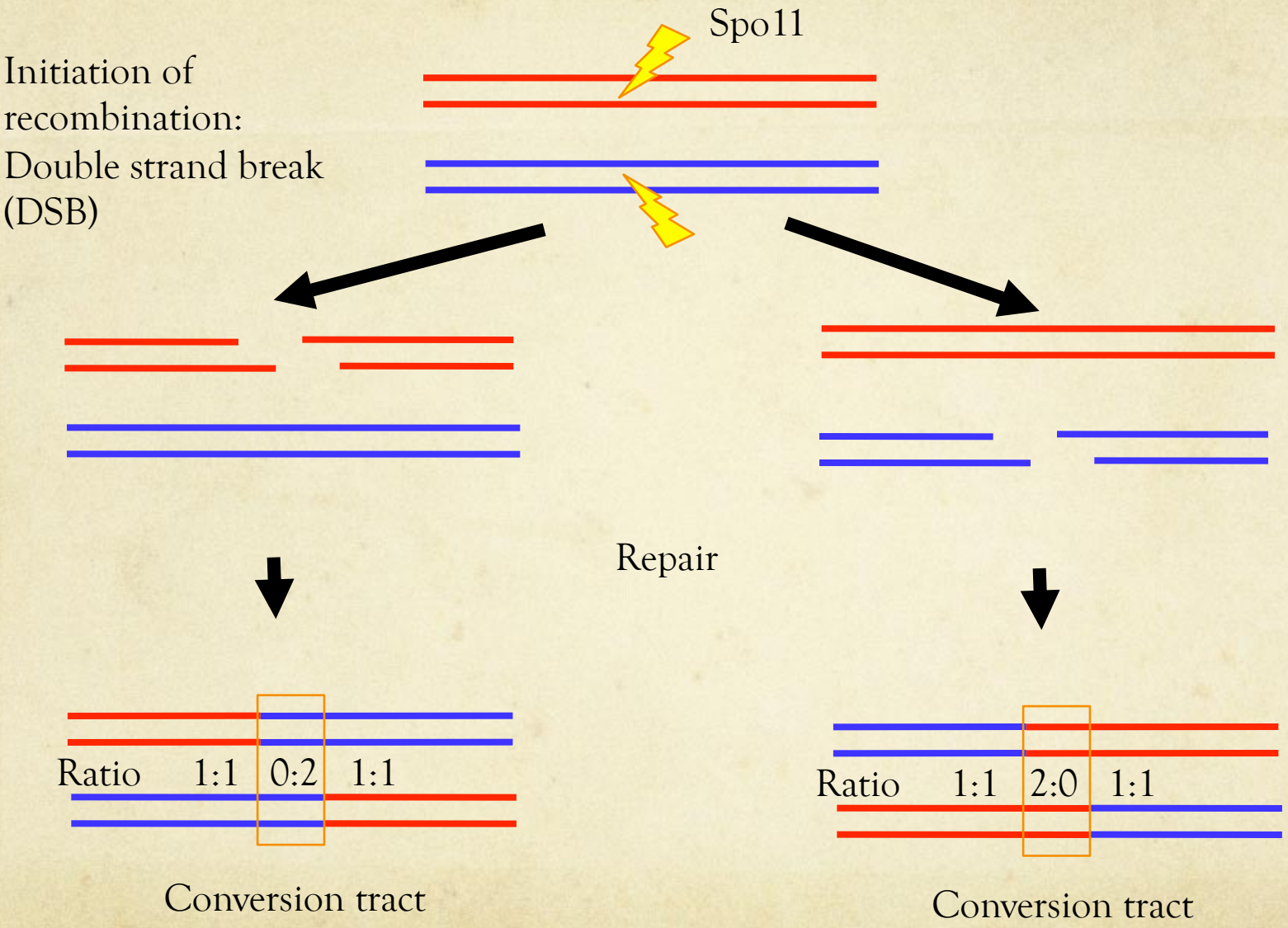
- A non-adaptive process that looks like selection
- Recombination affects allele frequencies in populations
- NOT a mutagenic effect of recombination



Biased Gene Conversion: an old story

- Lamb, B. C. and Helmi S. (1982) The extent to which gene conversion can change allele frequencies in populations. *Genet. Res.* **3 9** 199-217.
- Nagylaki T. (1983). Evolution of a finite population under gene conversion. *Proc. Natl. Acad. Sci. USA* 80:6278–681
- Walsh JB. (1983) Role of biased gene conversion in one-locus neutral theory and genome evolution. *Genetics.* 105:461-8.
- Bengtsson BO. (1986). Biased conversion as the primary function of recombination. *Genet. Res.* 47:77–80

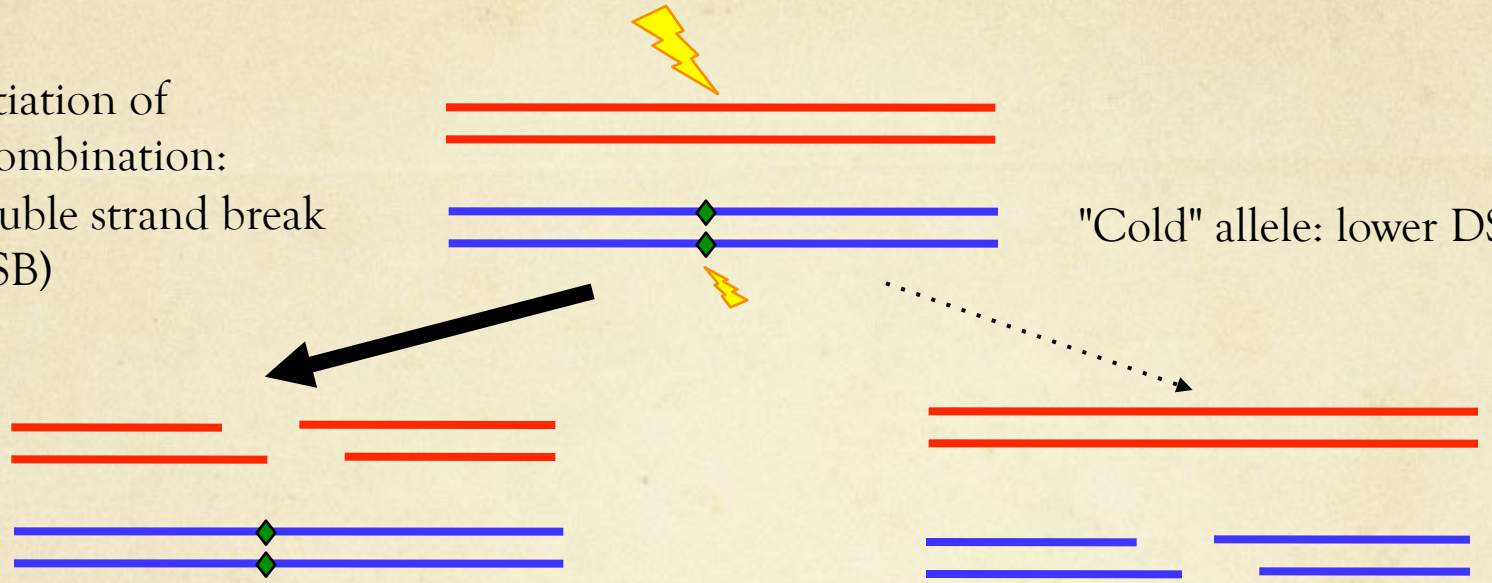
BGC mechanisms (1): initiation bias



At the population scale, if both haplotypes have the same rate of DSB formation, then they have the same probability to be transmitted to the next generation

BGC mechanisms (1): initiation bias

Initiation of recombination: Double strand break (DSB)



"Cold" allele: lower DSB frequency

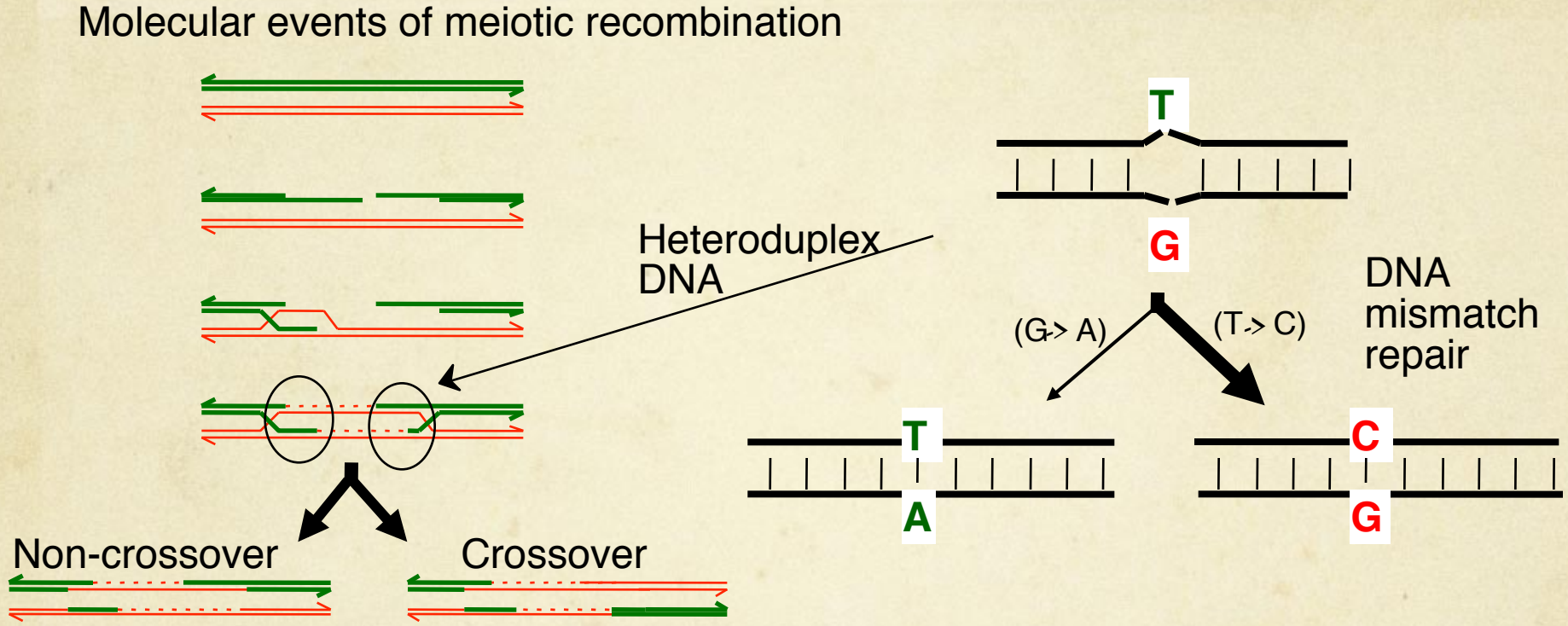
BGC in favor of cold alleles

Conversion tract

Conversion tract

Myers et al. 2010, Science

BGC mechanisms (2): mismatch repair bias



BGC in favor of GC alleles

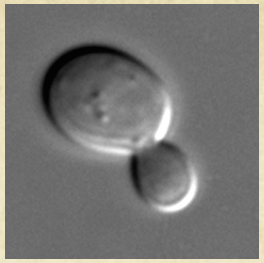
Lesecque et al. (2013) Mol Biol Evol

Two types of BGC

- Initiation bias: BGC in favor of cold alleles (i.e. alleles with lower rate of DSB formation, i.e. alleles with low recombination activity)
  - dBGC: DSB-induced Biased Gene Conversion

- Mismatch repair bias: BGC favors GC-alleles over AT-alleles
  - gBGC: GC-Biased Gene Conversion

gBGC in yeast



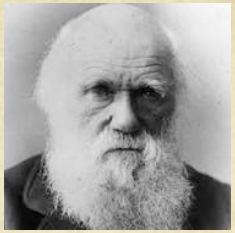
- Mancera et al. (*Nature* 2008): high resolution mapping of meiotic recombination products in yeast
- >6000 recombination events
- Gene conversion tracts involving GC/AT heterozygotes
- Gamete frequency expected in absence of BGC:

  freq. GC = freq. AT = 50%
- Observed gamete frequency:

  freq. GC=50.7% AT=49.3%

=> gBGC increases the frequency of GC alleles in populations => increases their probability of fixation

gBGC in human



- Williams et al. (*eLife* 2015): analysis of non-crossover in human pedigrees
- 103 recombination events
- Gene conversion tracts involving GC/AT heterozygotes
- Gamete frequency expected in absence of BGC:

  freq. GC = freq. AT = 50%
- Observed gamete frequency:

  freq. GC=68% AT=32%

gBGC in birds

- Smeds et al. (*Plos Genet* 2016): analysis of non-crossover in flycatcher pedigrees
- 229 recombination events
- Gene conversion tracts involving GC/AT heterozygotes
- Gamete frequency expected in absence of BGC:

  freq. GC = freq. AT = 50%
- Observed gamete frequency:

  freq. GC=59% AT=41%

Impact of gBGC on genome evolution ?

- The intensity of gBGC depends on:
  - b_0 : transmission bias (= mismatch repair bias)
    - $F_{GC} = ½(1+b_0)$
    - $F_{AT} = ½(1-b_0)$
    - *Example: human:*
      - $F_{GC} = 0.68$, $F_{AT} = 0.32$
      - $b_0 = 0.36$
  - r : probability to be involved in a gene conversion event (recombination rate, conversion tract length)
  - N_e : effective population size

Impact of gBGC on genome evolution ?

- Nagylaki (1983): gBGC behaves just like selection of a semidominant mutation
  - $b = b_0 \times r$: gBGC coefficient (<=> selection coefficient s)
  - N_e : effective population size

- The probability of fixation of AT→GC mutations is

$$P(AT \rightarrow GC) = \frac{1 - e^{-2b}}{1 - e^{-4Neb}}$$

- The probability of fixation of GC→AT mutations is

$$P(GC \rightarrow AT) = \frac{1 - e^{2b}}{1 - e^{4Neb}}$$

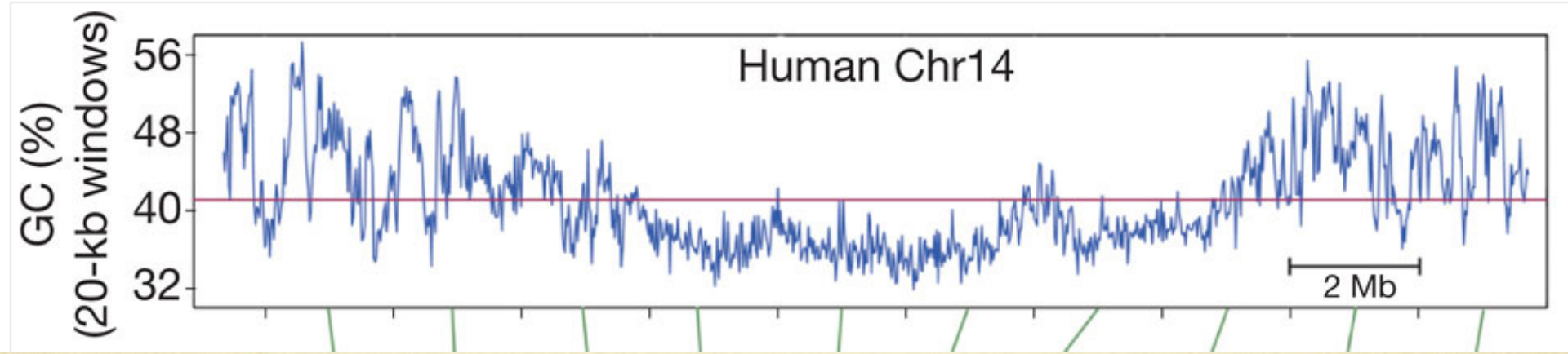
Impact of gBGC on genome evolution ?

- Population-scaled BGC coefficient: $B = 4\,N_e \times b$
  - $|B| \ll 1$: random genetic drift (neutral evolution)
  - $|B| \gg 1$: strong BGC
  - $|B| \sim 1$: nearly-neutral area

- **Loci under selective pressure** (S):
  - if $|B| \geq |S|$, **gBGC interferes with selection**, and can drive the fixation of deleterious alleles

Does gBGC affect genome evolution in mammals?

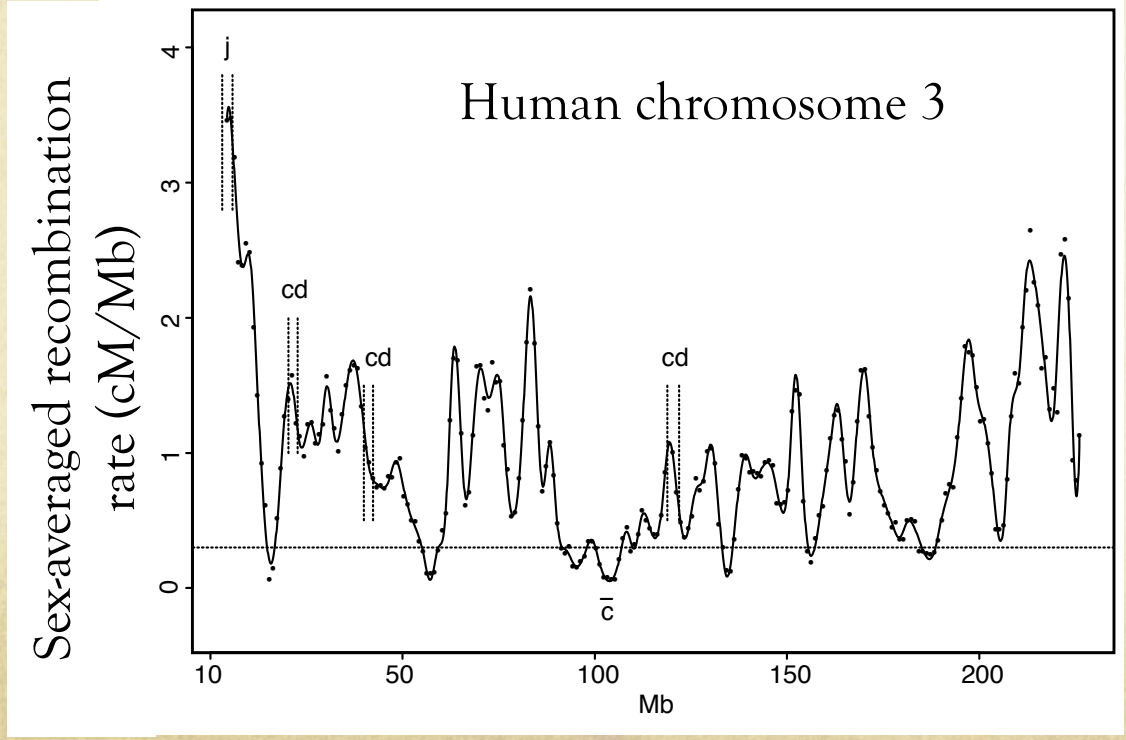
- Large-scale variation in GC-content along chromosomes (isochores) (Bernardi et al. 1985, *Science*)



Alföldi et al (2011) Nature 477:587–591

Does gBGC affect genome evolution in mammals?

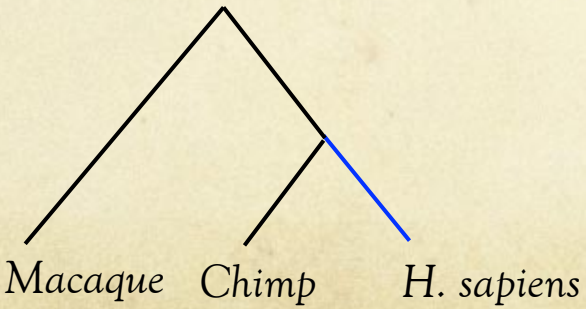
- Large-scale variation in recombination rate along chromosomes



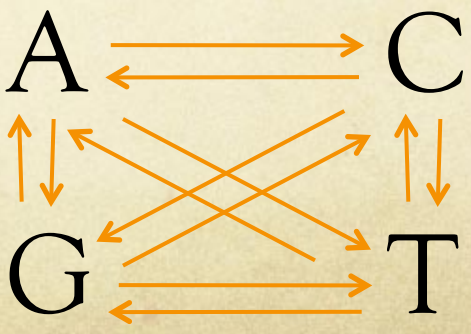
Kong et al. (2002)
Nat Genet. 31: 241–7

Does gBGC affect genome evolution in mammals?

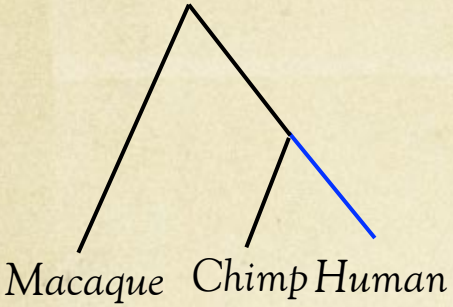
- Relationship between recombination rate and the evolution of GC-content ?
- Analysis of substitution patterns at neutral sites along the human lineage, since the divergence from chimpanzee



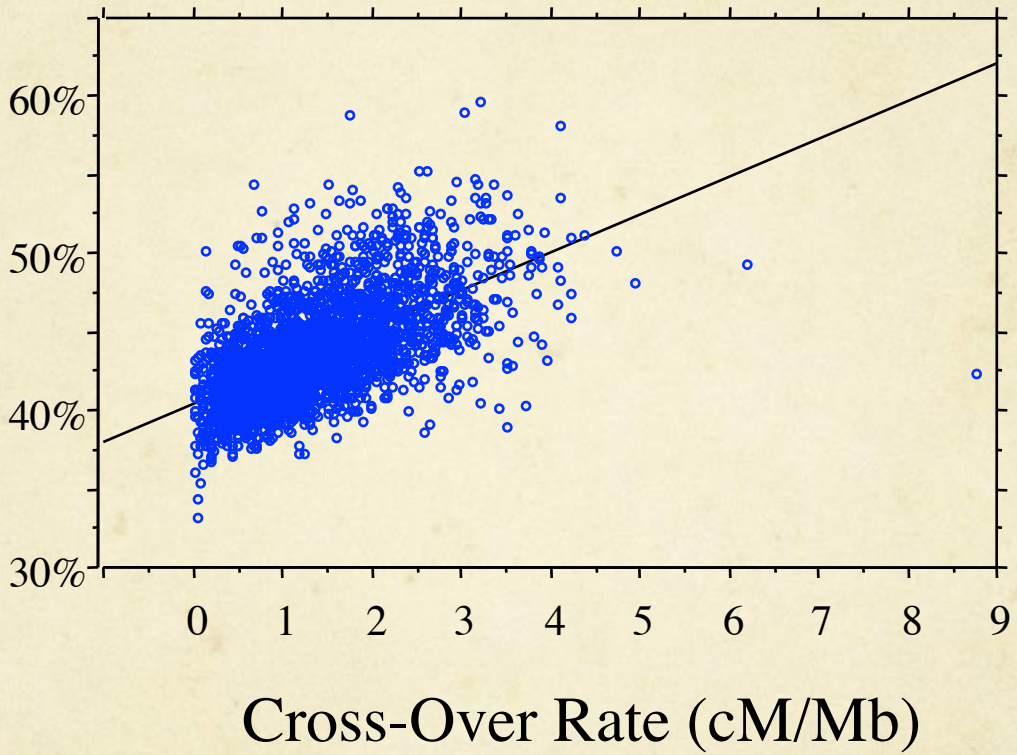
Whole genome alignments



Equilibrium GC-content and recombination



Equilibrium GC-content GC*



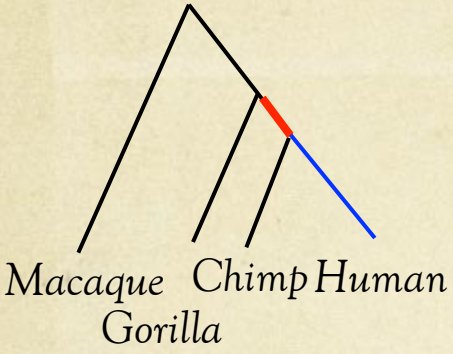
$R^2 = 36\%$

$p < 0.0001$

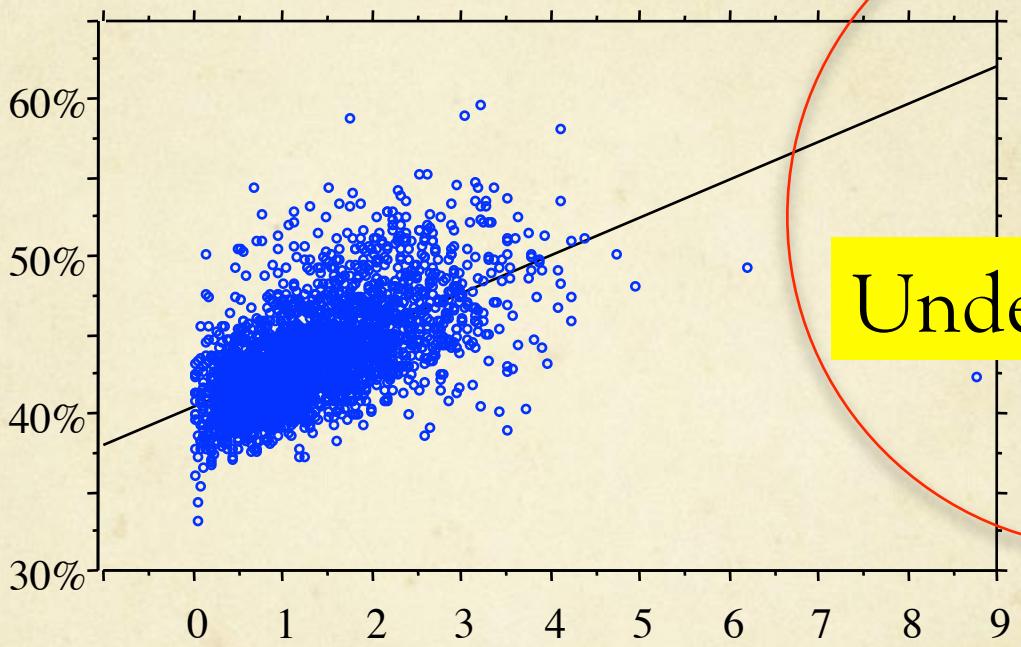
N = 2707 non-overlapping windows (1 Mb), from autosomes

Duret & Arndt (2008) Plos Genet

Equilibrium GC-content and recombination



Equilibrium GC-content GC*



$R^2 = 36\%$

$p < 0.0001$

Underestimate !

Munch K, Mailund T, Dutheil J, Schierup M. (*Genome Res.* 2013):

GC* vs crossover rate in human-chimp ancestral lineage:

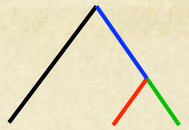
R^2=64%

GC-content and Recombination

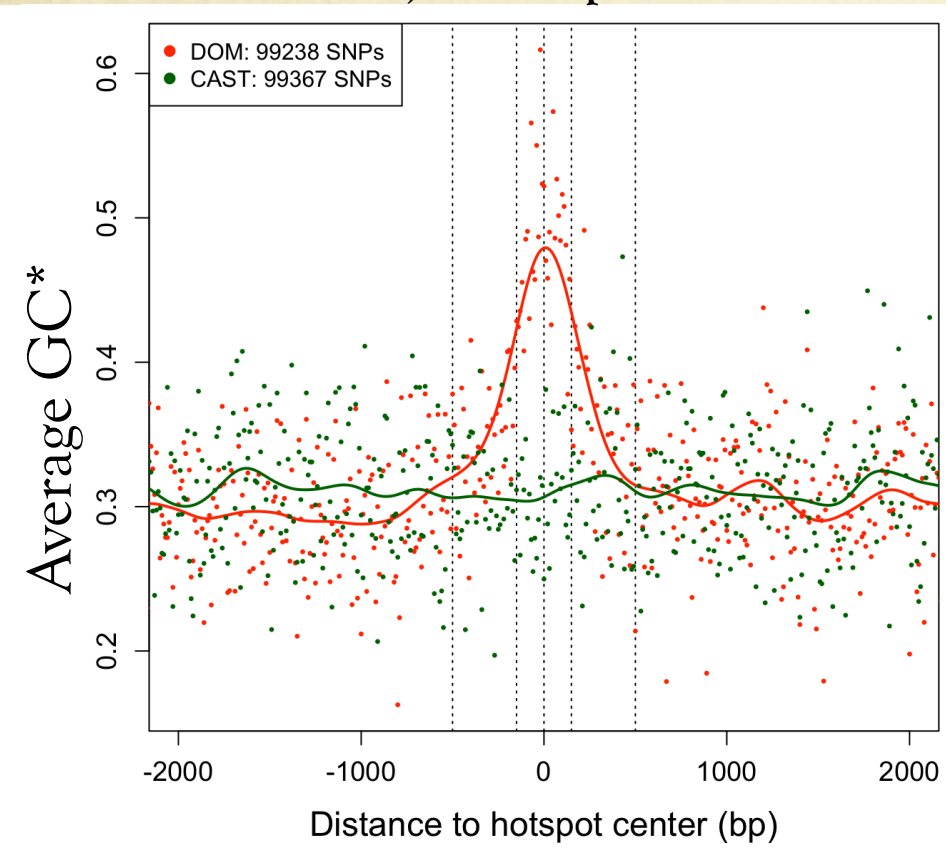
- Strong correlation: suggests direct causal relationship
- GC-rich sequences promote recombination ?
  - Gerton et al. (2000), Petes & Merker (2002), Spencer et al. (2006)
- Recombination promotes AT→GC substitutions ?

Fine scale, short term: substitutions patterns at recombinations hotspots

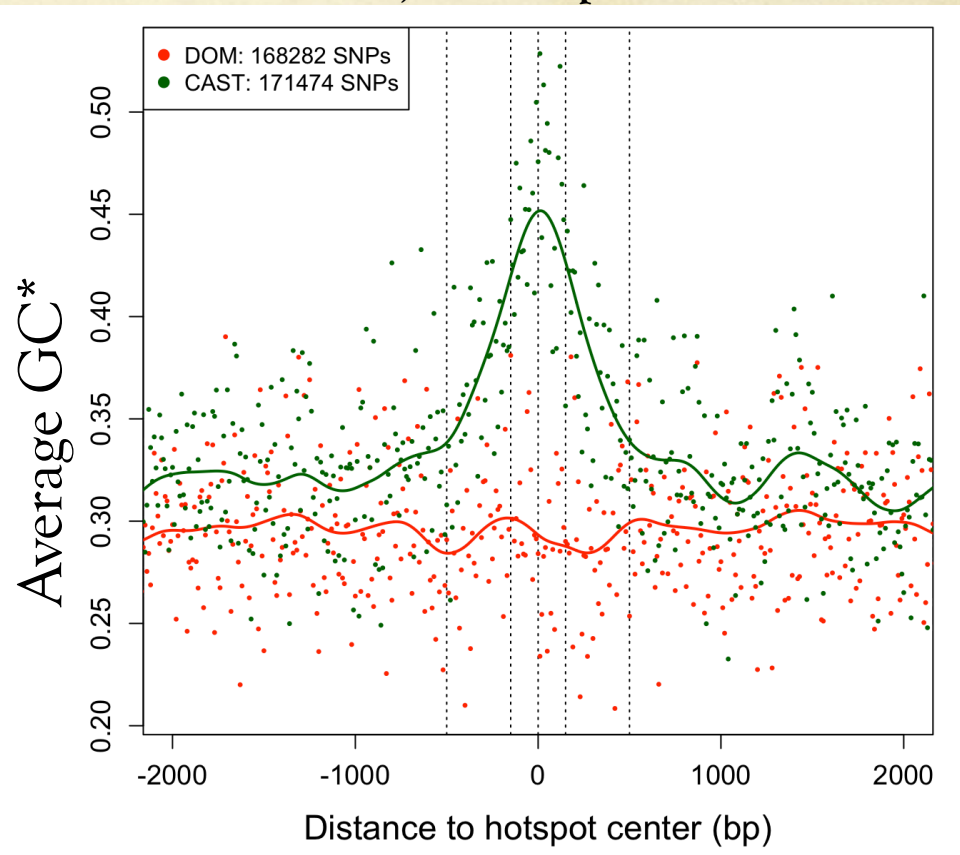
- *Mice sub-species: M. m. domesticus/castaneus:*
  - 0.5–1.0 million years of divergence, 99.2% sequence identity
  - Different sets of recombination hotspots (Baker et al. 2015)



DOM : N=3,244 hotspots



CAST : N=5,526 hotspots



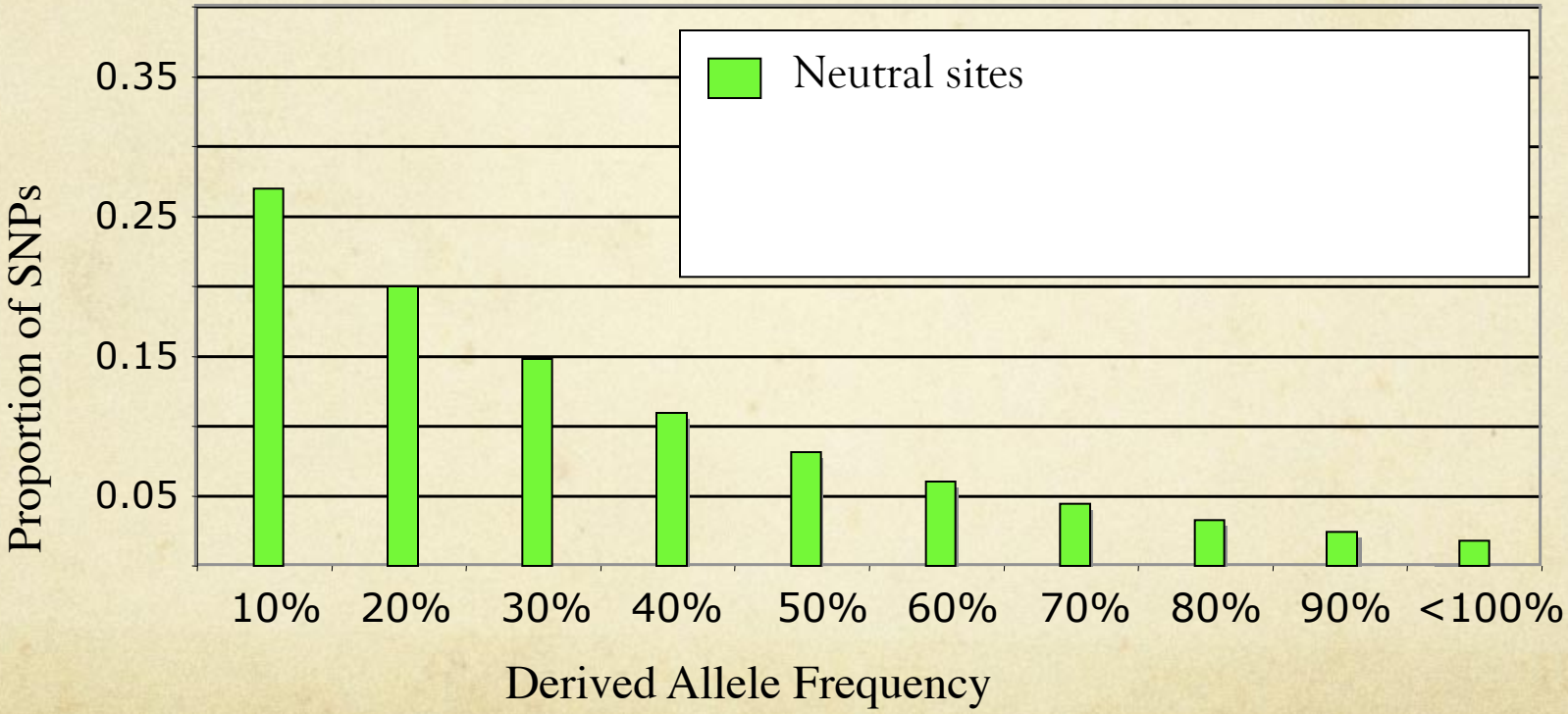
GC-content and Recombination

- Strong correlation: suggests direct causal relationship
- GC-rich sequences promote recombination ?
  - Gert[illegible] et al. (2000), Petes & Merker (2002), Spencer et al. (2006)
- Recombination promotes AT→GC substitutions ?

Impact of recombination on the evolution of GC-content: Mutagenic effect or fixation bias?

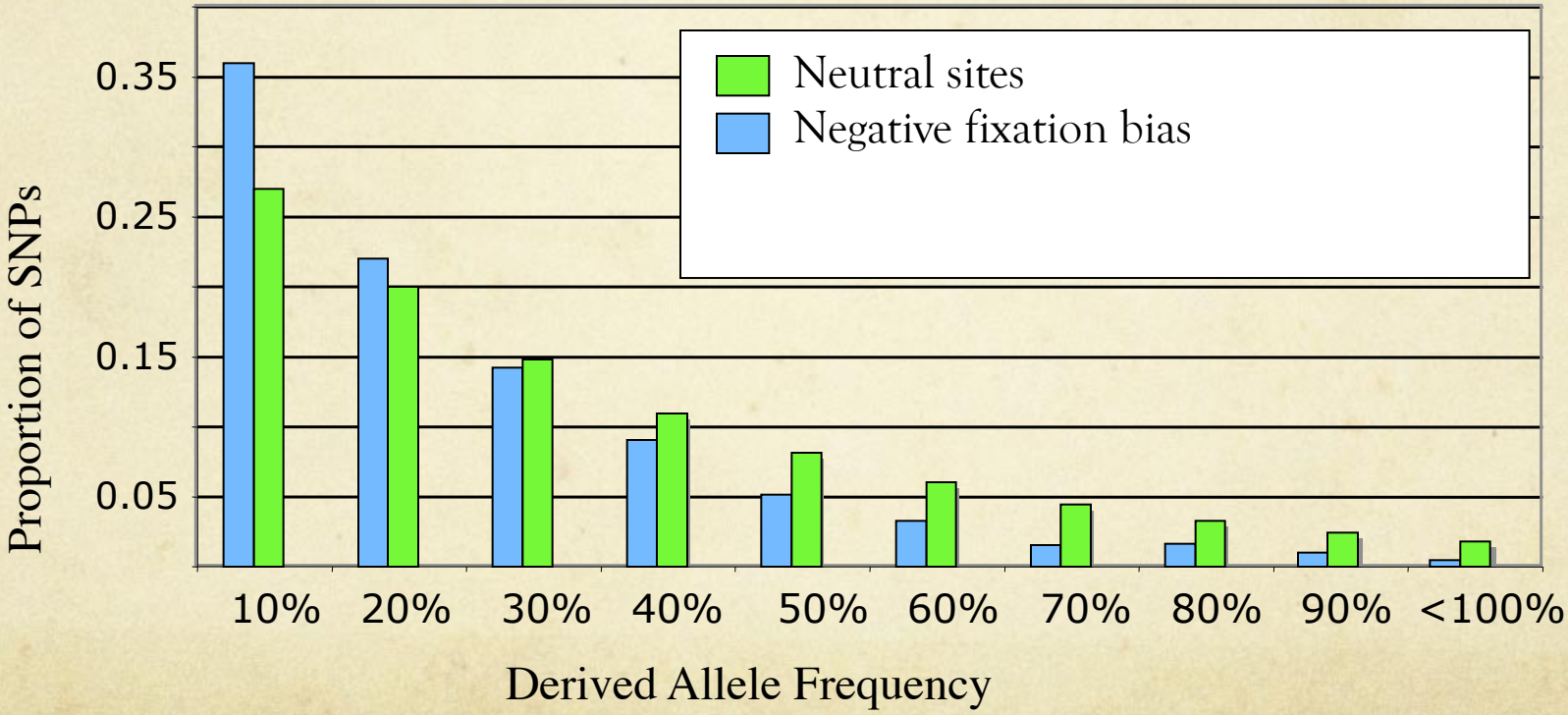
Detecting fixation bias by analysis of polymorphism data

- *Derived allele frequency spectrum*



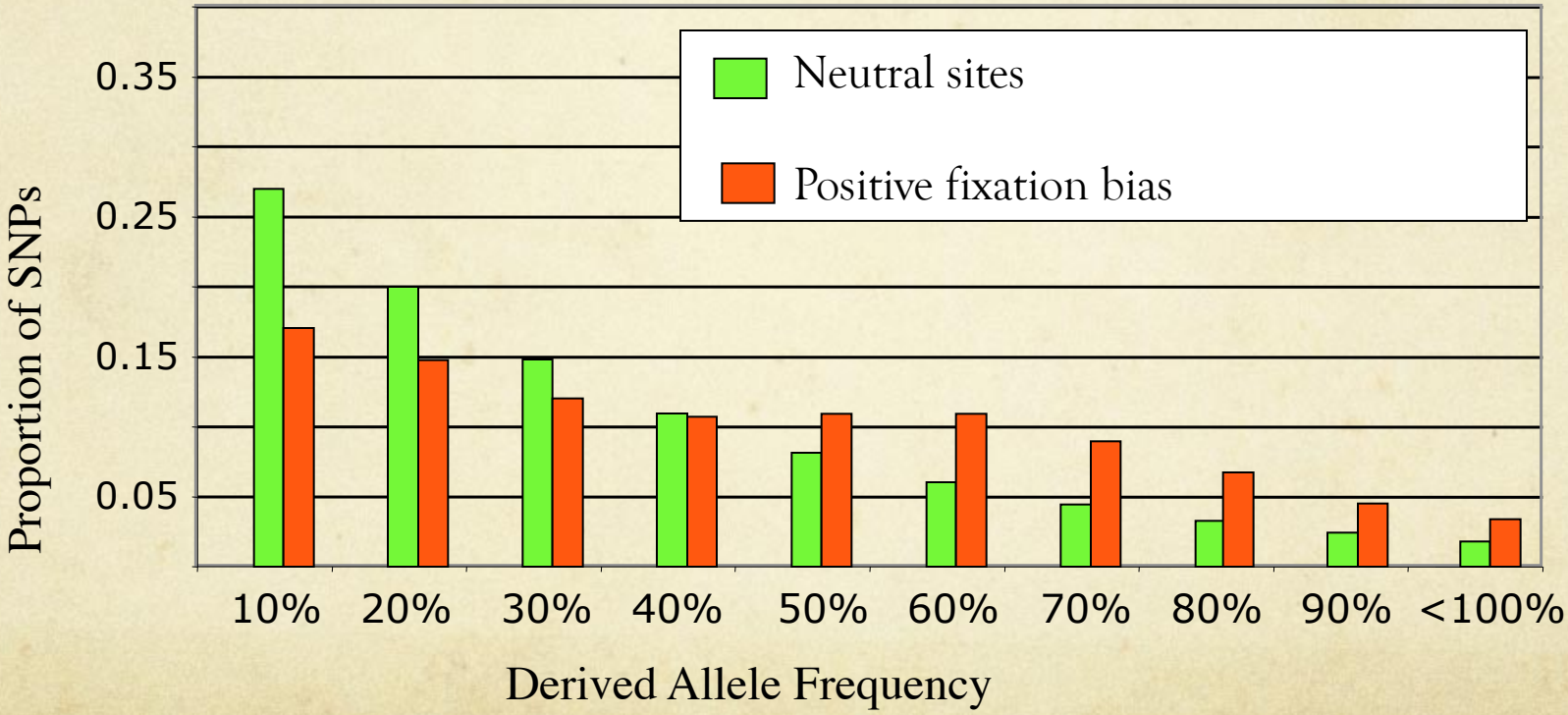
Detecting fixation bias by analysis of polymorphism data

- *Derived allele frequency spectrum*



Detecting fixation bias by analysis of polymorphism data

- *Derived allele frequency spectrum*



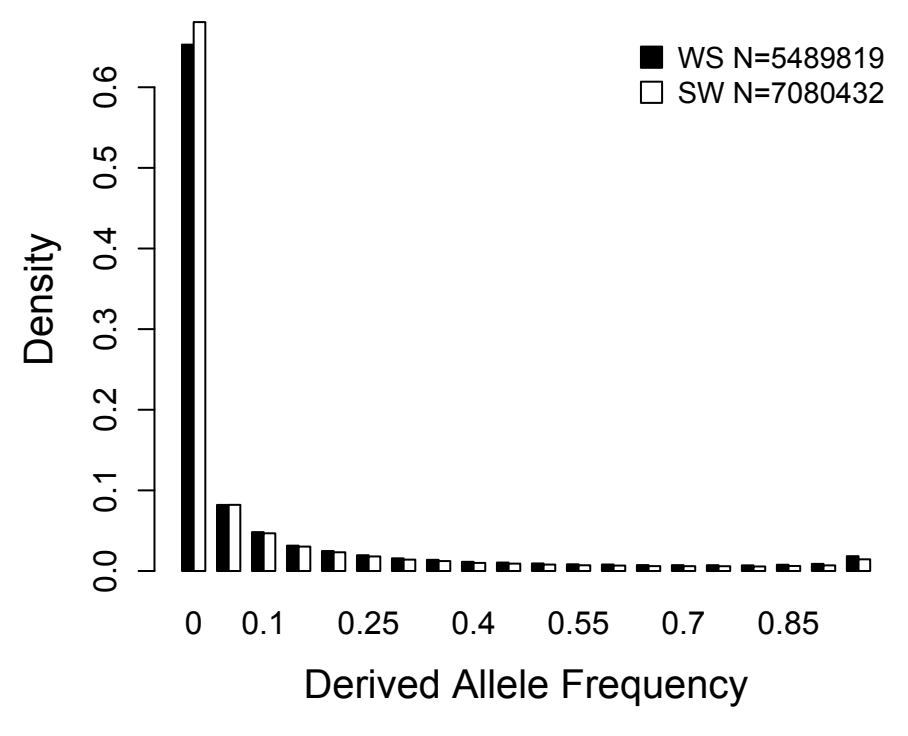
Mutagenic effect of recombination or fixation bias?

- 1000 Genomes Project SNP dataset:
  - Ancestral state inferred by comparison with outgroups (chimp, orangutan, macaca)
  - Analysis of derived allele frequency spectra

12×10^6 non-CpG WS/SW SNPs

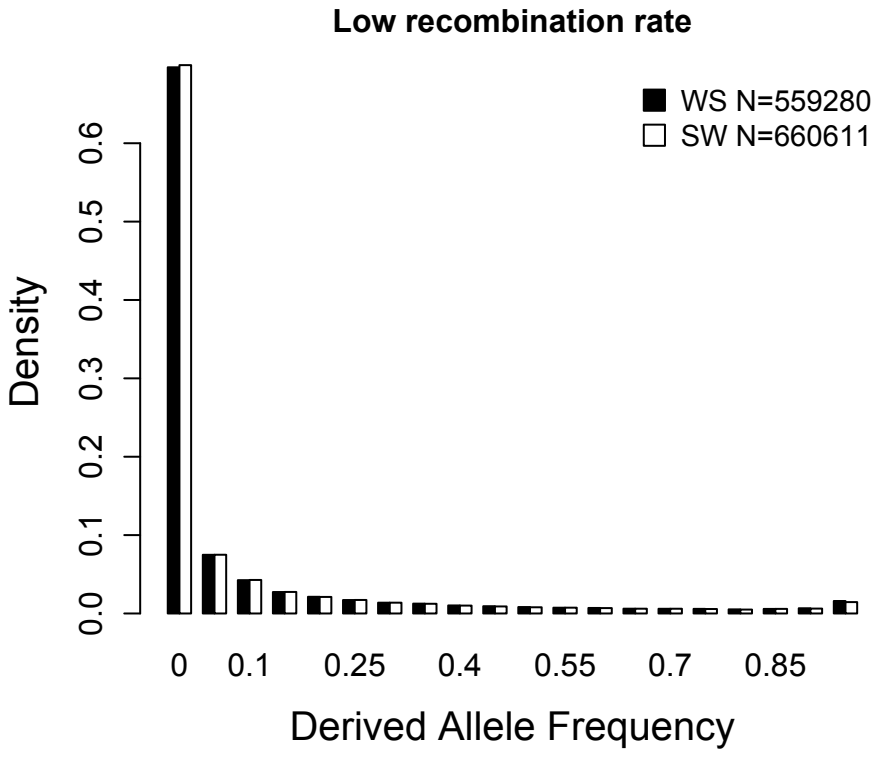
Non-coding regions

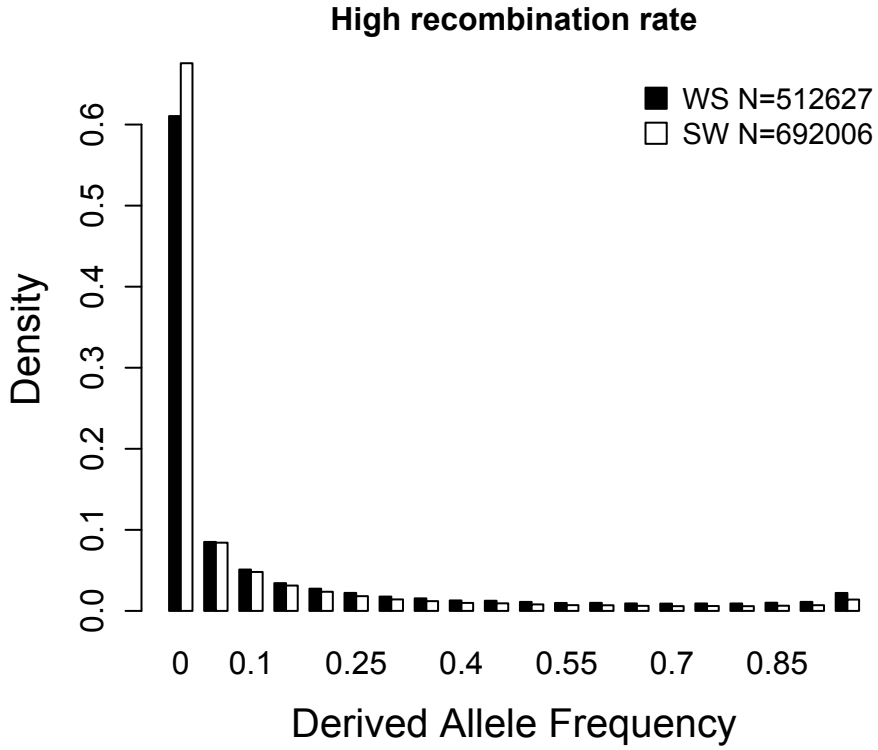
W (weak) = A or T
S (strong) = G or C



Mutagenic effect of recombination or fixation bias?

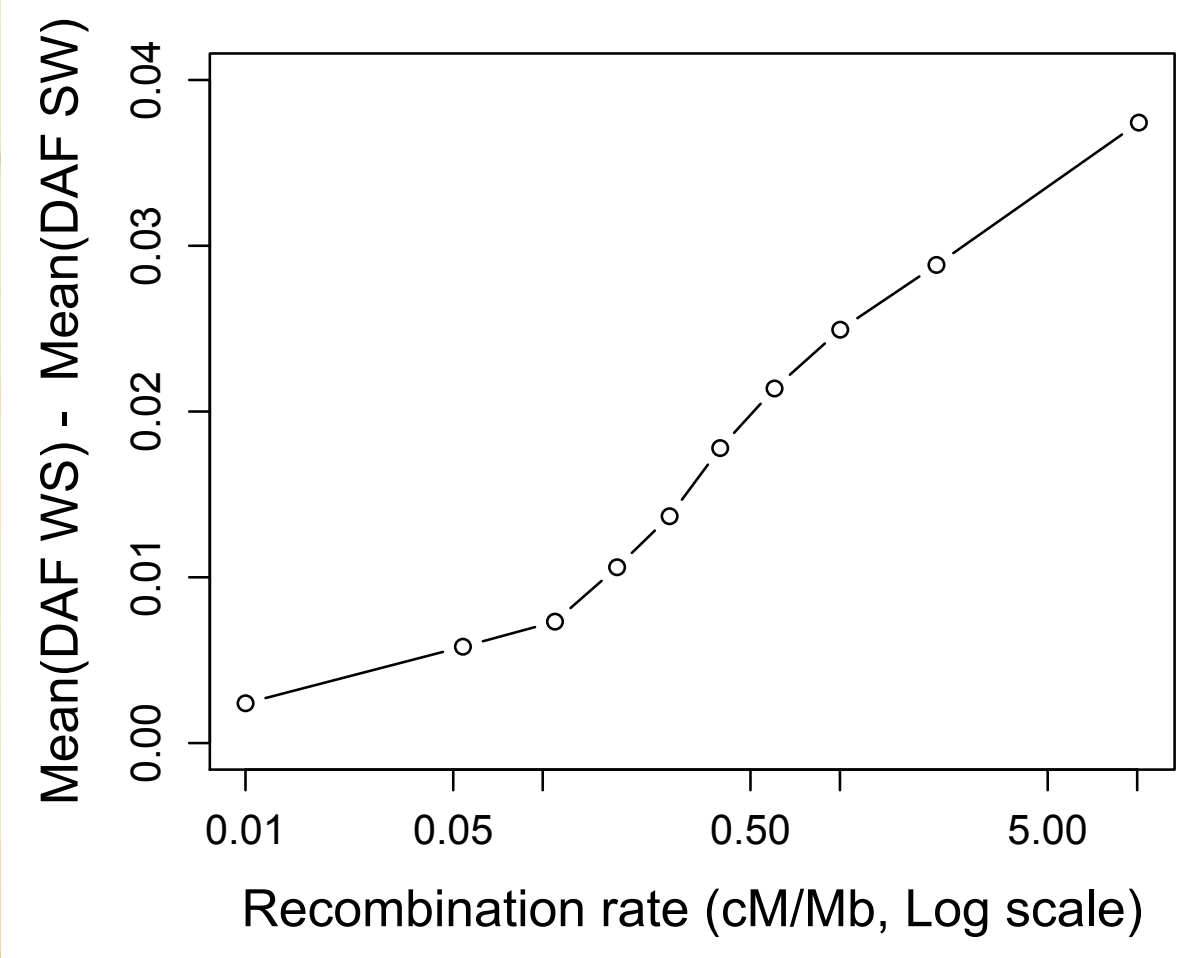
Analysis of derived allele frequency spectra





=> In regions of high recombination, GC alleles have a higher probability of fixation than AT alleles

The fixation bias in favor of GC-alleles increases with recombination



Evidence of gBGC in humans

- Analysis of non-crossovers in human pedigrees: transmission bias in favor of GC alleles (68:32)
- In non-coding regions (presumably neutral):
  - The evolution of GC-content is driven by recombination
  - GC-alleles segregate at higher frequency than AT-alleles
  - This fixation bias increases with recombination rate

Quantifying gBGC in the human genome

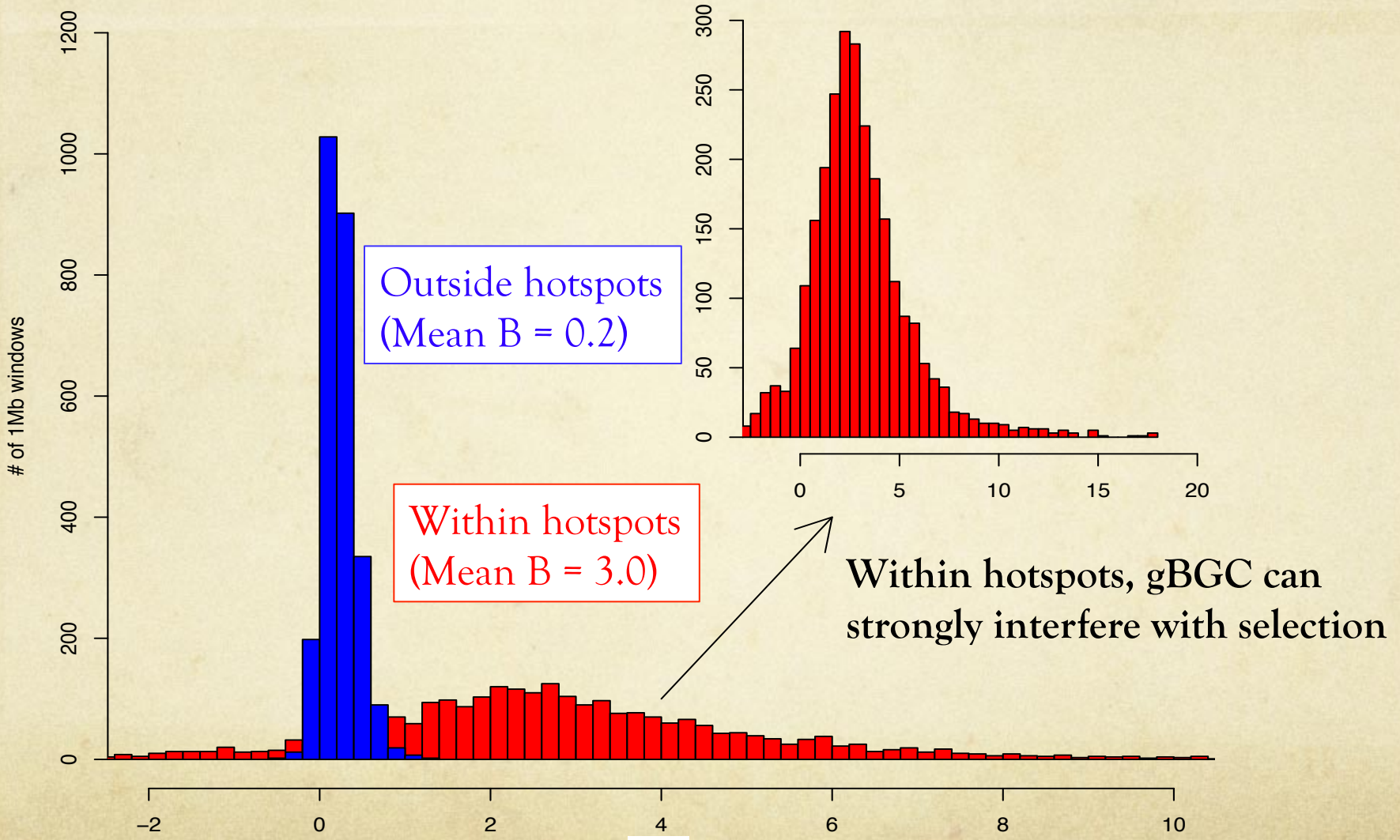
Sylvain Glémin

- Quantify the strength of gBGC from the comparison of the DAF spectra of WS and SW mutations
- Maximum likelihood framework, inspired by Eyre-Walker, Woolfit & Phelps (2006): quantify selection on non-synonymous mutations
- Using WW and SS SNPs as neutral markers (i.e. not affected by gBGC) to take demographic history into account
- Model parameters:
  - **Population-scaled gBGC coefficient ($B = 4 N_e b$)**
  - Mutational bias (WS/SW mutation rate)
  - **SNP polarization error rates**
- Accounting for heterogeneity of recombination rates (recombination hotspots)

Glémin et al. *Genome Res.* (2015)

Quantification of gBGC

- Estimate parameters in 1-Mb long genomic windows



Evidence of gBGC at recombination hotspots in birds

- Singhal et al. Science (2015) 350: 928–932. doi:10.1126/science.aad0843

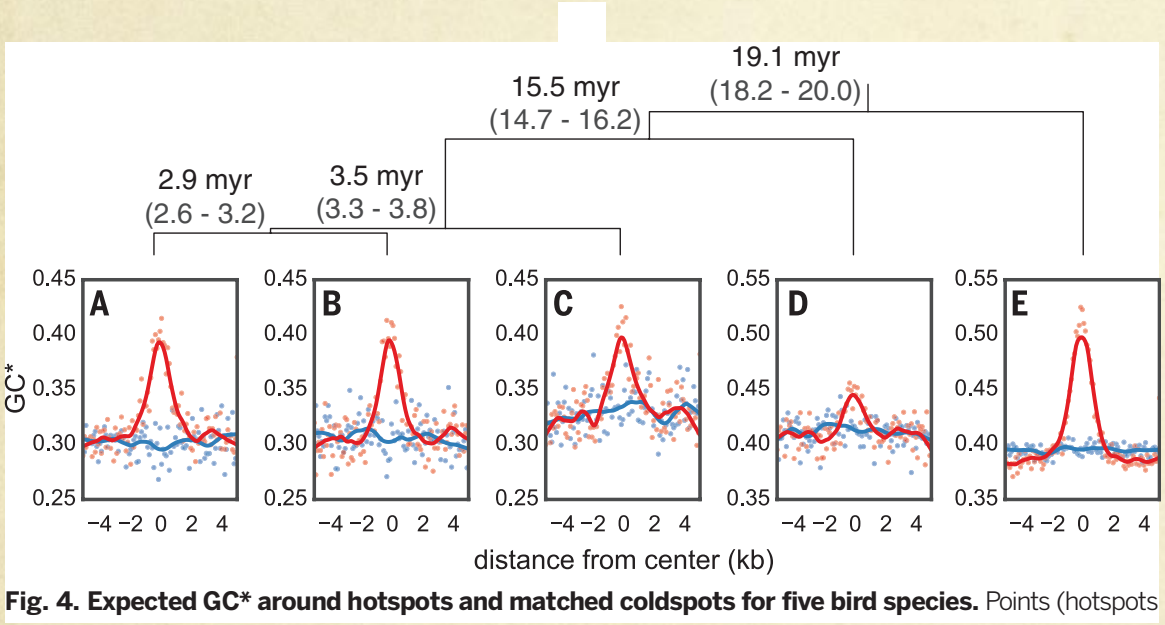
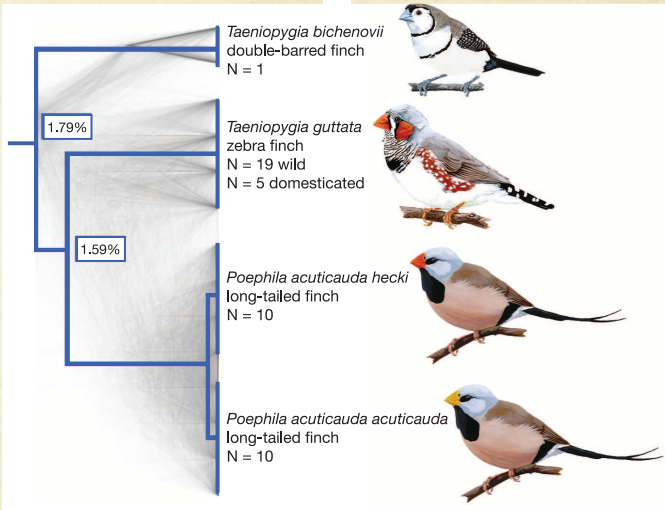
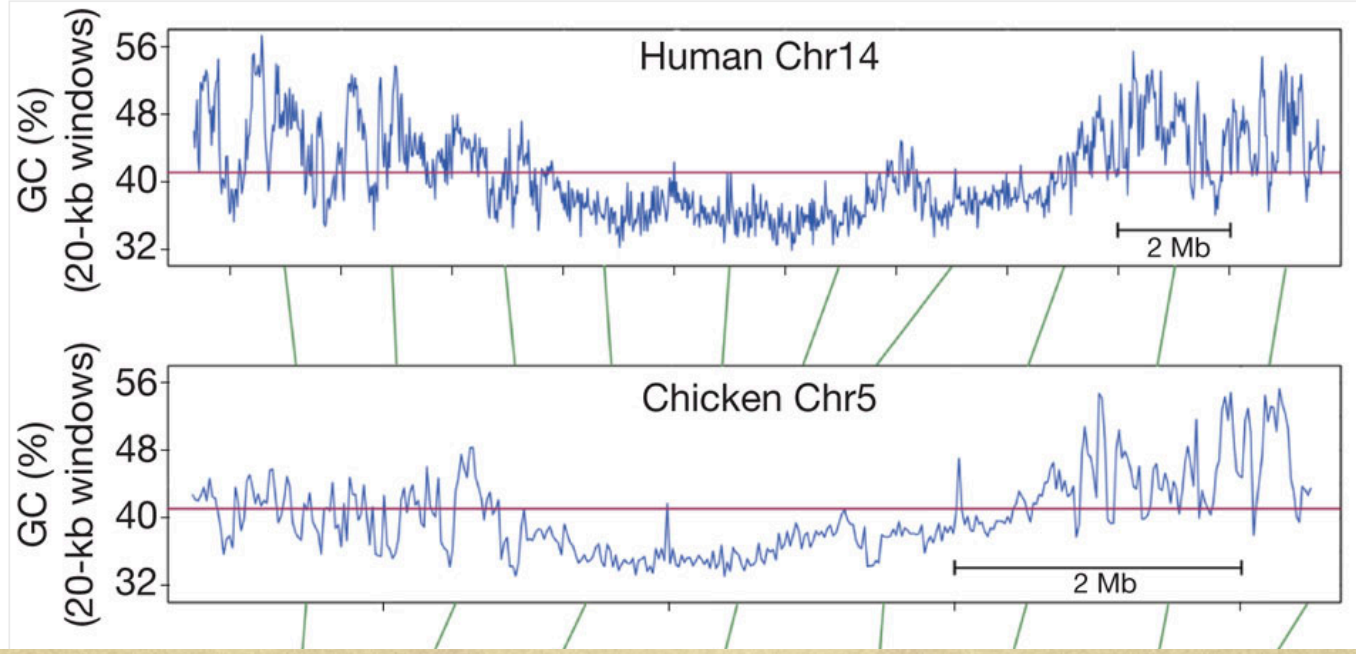


Fig. 4. Expected GC* around hotspots and matched coldspots for five bird species. Points (hotspots

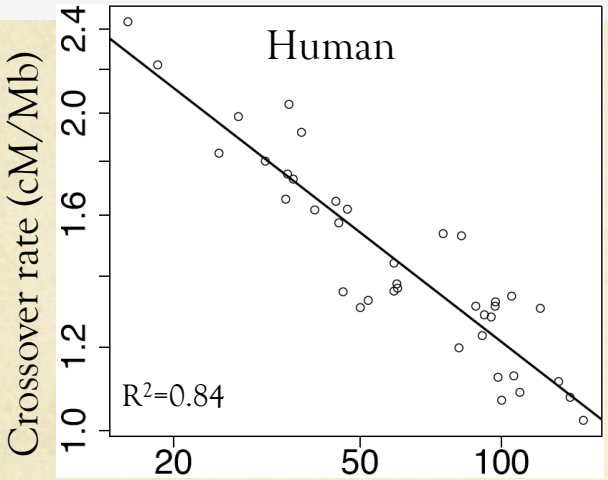


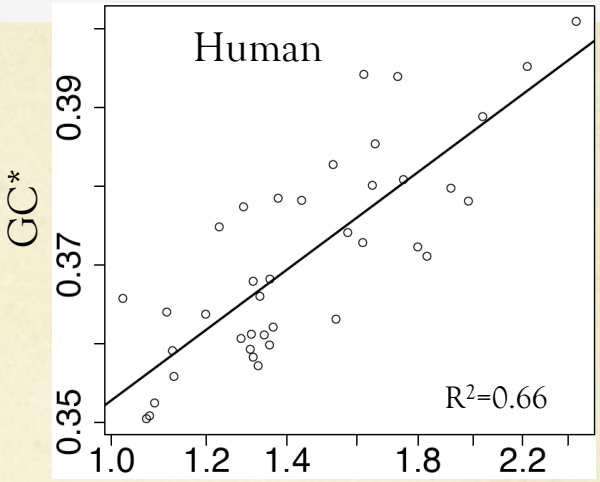
Summary

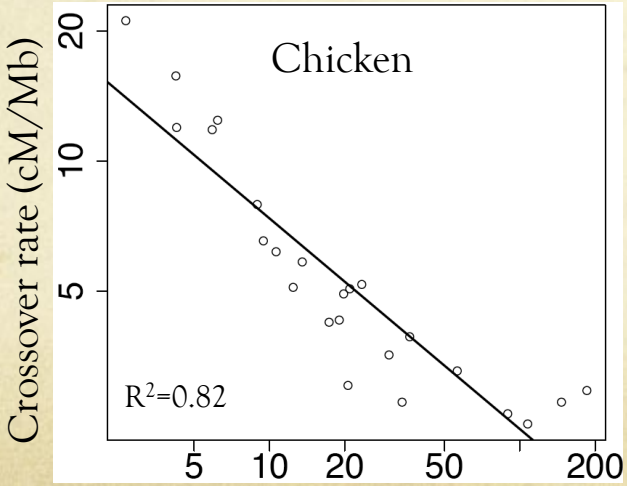
- Recombination hotspots are subject to strong gBGC (human, mice, birds)
- The density and intensity of hotpots varies along chromosomes => large-scale variations in recombination rates
- Major impact on the evolution of genomic landscapes in amniotes

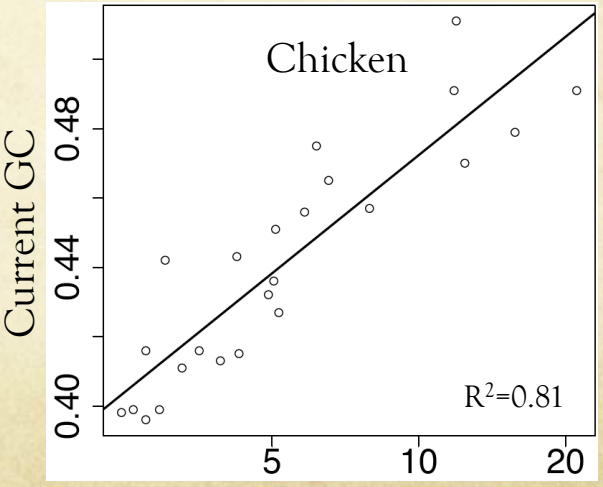


Chromosome size, recombination and GC-content



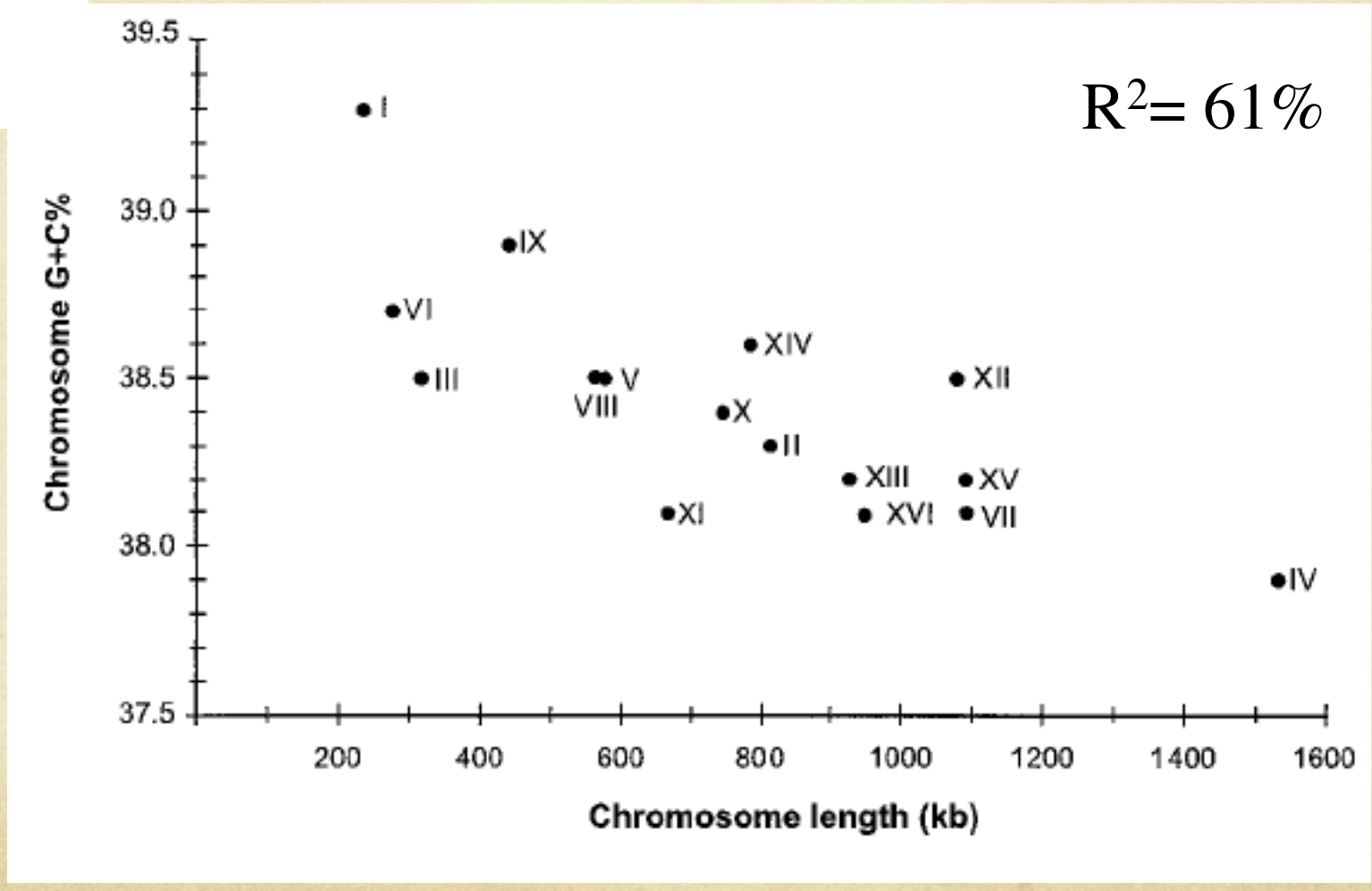






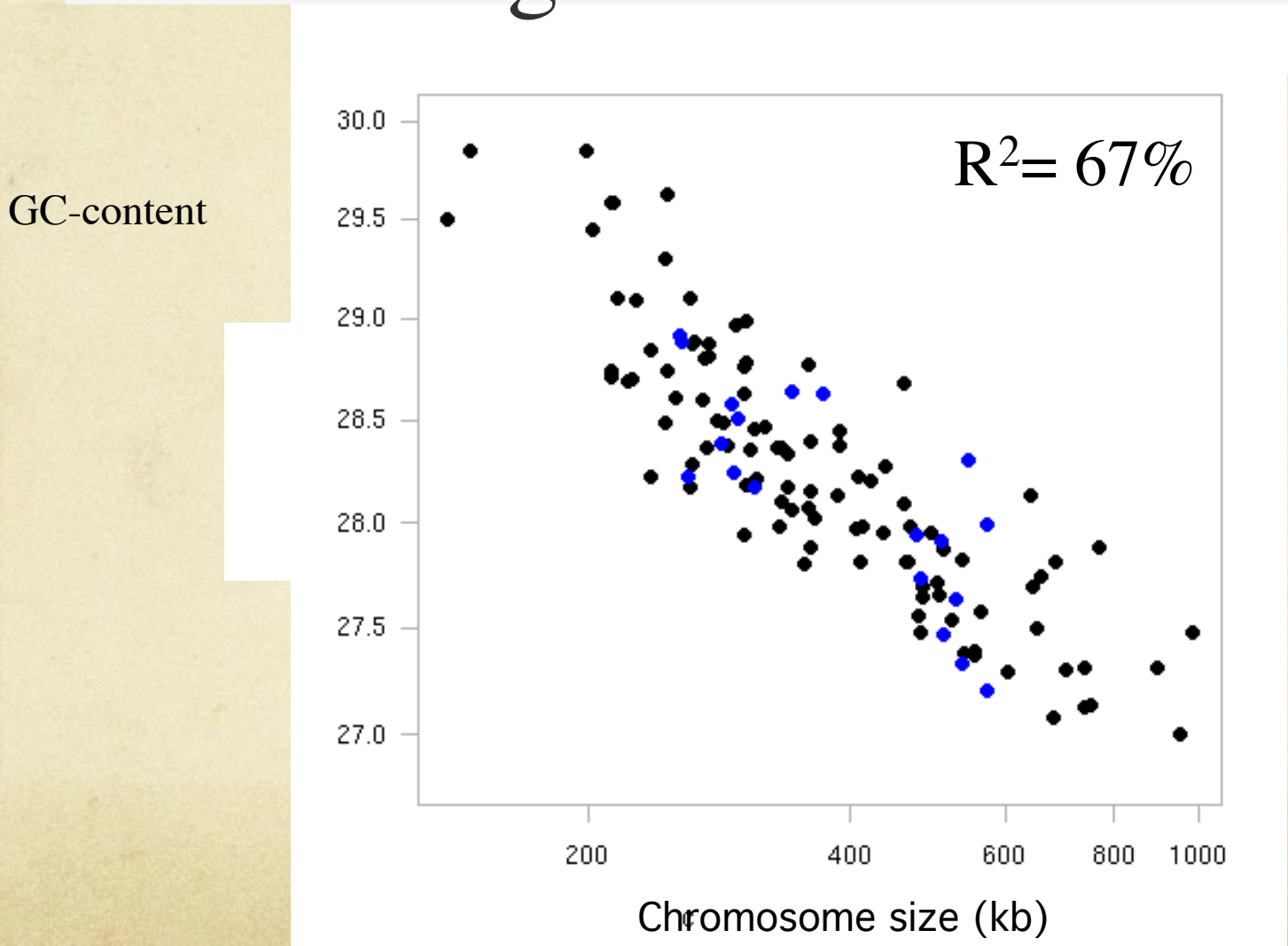
Recombination and GC-content: a universal relationship ?

G+C content *vs*. chromosome length: yeast



Bradnam et al. (1999) Mol Biol Evol

G+C content *vs*. chromosome length: *Paramecium*



Correlation between recombination rate and GC-content within eukaryotic genomes

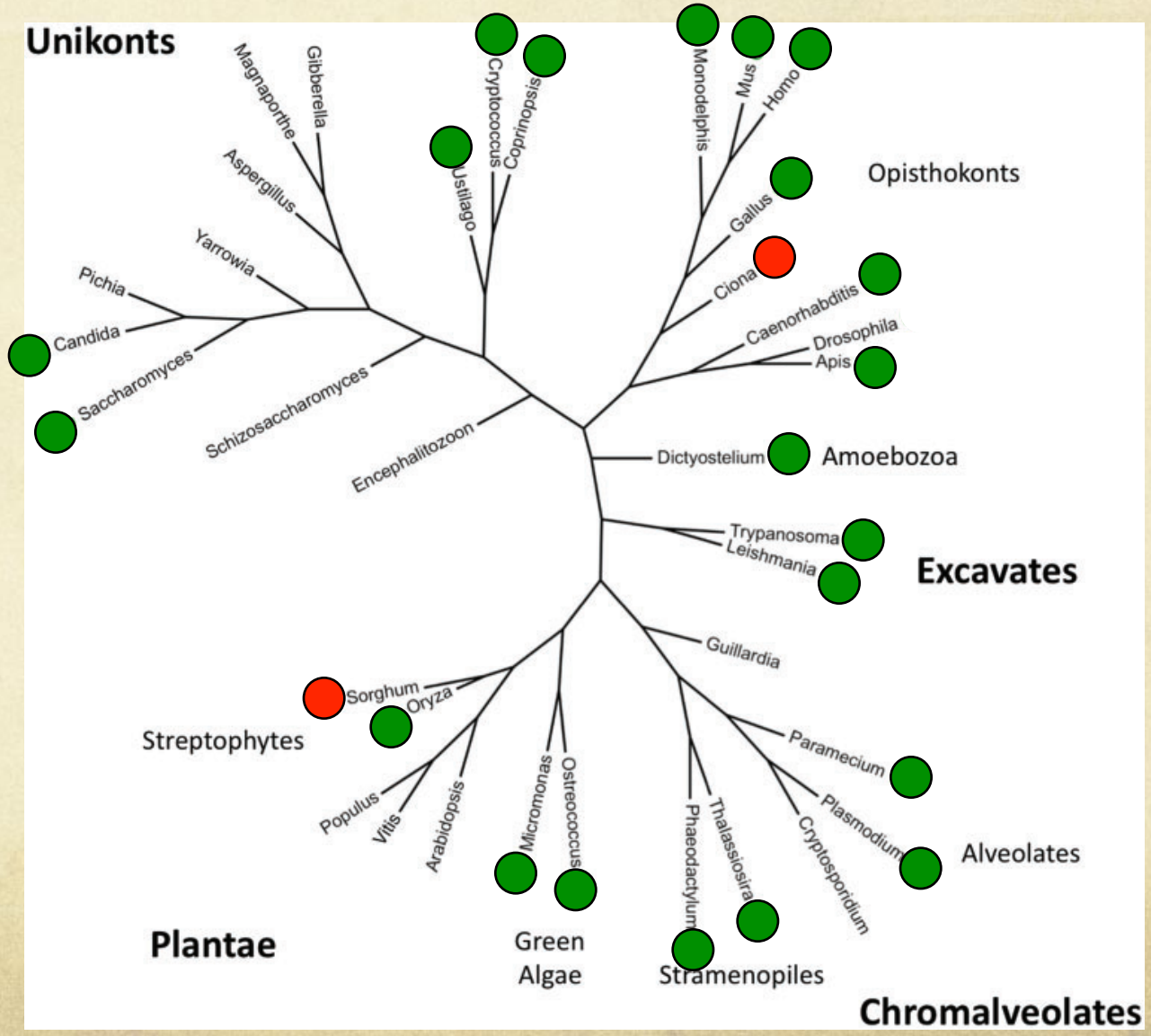
36 species with complete genomes and estimates of recombination rate

Negative correlation

Positive correlation

Pessia et al. (2012) GBE

Eugénie Pessia



Evolution of GC-content in bacterial genomes: the gBGC hypothesis expands

- Analysis of relationships between recombination rate and GC-content in bacterial species
- 21 species with ≥ 6 sequenced strains
- Detection of recombination events in multiple alignments
  - 7 clonal species : no sign of recombination
  - 14 species with > 10% recombinant genes
- Bin genes according to their GC-content at 3rd codon position (GC3) => compute the fraction of recombining genes per bin

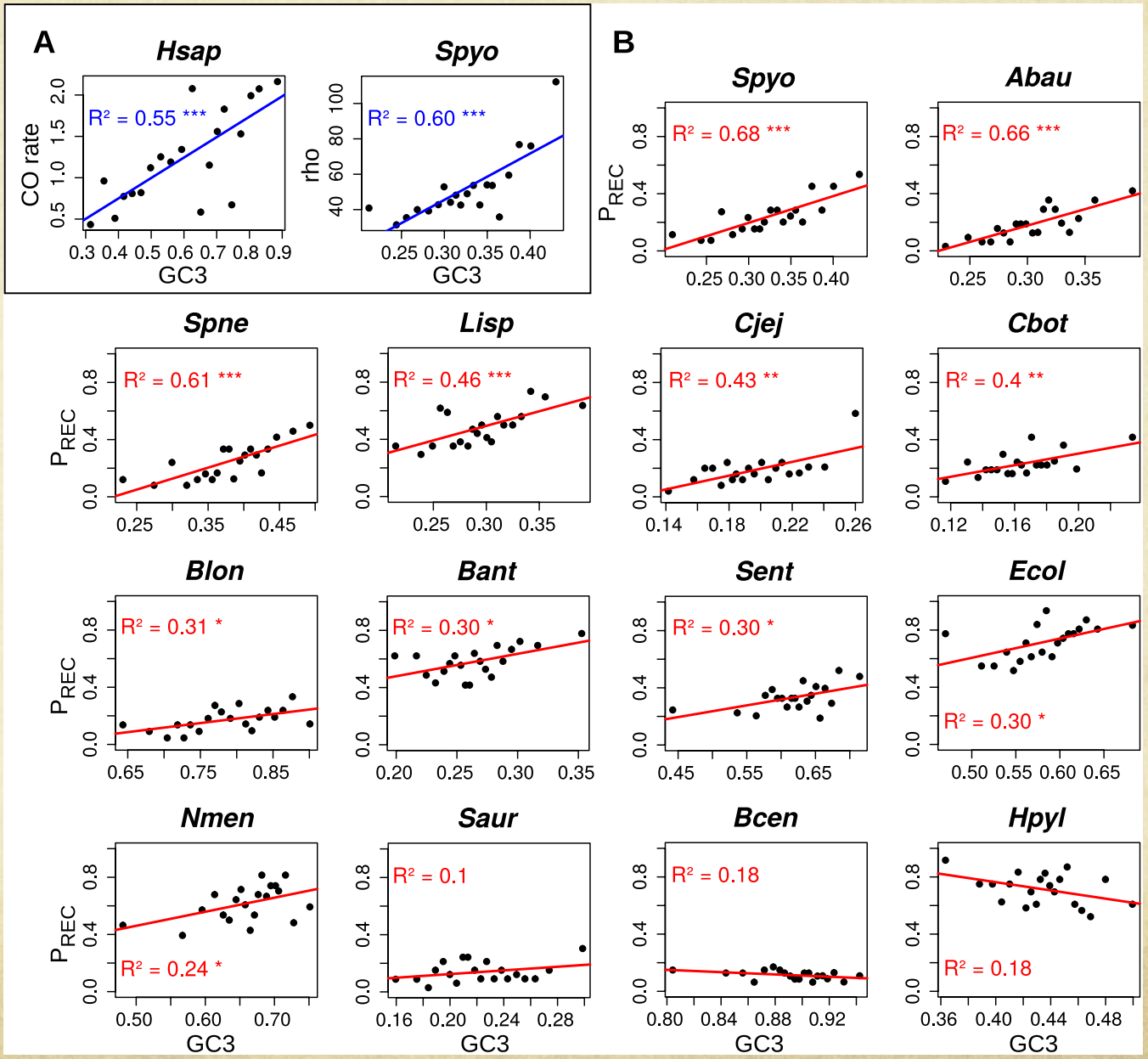
Lassale et al. (2015) Plos Genet

Frequency of recombination vs. GC-content

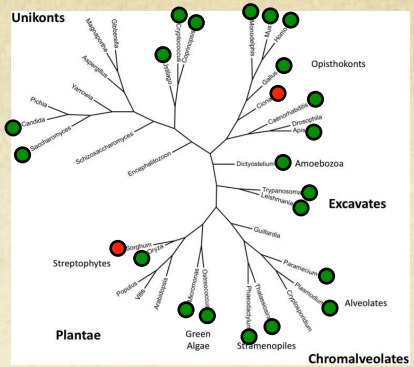
14 bacterial species

11 with a significant positive correlation

Florent Lasalle



Summary



- gBGC is probably a widespread process, both in eukaryotes and bacteria
- Why is the conversion bias always in favor of GC?
  - mutation pattern: universal bias towards AT
  - Selection for lower mutation rate may favor GC-biased mismatch repair => gBGC ??
- gBGC explains how genomes can remain relatively GC-rich, despite a universal AT-biased mutation pattern (no need to invoke selection on genomic GC-content)

Two types of BGC

- Mismatch repair bias: BGC favors GC-alleles over AT-alleles
  - gBGC: GC-Biased Gene Conversion

- Initiation bias: BGC in favor of cold alleles (i.e. alleles with lower rate of DSB formation, i.e. alleles with low recombination activity)
  - dBGC: DSB-induced Biased Gene Conversion

The recombination landscape in the human genome

- Non-uniform distribution: 80% of crossovers occur in 10% of the genome *(McVean et al. 2004 Science)*
- Recombination hotspots (< 2 kb)

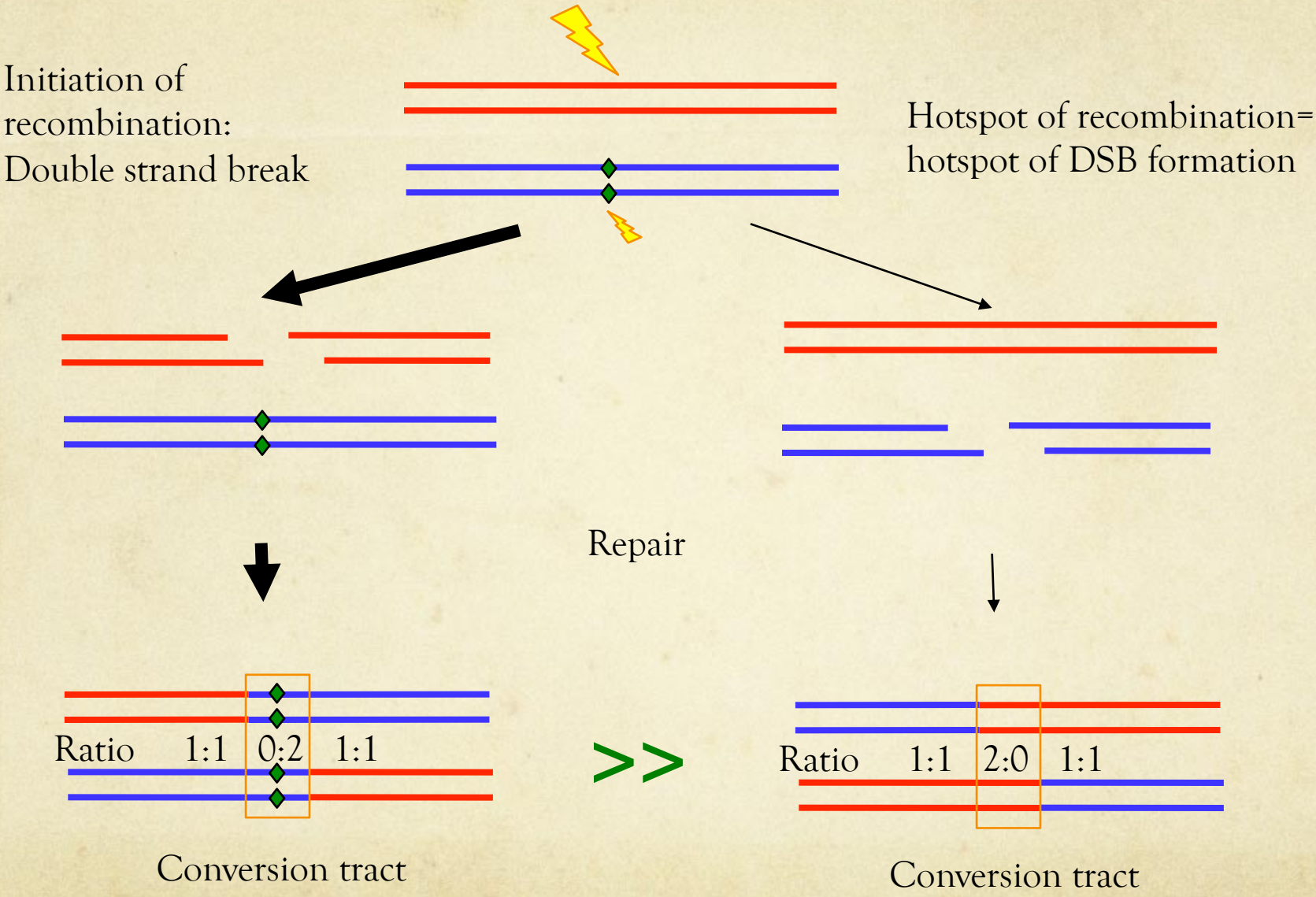
Recombination hotspots in the human genome

- >30,000 recombination hotspots (< 2 kb)
- Detected by analysis of linkage disequilibrium in human populations (HapMap)
  - historical hotspots
- Identification of a 13-bp sequence motif over-represented in recombination hotspots:

  Core HM motif: **CCTCCCTNNCCAC**

Myers et al. 2008, Nat. Genet

The recombination hotspot paradox

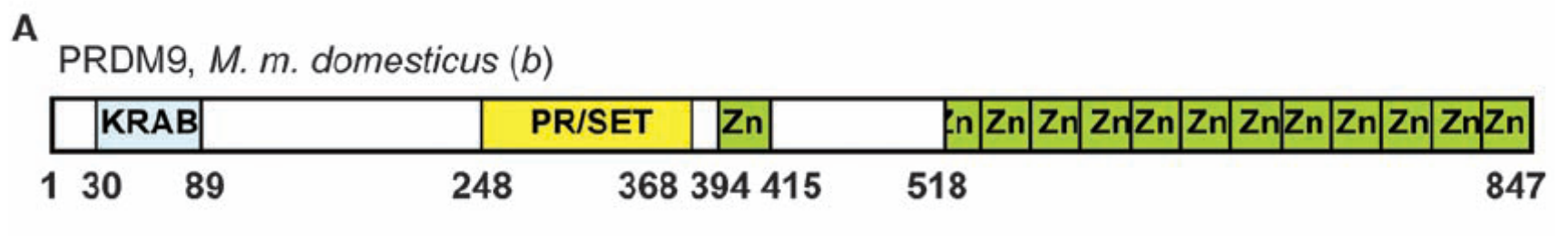


dBGC favors cold alleles

The recombination hotspot paradox

- dBGC: hot alleles get converted by cold alleles
- The fate of recombination hotspots is to disappear from genomes
- Hot-spots have a short lifespan: positions are not conserved between Human and Chimp (Ptak et al., Nat. Genet. 2005; Winckler et al. Science 2005; Auton et al. Science 2012)
- How are recombination hotspots maintained on the long term?

PRDM9: a major determinant of meiotic recombination hotspots

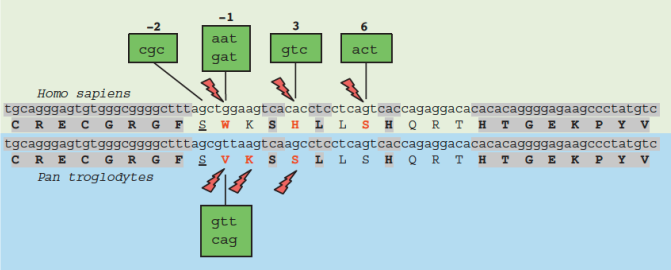


- Zinc-finger DNA-binding protein
- Allelic variation of PRDM9 in the DNA-binding domain
- Allelic variation correlates with variation in recombination hotspot activity
- The major allele (A) recognizes specifically the 13-bp motif

Baudat et al., Science 2010; Myers et al. Science 2010

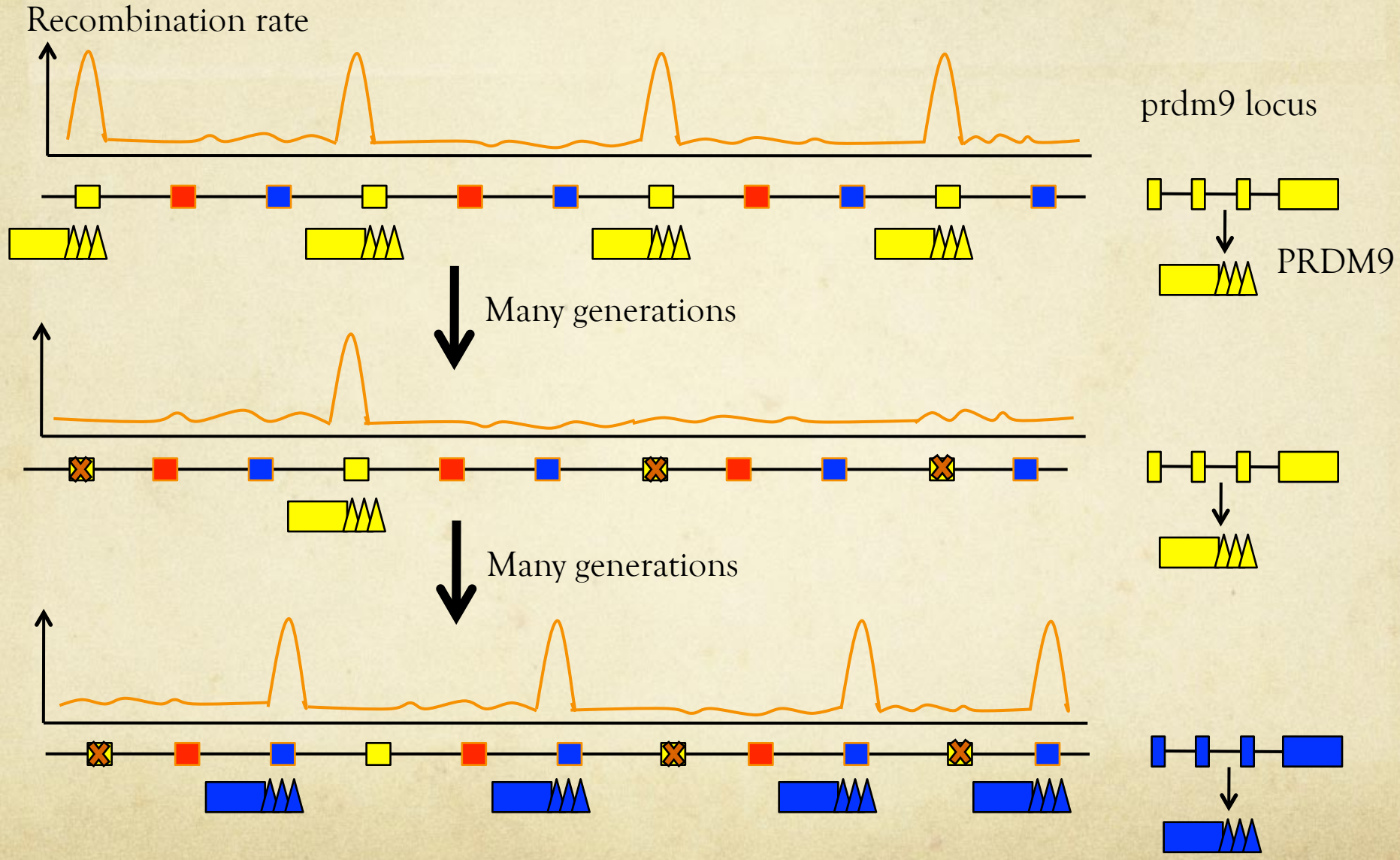
PRDM9: positive selection in many metazoans

Positive selection on DNA-binding sites (Oliver et al. 2009)



C. Ponting (2011) *Trends Genet*

Rapid hotspot turnover: the Red Queen model



Baudat et al., Science 2010; Myers et al., Science 2010

Rapid hotspot turnover: the Red Queen model

Major PRDM9 allele: Y

Loss of PRDM9_Y targets

Loss of PRDM9_X targets

Loss of fitness of individuals having PRDM9 allele X

New PRDM9 allele (Y) increase in frequency

Major PRDM9 allele: X

Myers et al., Science 2010

Testing the Red Queen Model

- Is this model realistic ?
- What is the strength of dBGC on PRDM9 target motifs (HM: **CCTCCCTNNCCAC**) ?
- What is the dynamics of hotspots turnover ?
- When did the A allele of PRDM9 start being active?

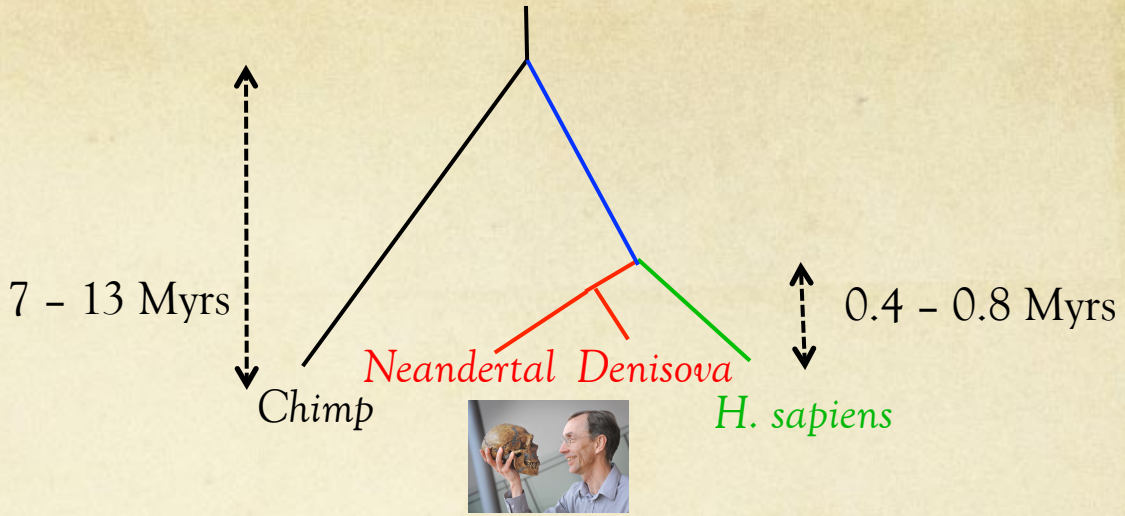
If a motif is targeted by PRDM9

⇔

Loss of this motif by conversion

=> Analyze the loss of HM motifs in the human lineage

Approach



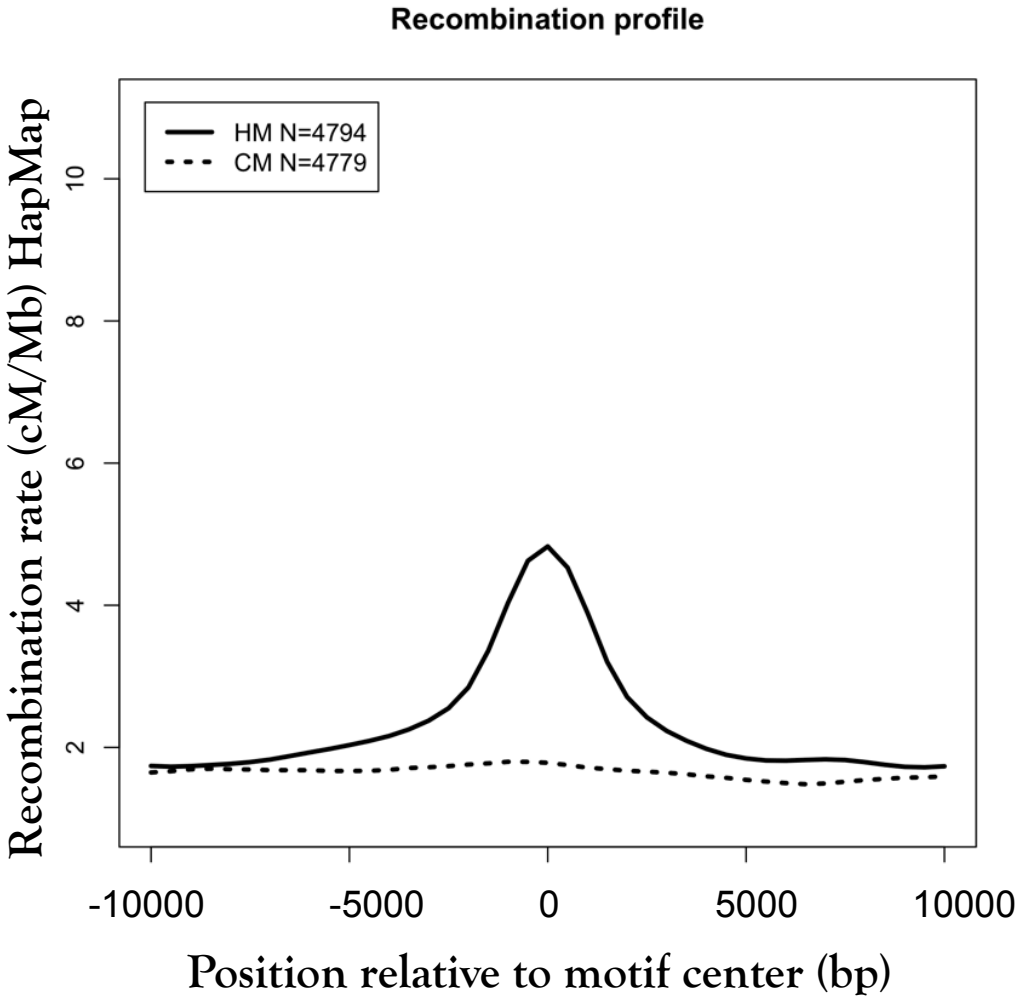
- Identify motifs that were present in the human/chimp ancestor
  - Sequences of the human/chimp ancestor reconstructed from 6 primates genome alignments (Human, Gorilla, Chimpanzee, Orangutan, Macaque, Marmoset)
- Detect mutations along the human lineage, and date them relative to the divergence Sapiens/Denisova
  - High coverage Denisova genome sequences, aligned on the human reference genome (Reich et al. 2010, Meyer et al. 2012)
- Analyze their allelic frequency in human populations
  - Human SNPs from the 1000 genomes project (Durbin et al. 2010)

Hotspot motif (HM) and control motif (CM)

- Motifs:

CC*TCCCTNNCCAC* (HM)

C*TTCCCTNNCCAC* (CM)



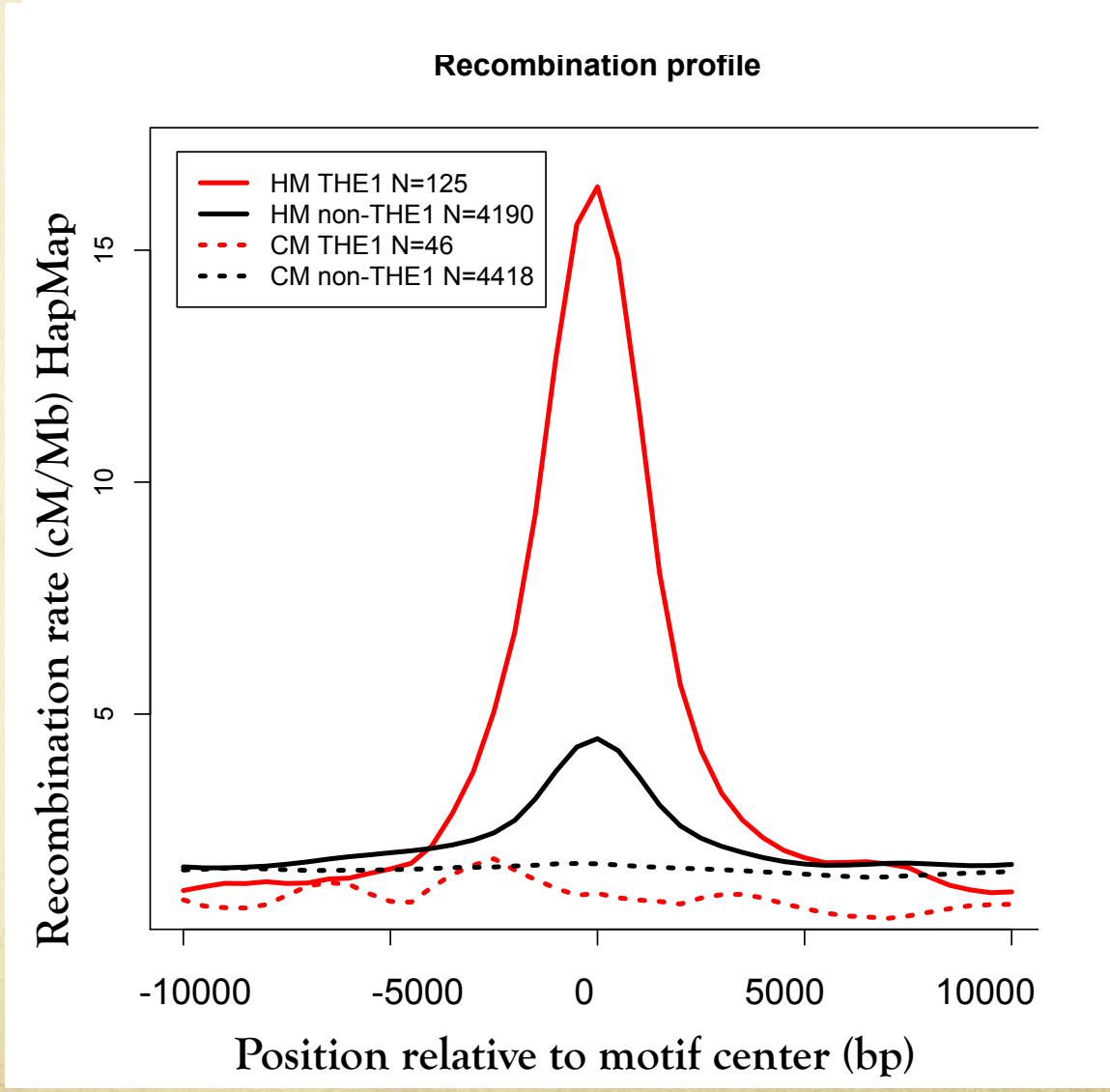
Hotspot motif (HM) within THE1 transposable elements

- Myers et al. (2008): very high recombination rate at HM motifs located within THE1 elements

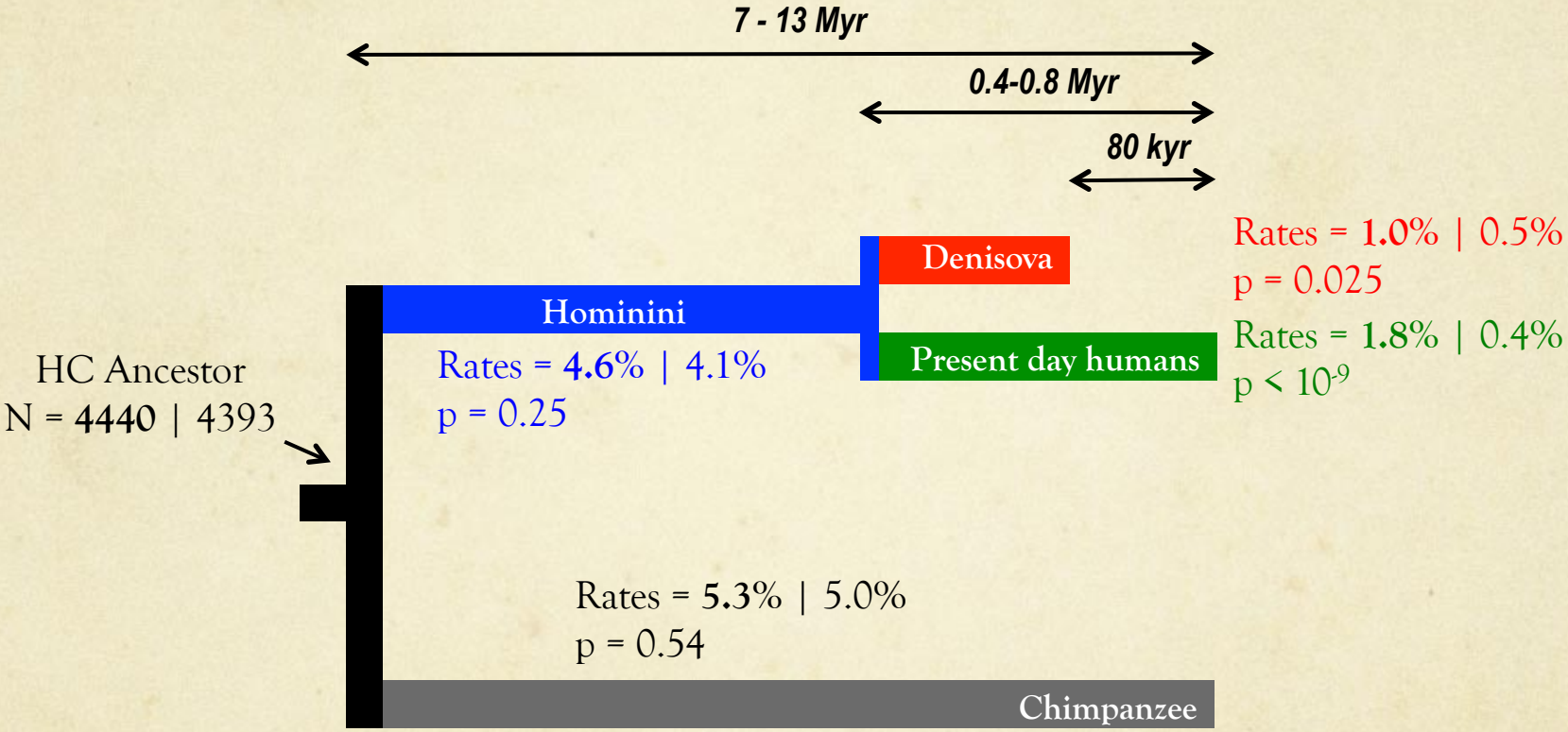
- Motifs:

CCTCCCTNNCCAC (HM)

CTTCCCTNNCCAC (CM)



Tracking PRDM9 HM motif losses across the human lineage

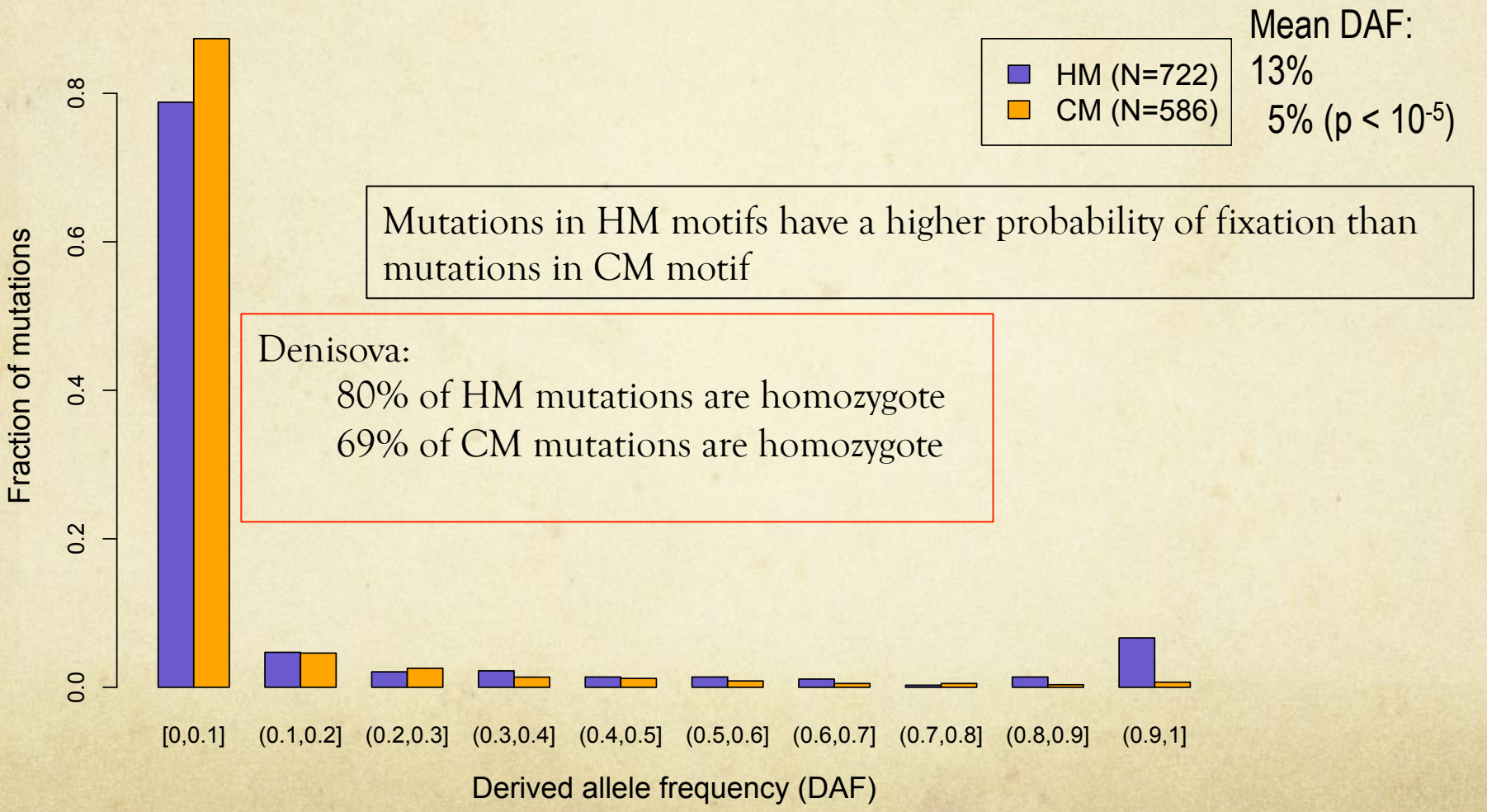


HM | CM

CCTCCCTNNCCAC | CTTCCCTNNCCAC

Mutagenic effect of recombination or fixation bias?

Analysis of derived allele frequency spectra



Mutations in HM motifs have a higher probability of fixation than mutations in CM motif

Denisova:

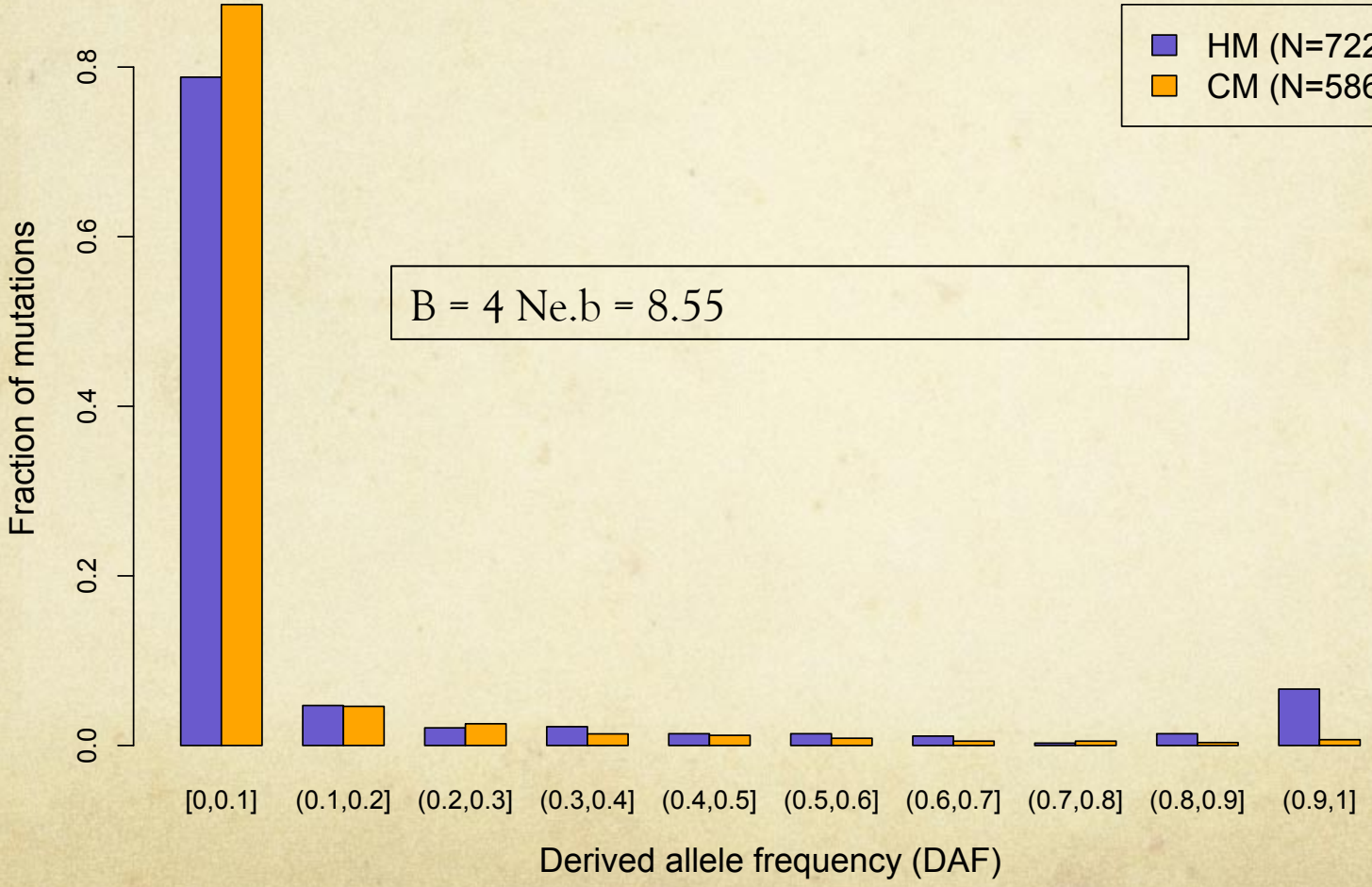
80% of HM mutations are homozygote

69% of CM mutations are homozygote

Strength of the fixation bias

Analysis of derived allele frequency spectra

Fitting a population genetics model to the DAF spectra (Eyre-Walker, Woolfit & Phelps, 2006)

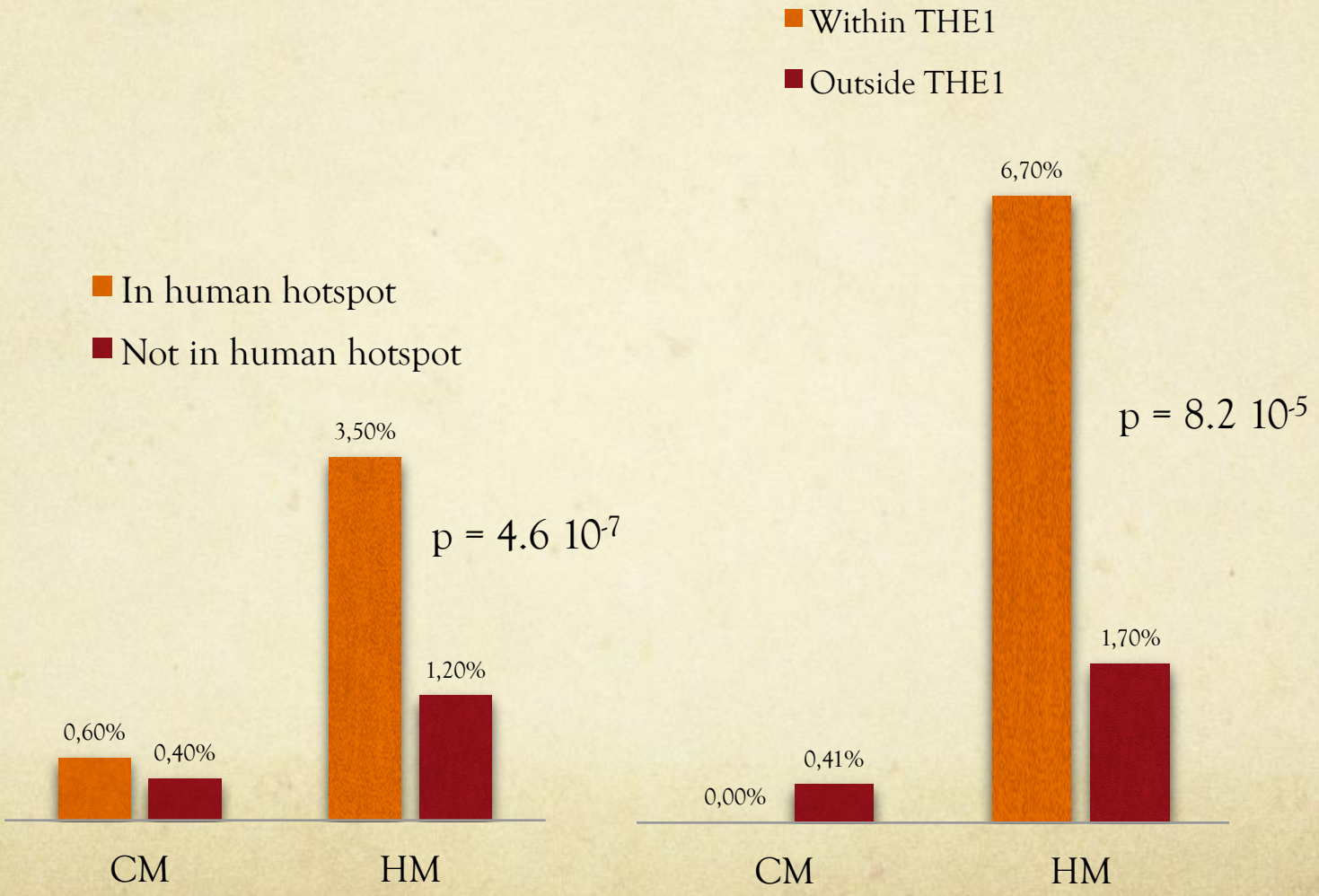


Sylvain Glémin
(Montpellier)

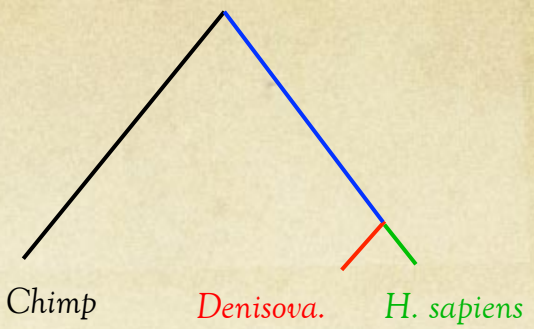
Variations in the strength of dBGC on HM motifs

- The HM motif is significantly enriched in hotspots, but 80% of them are NOT in hotspot
- The strength of dBGC at a given locus depends on the net difference in recombination rate between the hot allele and the colder allele
- => HM motifs located located in loci of low recombination are expected to be subject to weaker dBGC

HM losses in the human branch are more frequent at highly recombining loci



Conclusion (1)

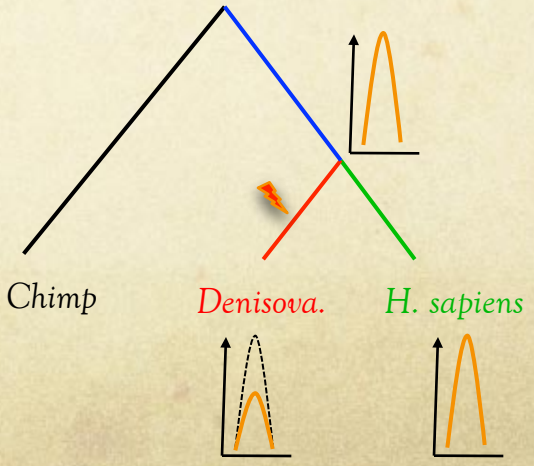


- High rate of losses of the HM motif in the human lineage
- Due to a fixation bias (not mutation)
- Stronger fixation bias for motifs located in highly recombining loci
- => consistent with the dBGC drive model: conversion bias favoring mutations that disrupt recombination hotspots
- The loss of the HM motif started before the divergence between *H. sapiens* and *Denisova*
- HM (or a similar motif) has been the target of PRDM9 for >0.8 Myrs

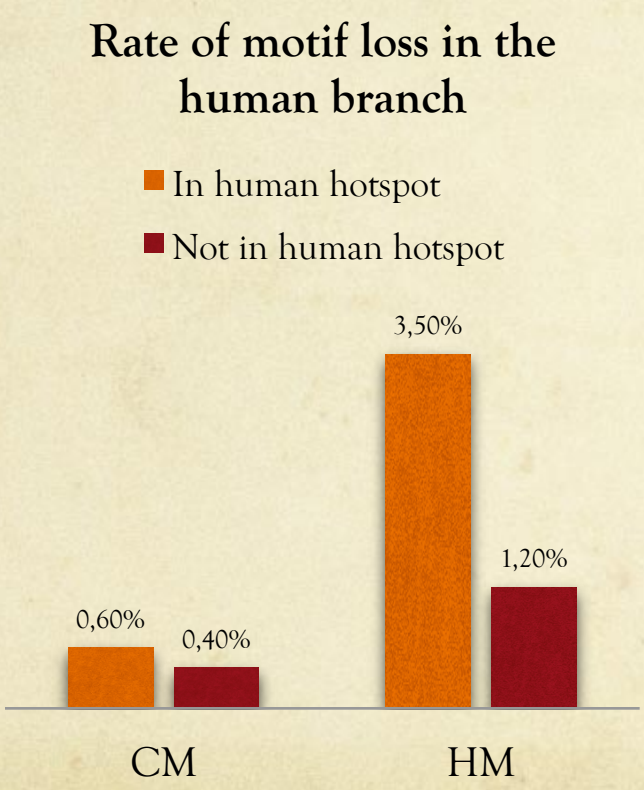
Recombination hotspots: conserved between Denisova and Sapiens?

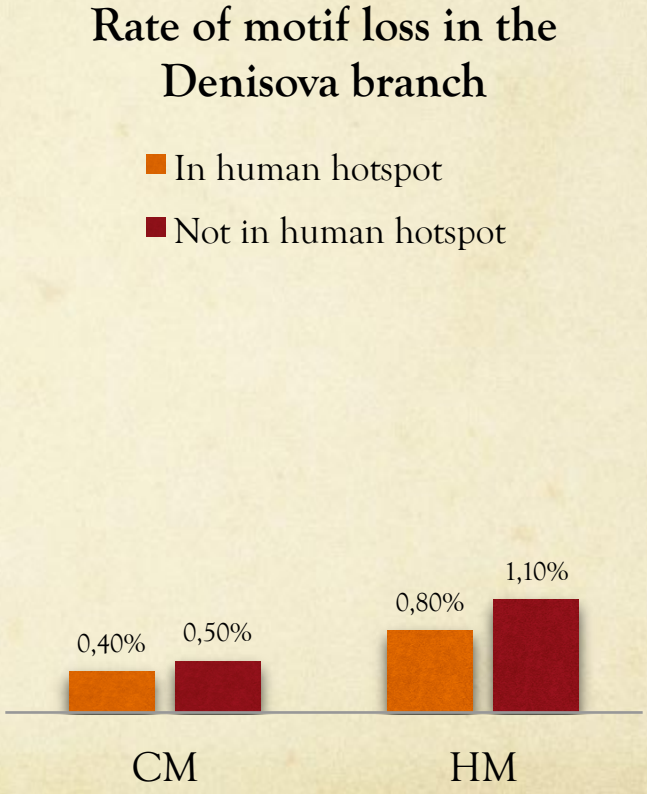
- HM motifs located in human hotspots accumulate more mutations

- If recombination hotspots were shared between sapiens and Denisova, then, motifs located at loci corresponding to human hotspots should also accumulate more mutations in the Denisova branch

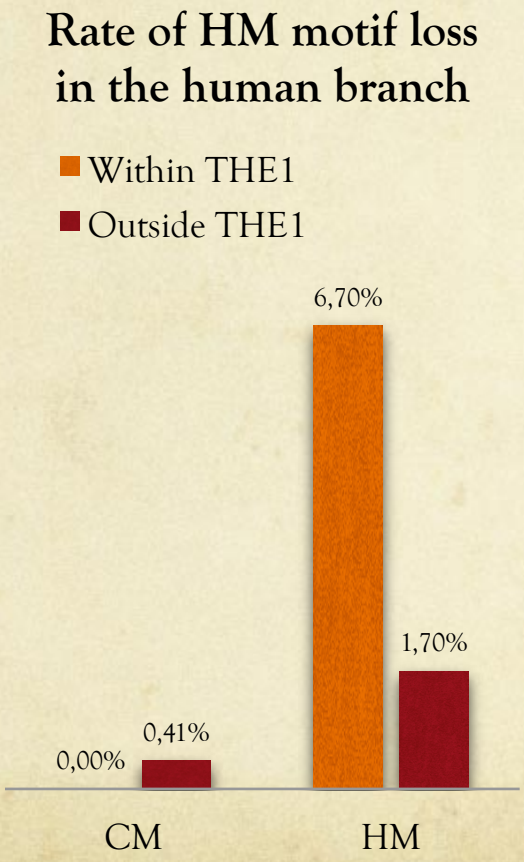


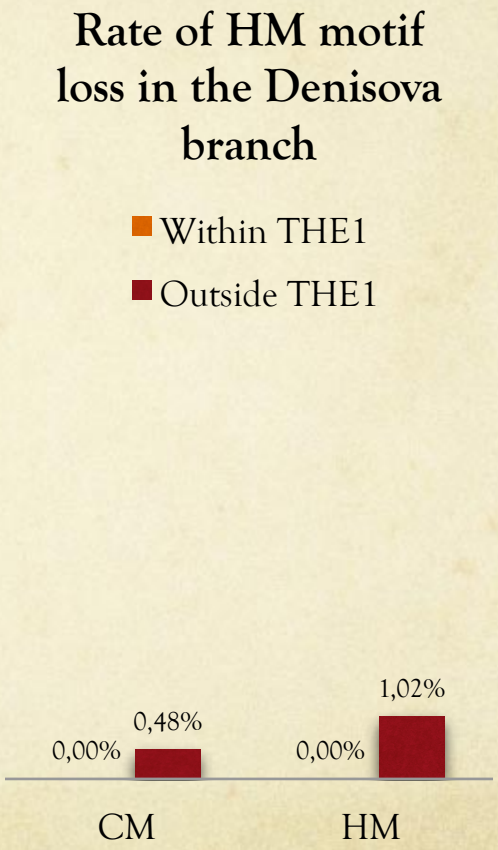
Recombination hotspots: conserved between Denisova and Sapiens?





Recombination hotspots: conserved between Denisova and Sapiens?

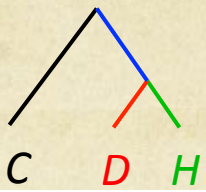




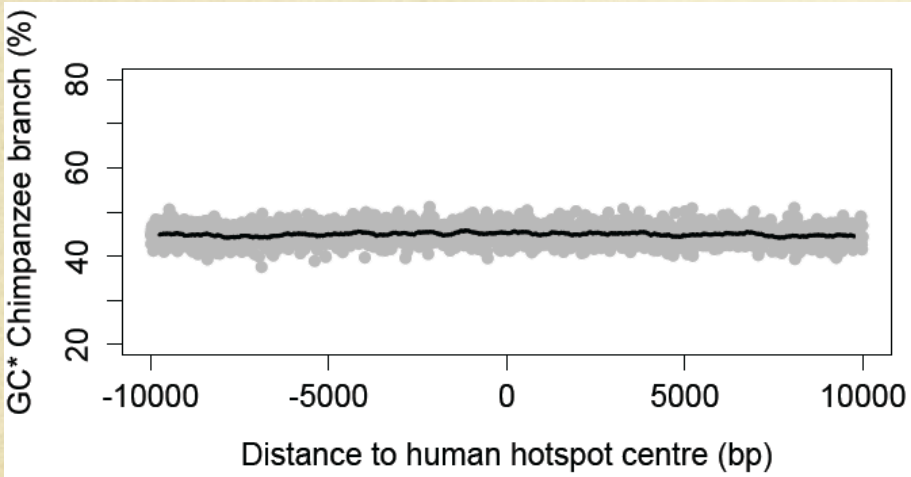
Recombination hotspots: conserved between Denisova and Sapiens?

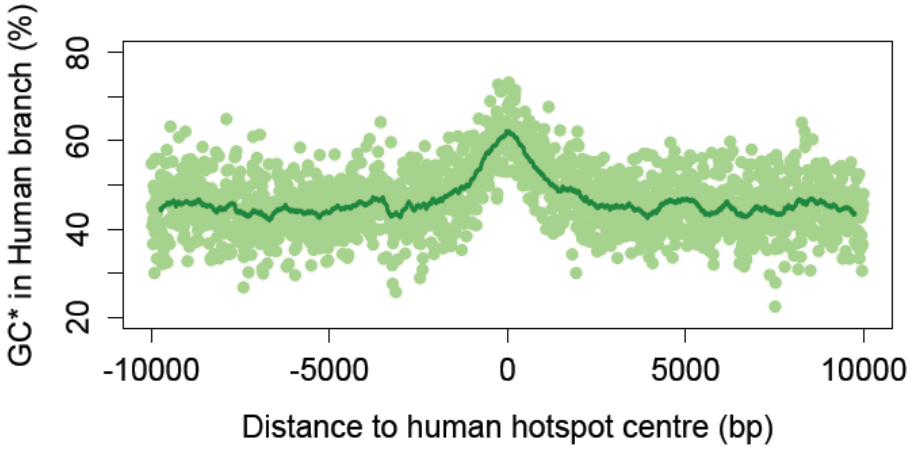
- HM motifs located in loci corresponding to human hotspots show no evidence of strong dBGC in Denisova.
- This suggests that recombination hotspots were not shared between Denisova and Sapiens
- Can we get more evidence to support this conclusion?

Dating the activity of recombination hotspots: looking for the signature of gBGC

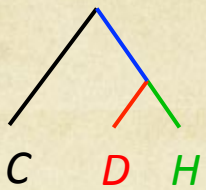


N=32,987 human recombination hotspots (HapMap)

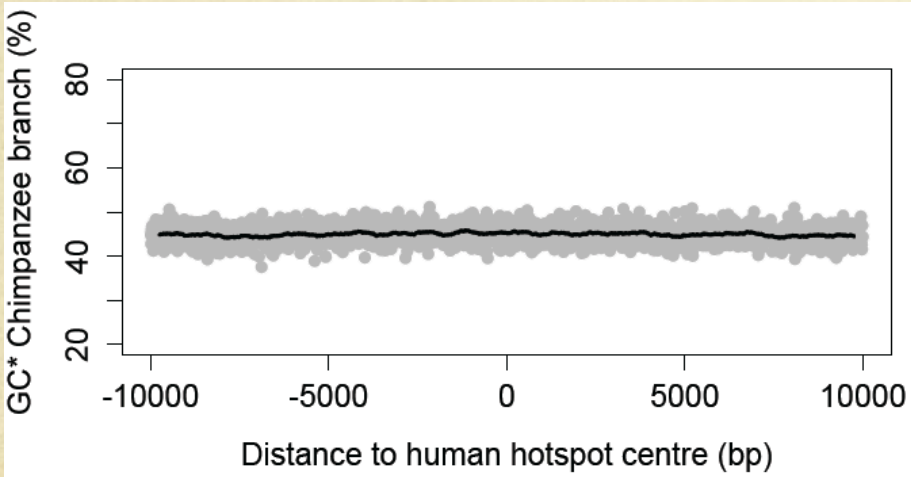


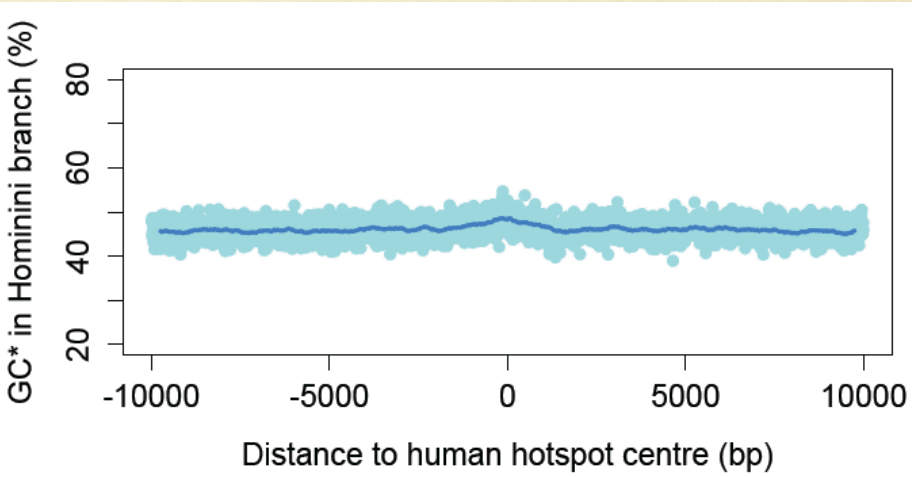


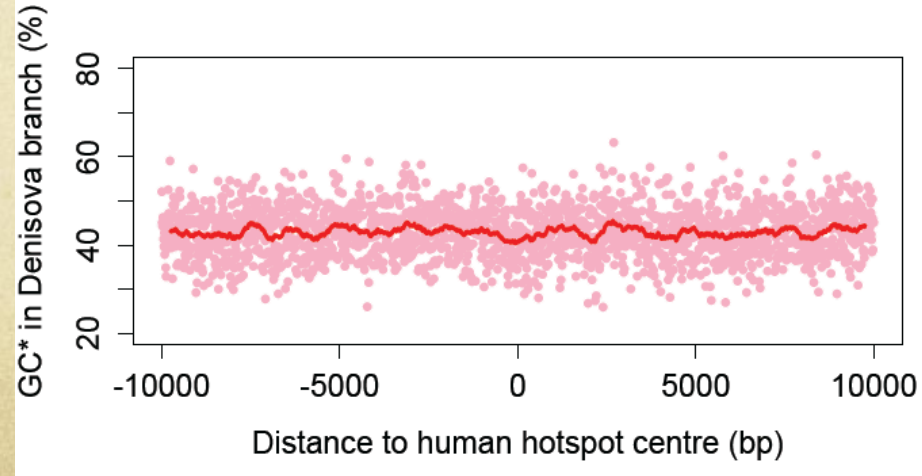
Dating the activity of recombination hotspots: looking for the signature of gBGC

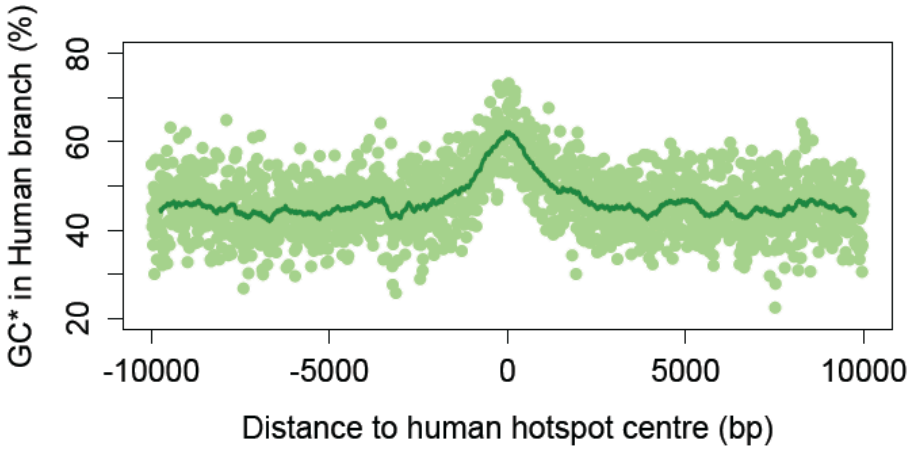


N=32,987 human recombination hotspots (HapMap)









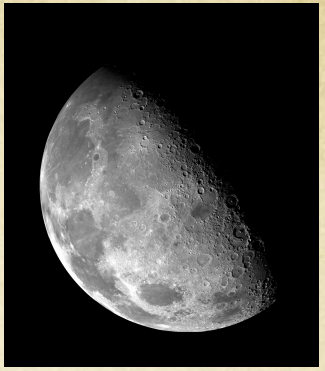
Evolution of recombination hotspots

- Human recombination hotspots are recent (< 1 Myr)
- Recombination hotspots are not shared between modern humans and Denisovans: **the turnover of recombination hotspots can be very rapid**
- The predicted lifespan of the strongest PRDM9 targets (which concentrate most of recombination events) is extremely short
- **Consistent with the Red Queen hypothesis :**
  - **rapid evolution of hotspots caused by dBGC-driven erosion of PRDM9 targets**

Evolution of recombination hotspots

- PRDM9 orthologs found in many vertebrates (and possibly in other metazoan)
- Plants and fungi:
  - no PRDM9
  - Recombination hotspots = regions of open chromatin (promoters)
  - No red queen process: stable hotspots (Lam et al. 2015)
- Birds, canids (dogs):
  - no PRDM9
  - Recombination hotspots = regions of open chromatin (promoters)
  - No red queen process: stable hotspots (Singhal et al. 2015)

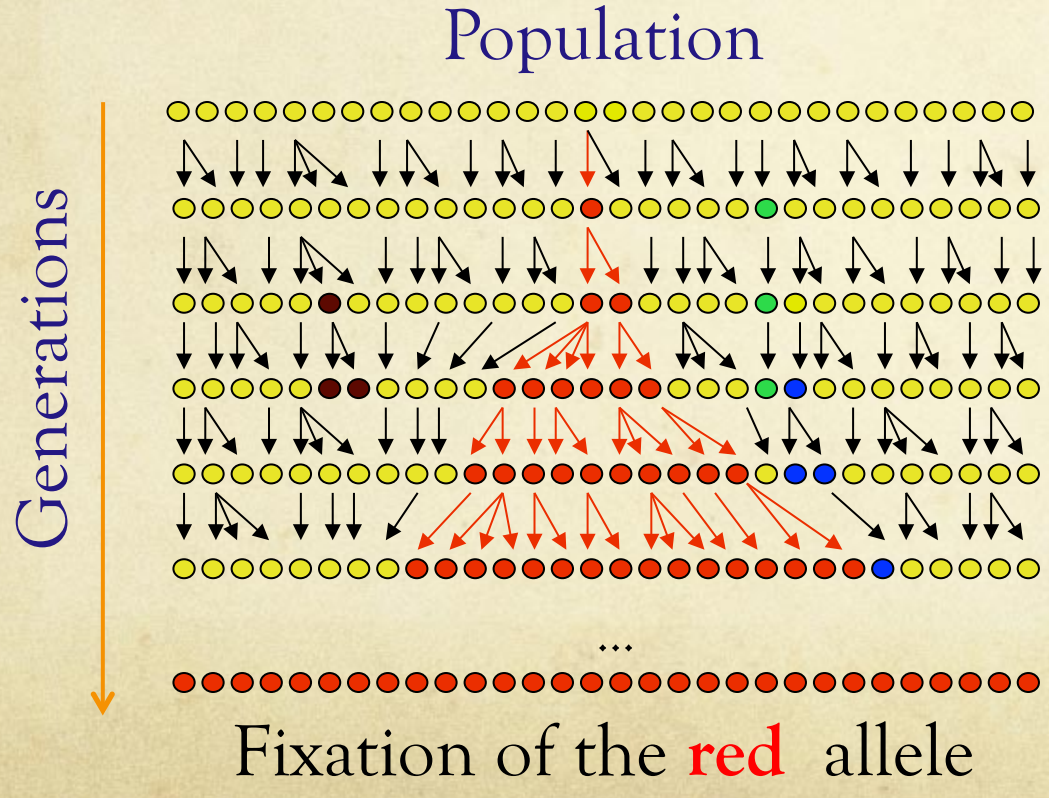
The impact of recombination on genome evolution



- Recombination disrupts linkage between selected sites
  - => reduce Hill-Robertson interference, increase selection efficacy
  - On the long-term, recombination is required for adaptation
- Biased gene conversion: the dark side of recombination
  - gBGC drives the evolution of GC-content
  - dBGC drives the evolution of recombination hotspots (PRDM9 dependent)

Evolution

- Mutation => new alleles
- Changes of allele frequencies over generations



- ✓ Genetic drift
- ✓ Natural selection
- ✓ Biased Gene Conversion

Acknowledgements

Dominique Mouchiroud

Yann Lesecque

Vincent Daubin

Anouk Necsulea

Florent Lasalle

Nicolas Lartillot

Nicolas Galtier
(Montpellier)

Sylvain Glémin (Montpellier, France)

Eugénie Pessia