

# The Coalescent

## Evolution Backward in Time

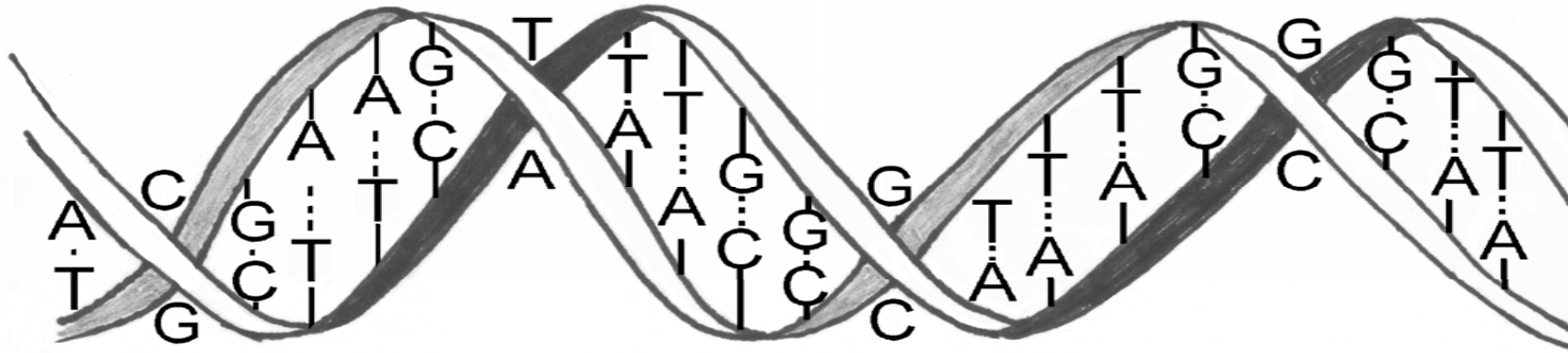
Joachim Hermisson

*Mathematics and Biosciences Group*

Mathematics & MFPL, University of Vienna

# Introduction to the Coalescent

data, data, data, ...

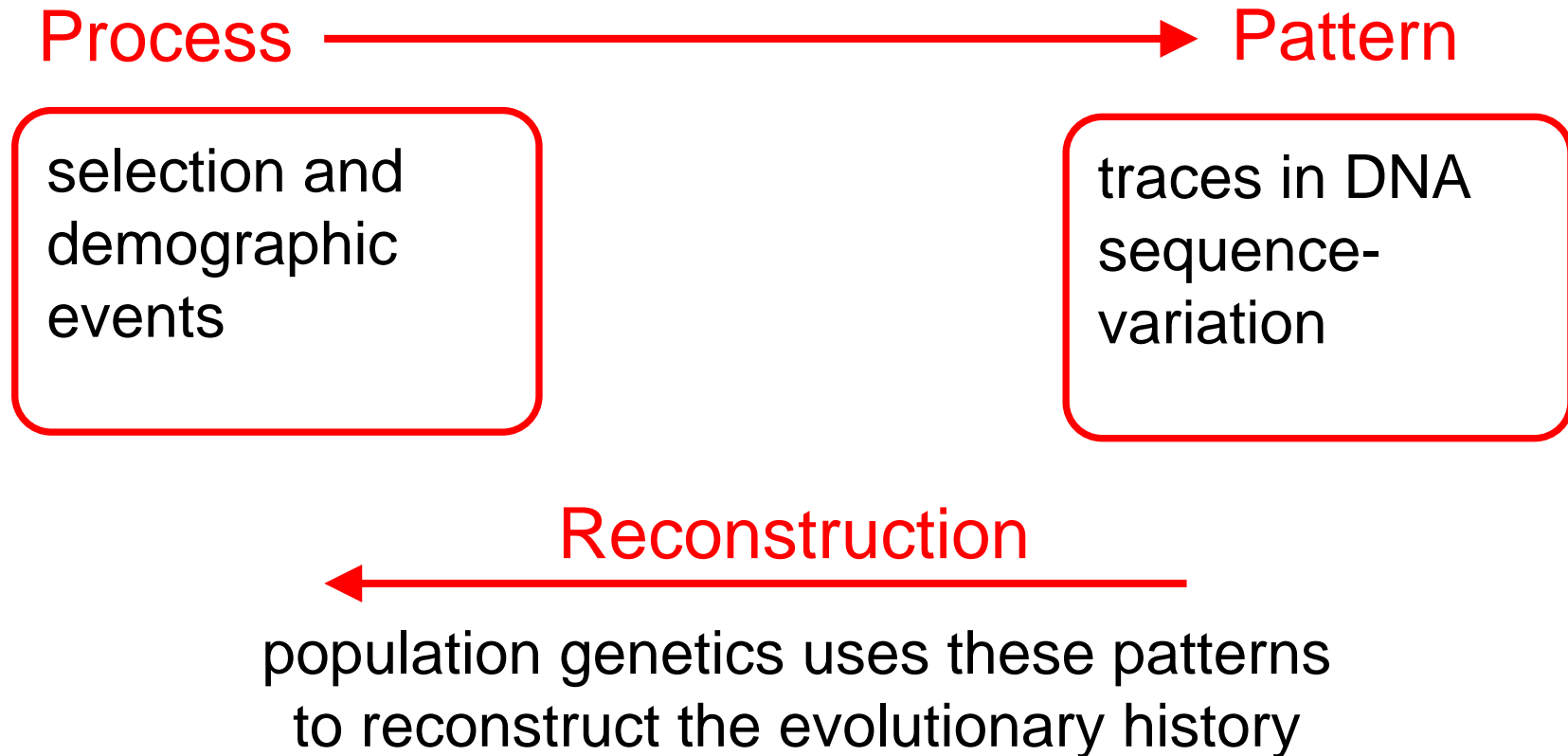


Massive accumulation of DNA sequence data

- 1980's: 3-4 years PhD projects to sequence a single gene (some 1000 base pairs)
- 1990 – 2003: *Human Genome Project* (~ 3 billion ( $10^9$ ) bases) expected: 3 billion \$, final: ~ 300 Mio \$
- 2010: **1000 Genome Project** ... in 3 years, 4000 \$ – 10000 \$ per genome, soon < 1000 \$  
1000 genomes also for *Drosophila*, *Arabidopsis* ...

# Patterns of Evolution

Reconstruction of evolutionary history  
from DNA sequence data



# Patterns of Evolution

Reconstruction of evolutionary history  
from DNA sequence data



pattern of a so-called “**selective sweep**”

# Patterns of Evolution

## "Selective Sweep"

~~AGATT CAGCCTAGACTTAGGGTGATGC~~  
~~ACATT CAGCCTAGACTTAGATGTTGC~~  
~~AGATT CAGCCTAGACTTAGGGTGTTGC~~  
~~ACATT AAGCGTAGACTTAGGGTGTTGC~~  
~~ACATT AAGCCTAGACTTAGATGTTGC~~  
~~ACATT AAGCCTAGACATAGGGTGTTGC~~

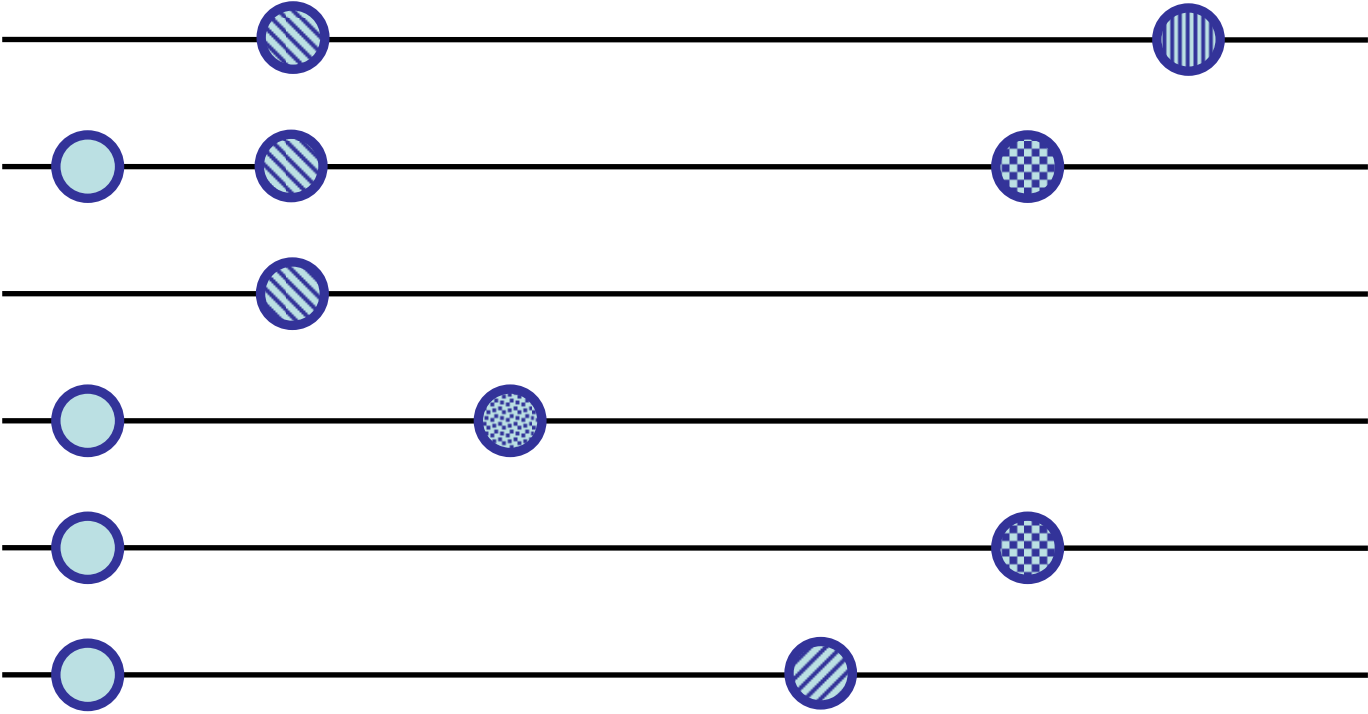
# Patterns of Evolution

## "Selective Sweep"

~~AGATT**C**AGCCTAGACTTAGGGT**G**ATGC~~  
~~**C**ATT**C**AGCCTAGACTTAG**A**TGTTGC~~  
~~AGATT**C**AGCCTAGACTTAGGGT**G**T**T**GC~~  
~~**C**ATT**A**AG**C****G**TAGACTTAGGGT**G**T**T**GC~~  
~~**C**ATT**A**AG**C**C**T**AGACTTAG**A**TGTTGC~~  
~~**C**ATT**A**AG**C**C**T**AGAC**A**TAGGGT**G**T**T**GC~~

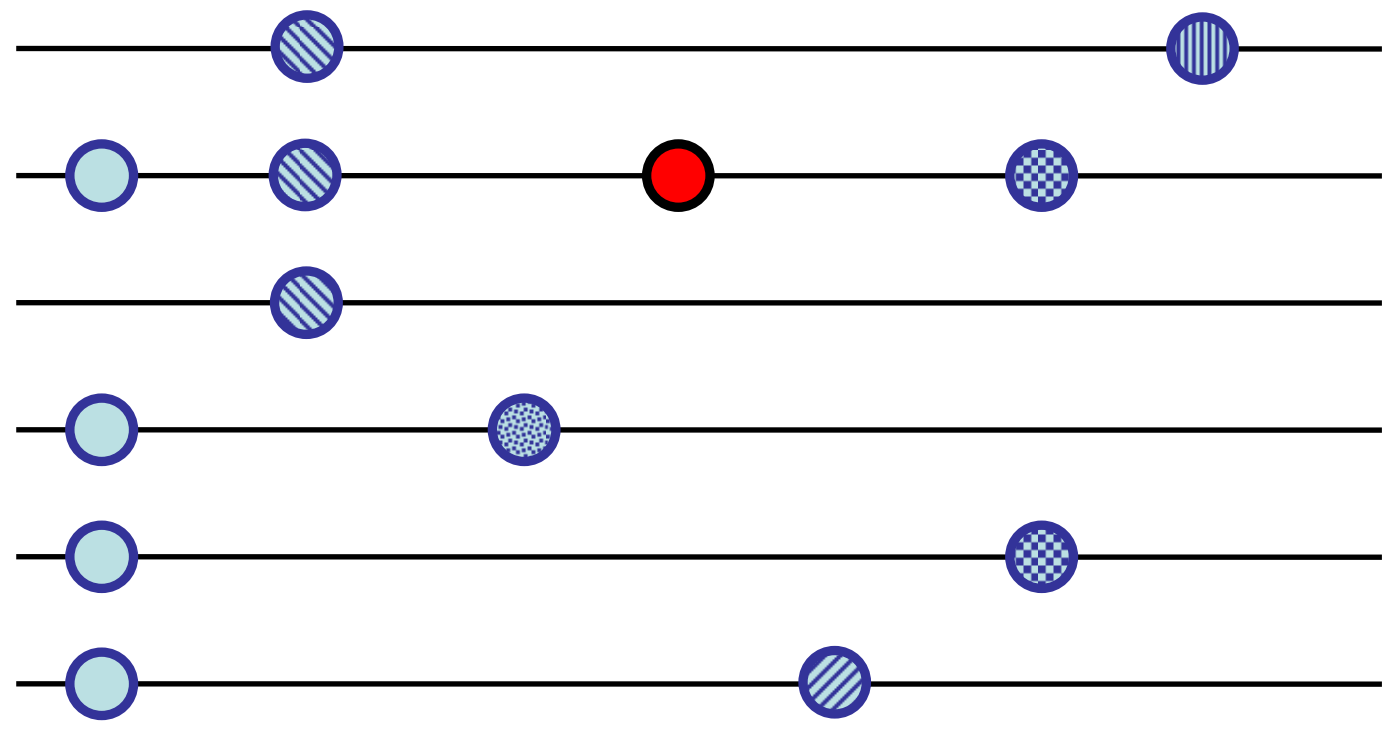
# Patterns of Evolution

## "Selective Sweep"



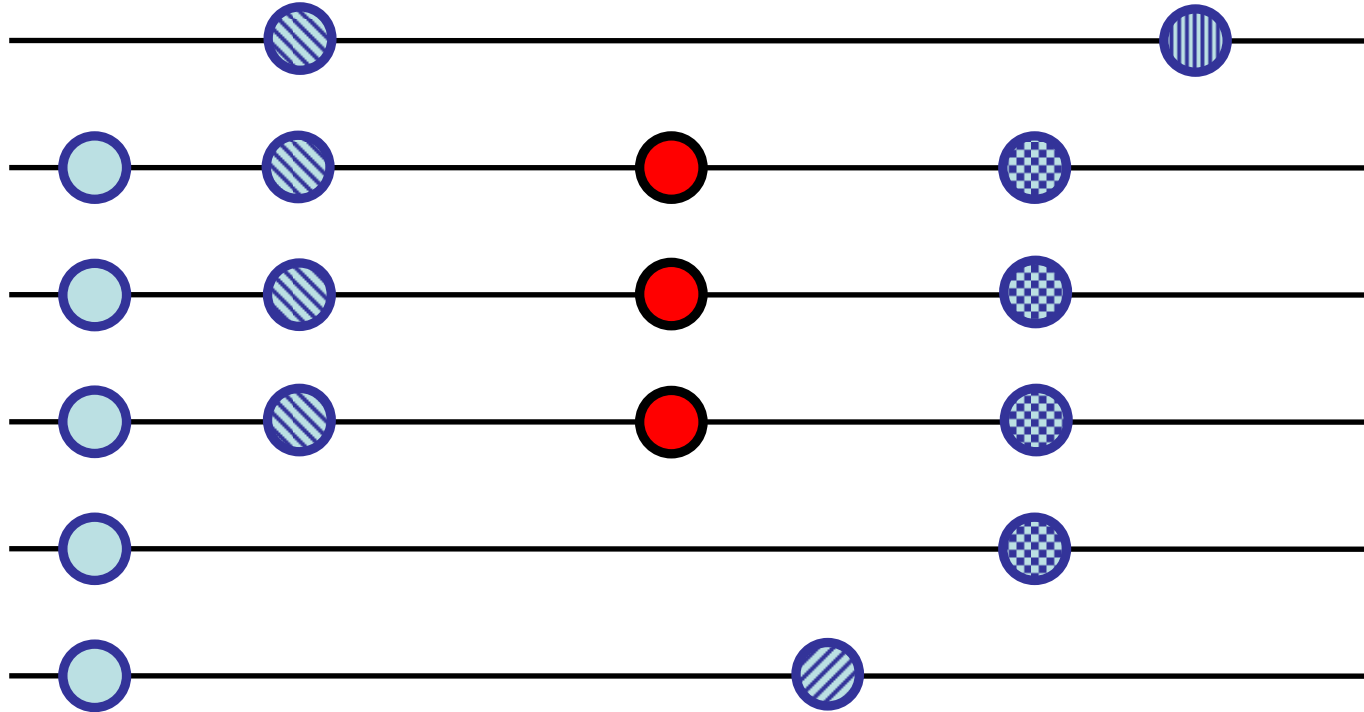
# Patterns of Evolution

## "Selective Sweep"



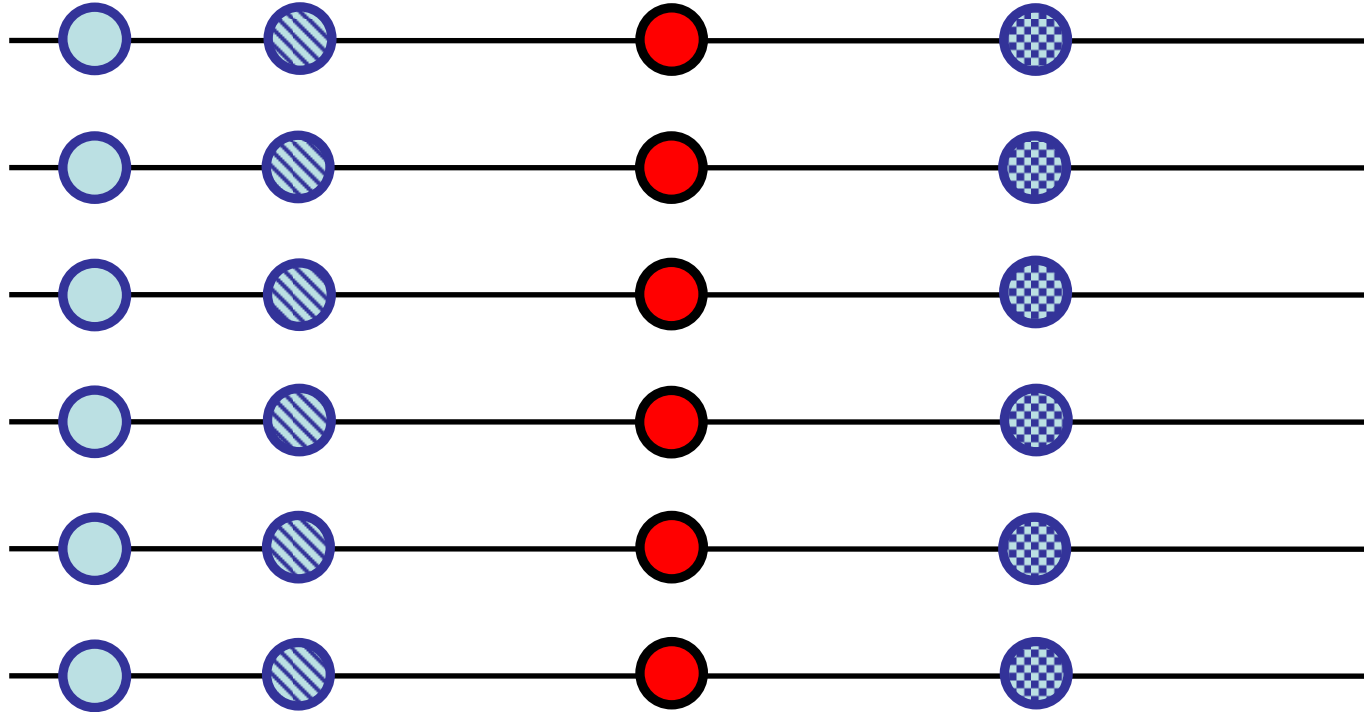
# Patterns of Evolution

## "Selective Sweep"



# Patterns of Evolution

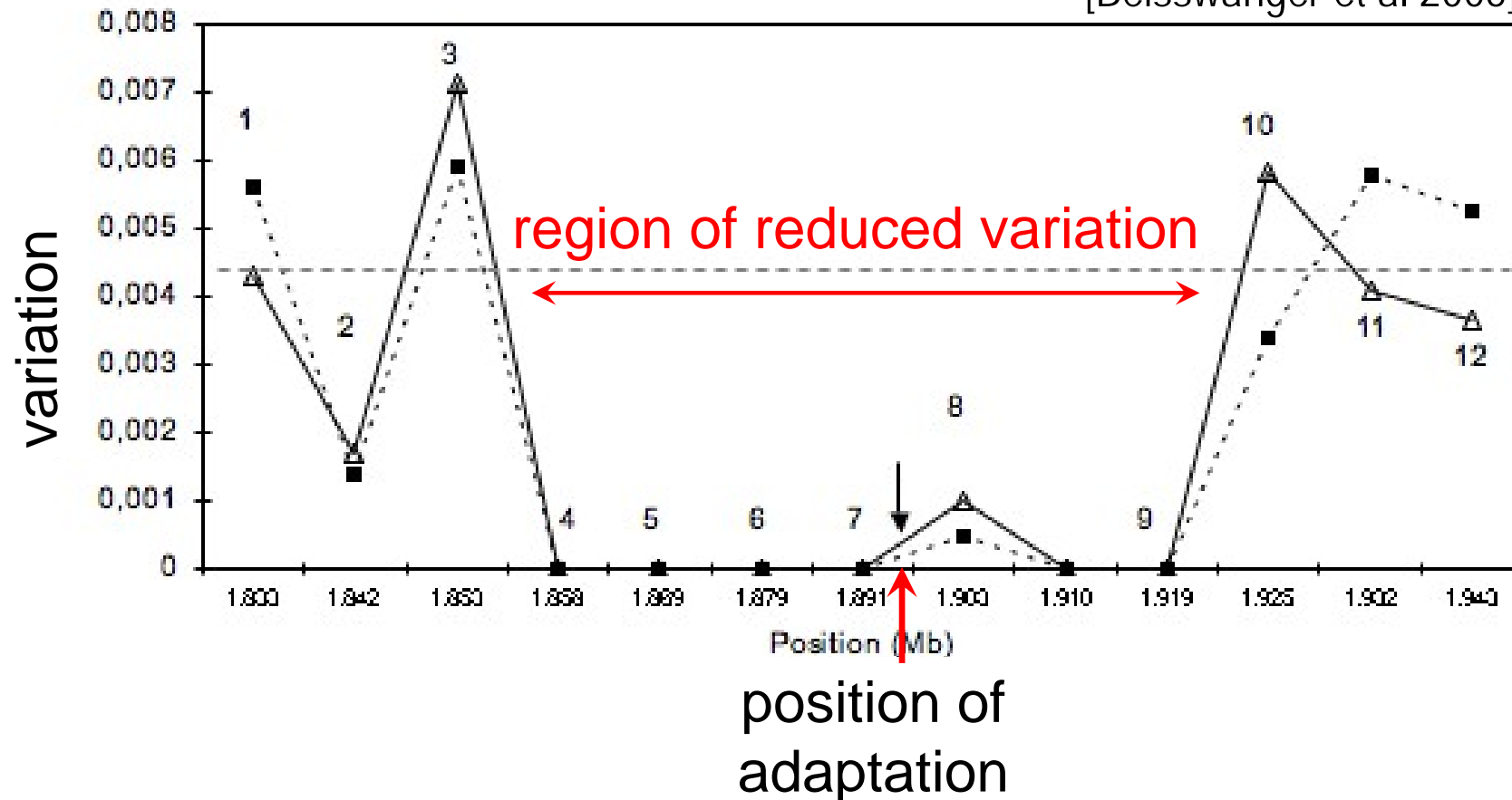
## "Selective Sweep"



# Patterns of Evolution "Selective Sweep"

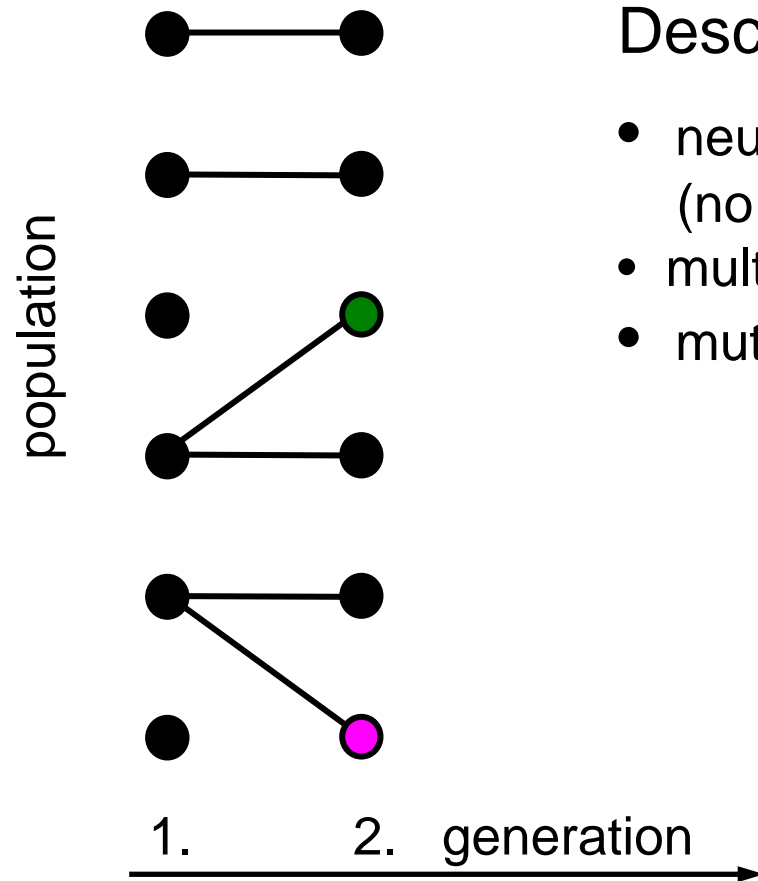
in the fruit fly *Drosophila melanogaster*

[Beisswanger et al 2005]



# Patterns of Evolution

## Wright-Fisher model

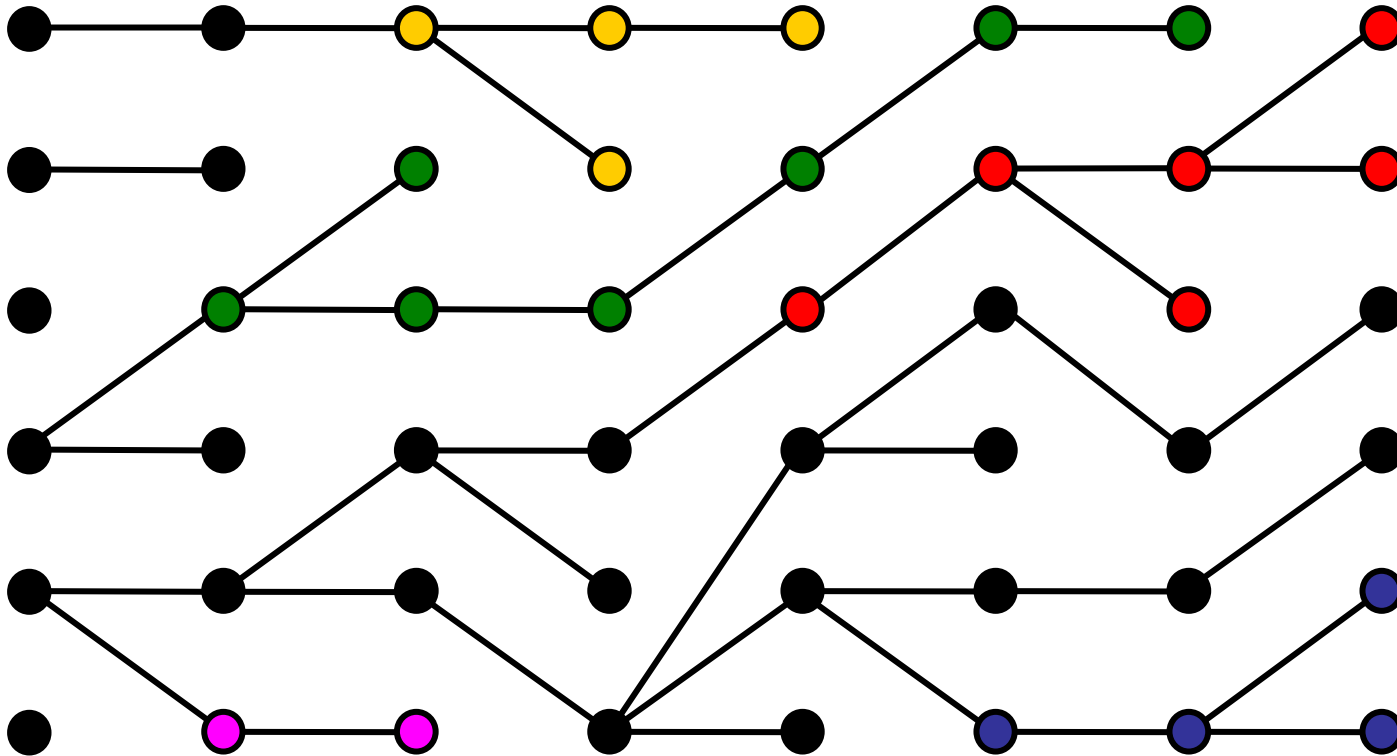


### Description of genetic variation

- neutral evolution forward in time (no selection)
- multinomial offspring distribution
- mutation with probability  $u$

# Patterns of Evolution

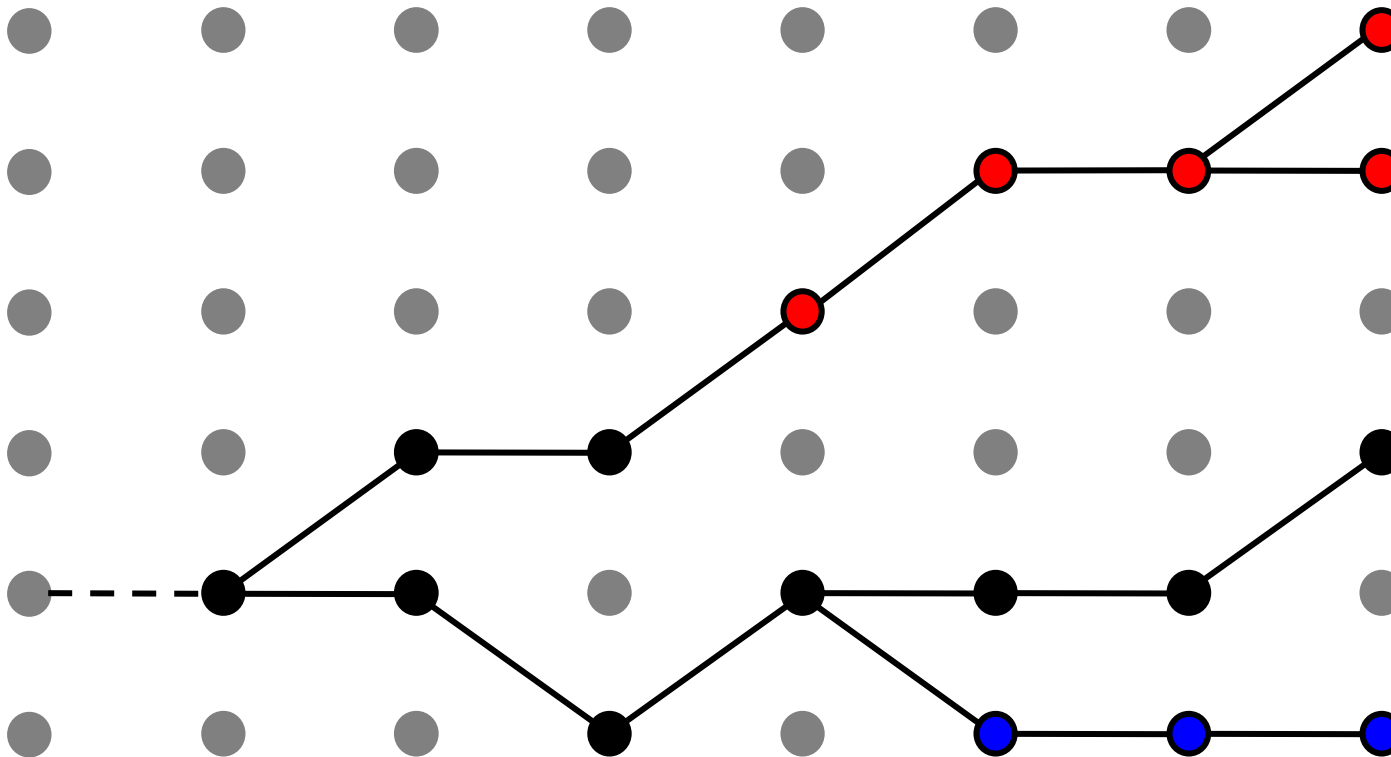
## Wright-Fisher model



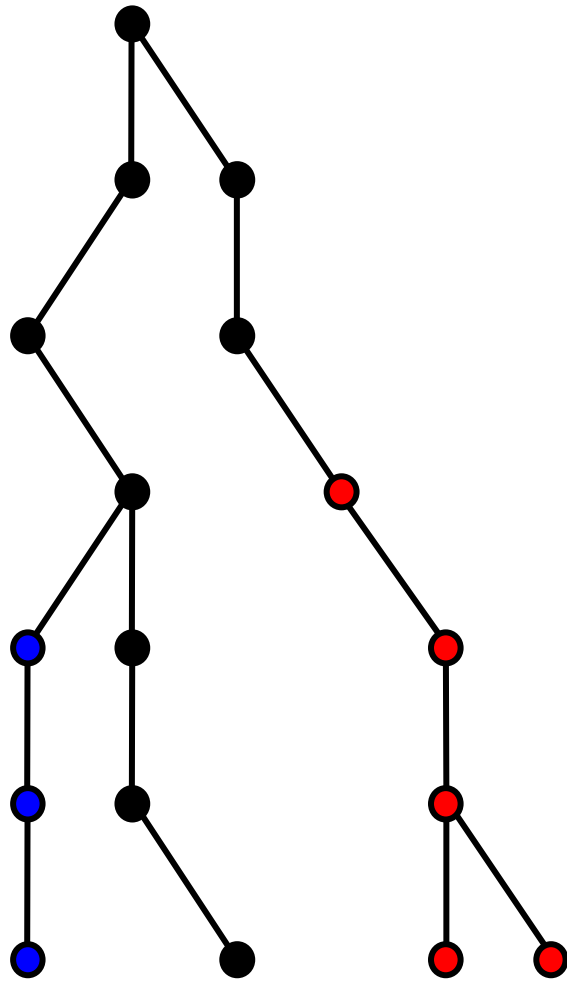


# Patterns of Evolution

## Wright-Fisher model

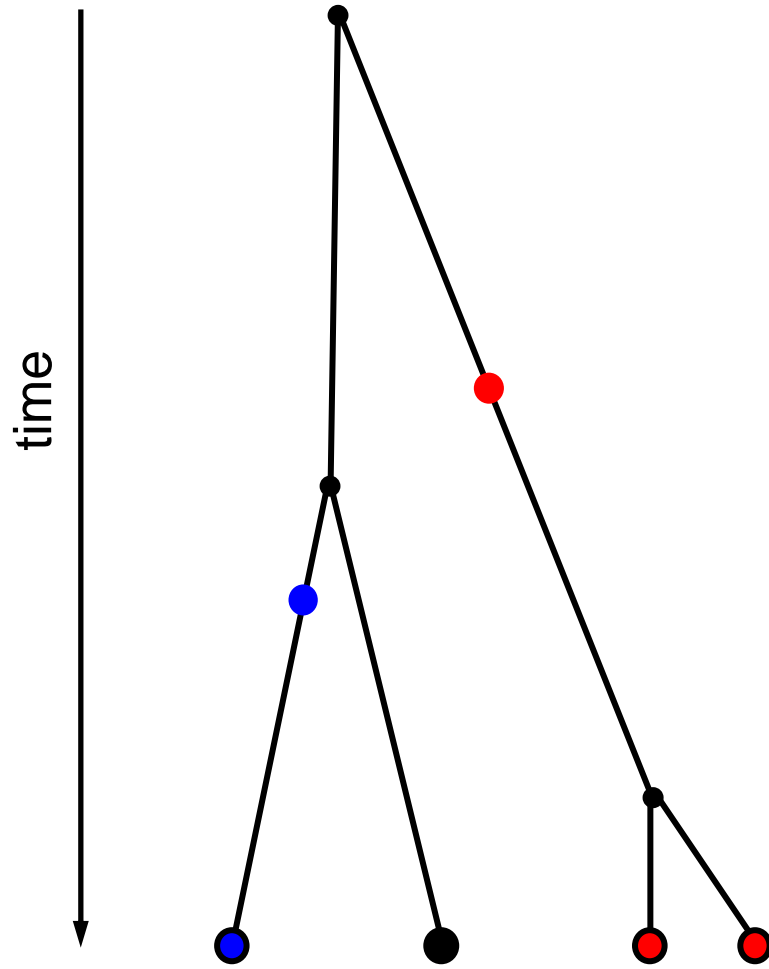


# Patterns of Evolution coalescence process



All Information about the genetic variation pattern is contained in the sample genealogy.

# Patterns of Evolution coalescence process

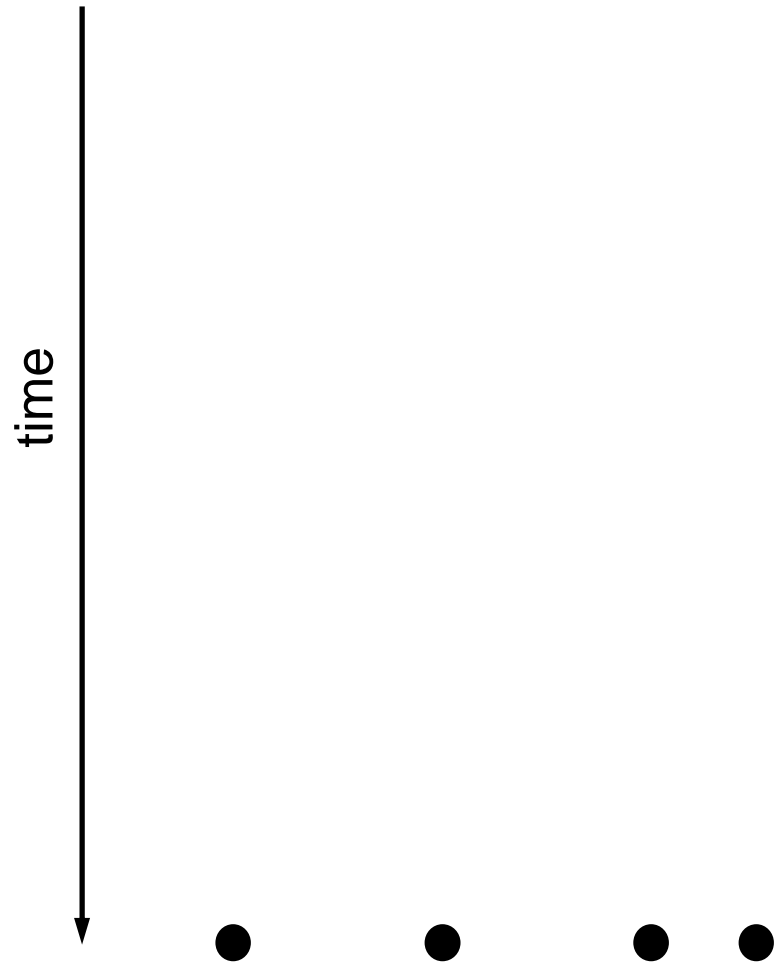


All Information about the genetic variation pattern is contained in the sample genealogy.

Construct a process to generate genealogies:  
„coalescence-process“

# Patterns of Evolution

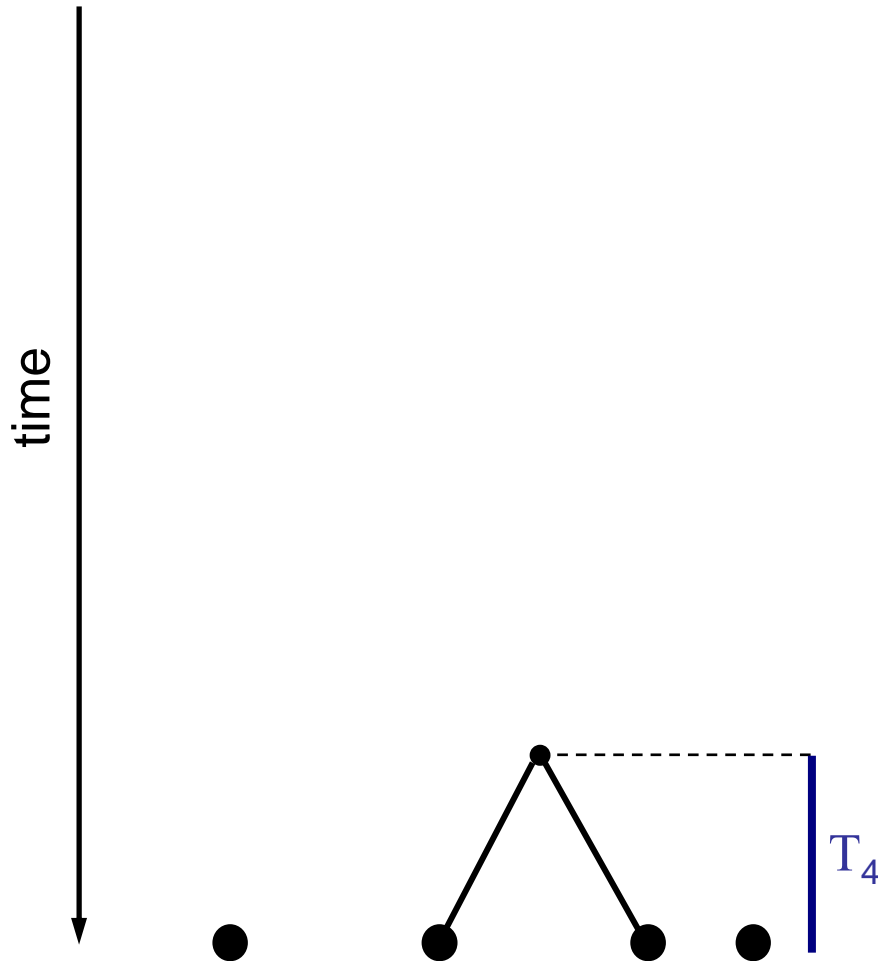
## coalescence process



1. start with the sample and proceed backward in time

# Patterns of Evolution

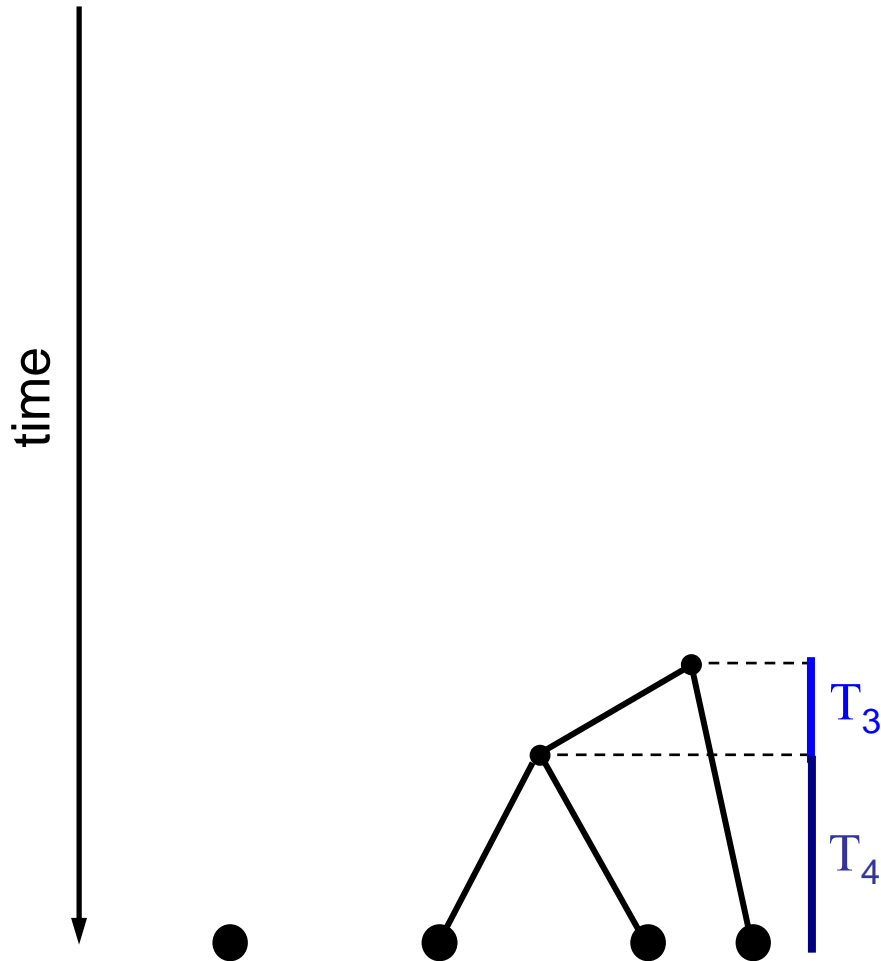
## coalescence process



1. start with the sample and proceed backward in time
2. two randomly drawn genealogical lines merge after time-interval  $T_4$  (exponentially distributed): „**coalescence**“

# Patterns of Evolution

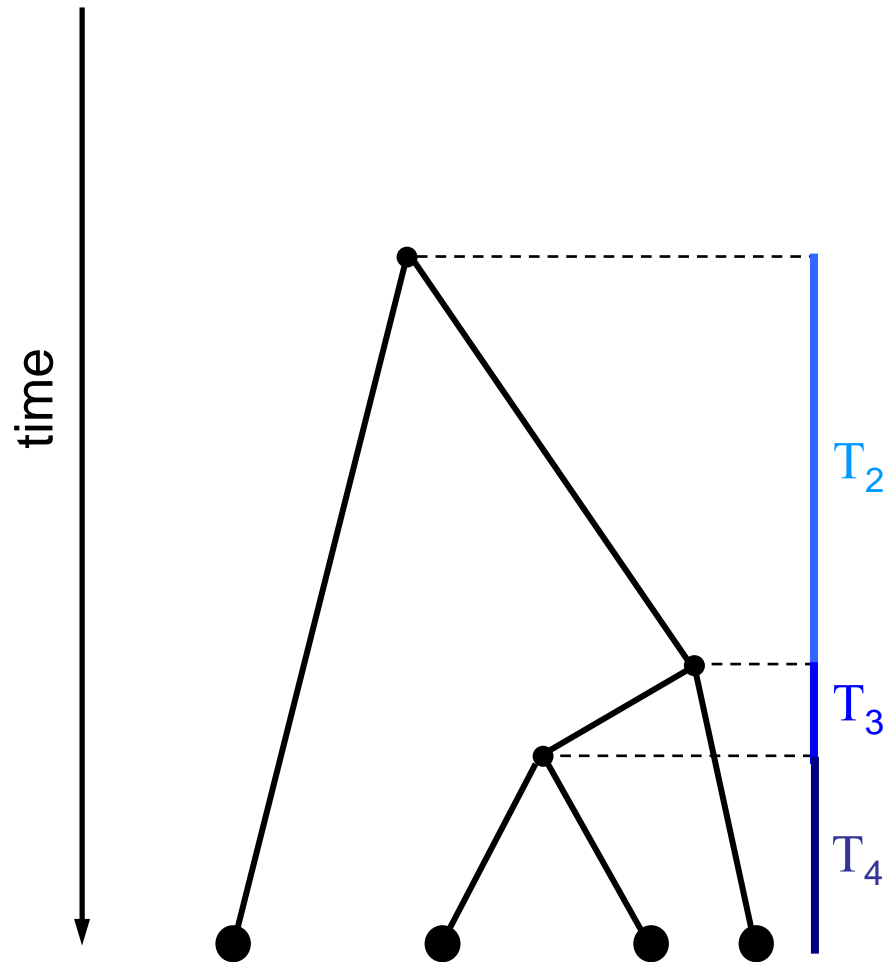
## coalescence process



1. start with the sample and proceed backward in time
2. two randomly drawn genealogical lines merge after time-interval  $T_4$  (exponentially distributed): „**coalescence**“
3. continue process up to the most recent common ancestor

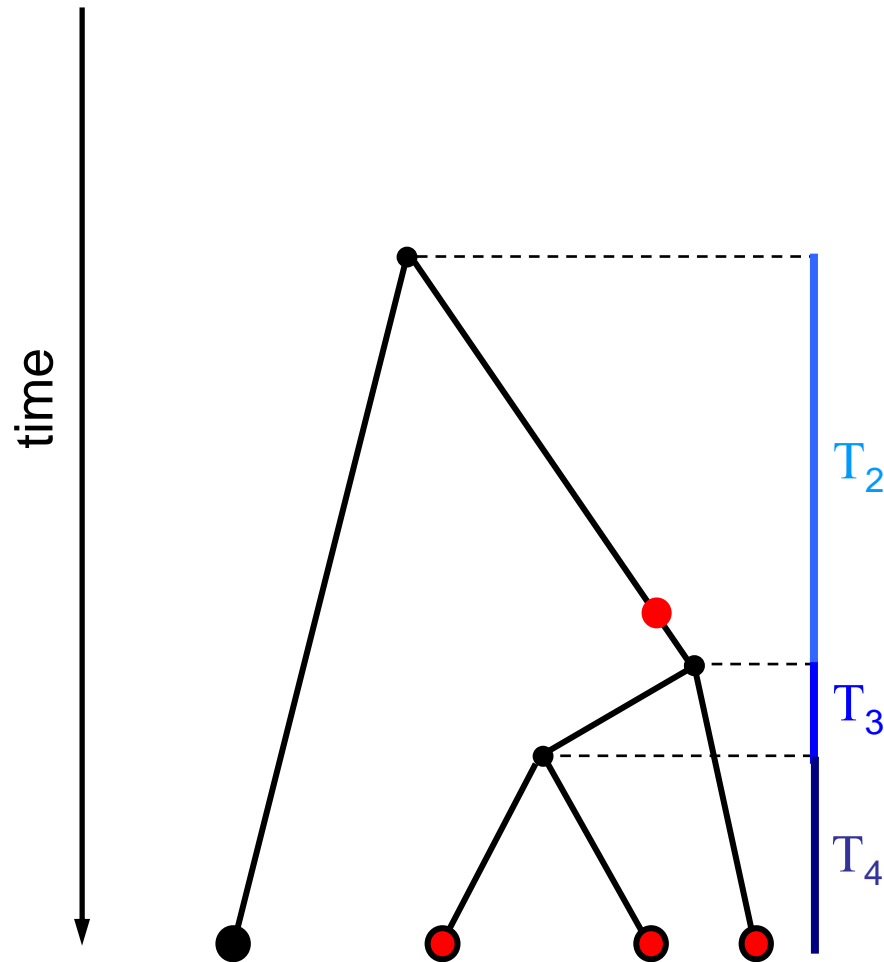
# Patterns of Evolution

## coalescence process



1. start with the sample and proceed backward in time
2. two randomly drawn genealogical lines merge after time-interval  $T_4$  (exponentially distributed): „**coalescence**“
3. continue process up to the most recent common ancestor

# Patterns of Evolution coalescence process

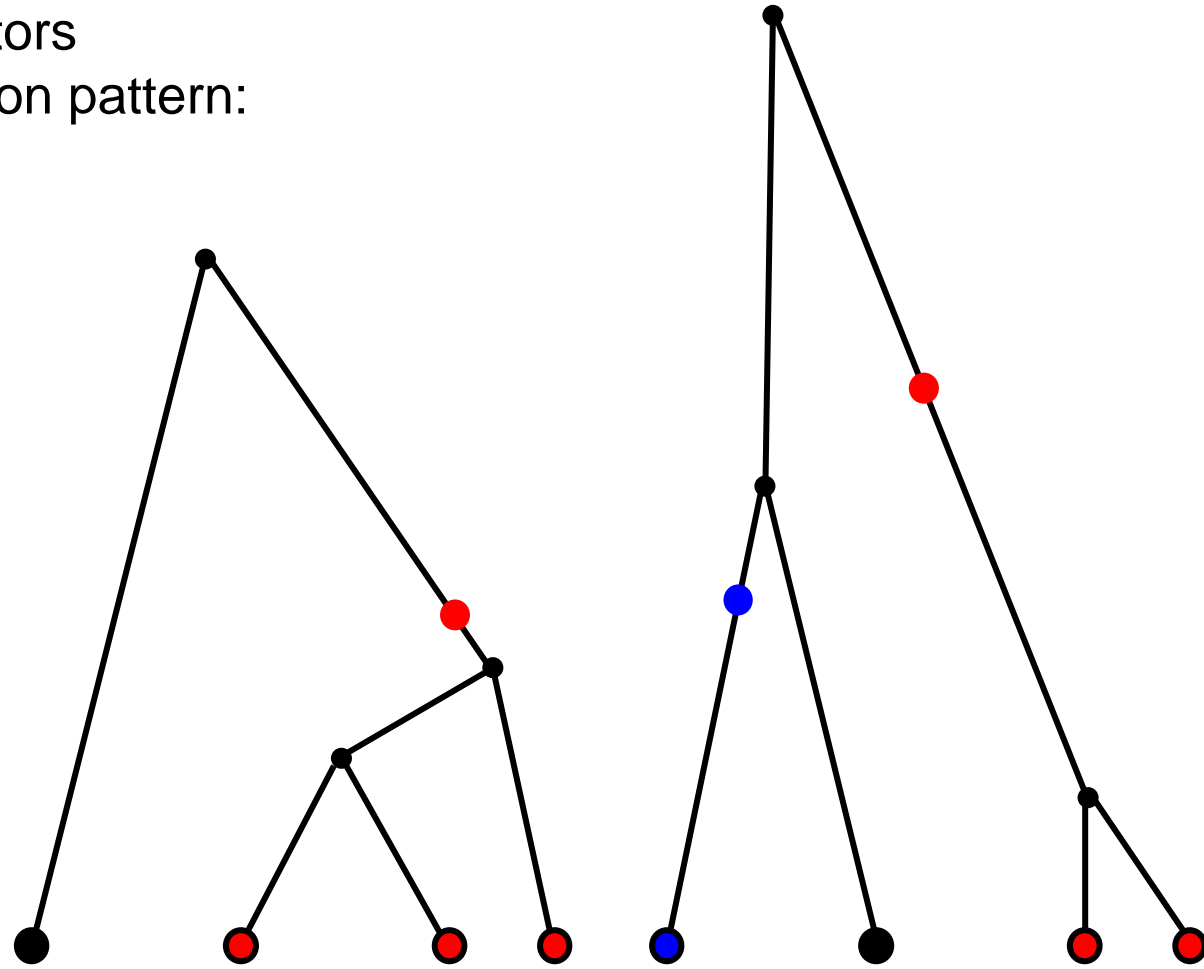


1. start with the sample and proceed backward in time
2. two randomly drawn genealogical lines merge after time-interval  $T_4$  (exponentially distributed): „**coalescence**“
3. continue process up to the most recent common ancestor
4. add **mutationen** (Poisson distributed)

# Patterns of Evolution coalescence process

Three stochastic factors  
determine the variation pattern:

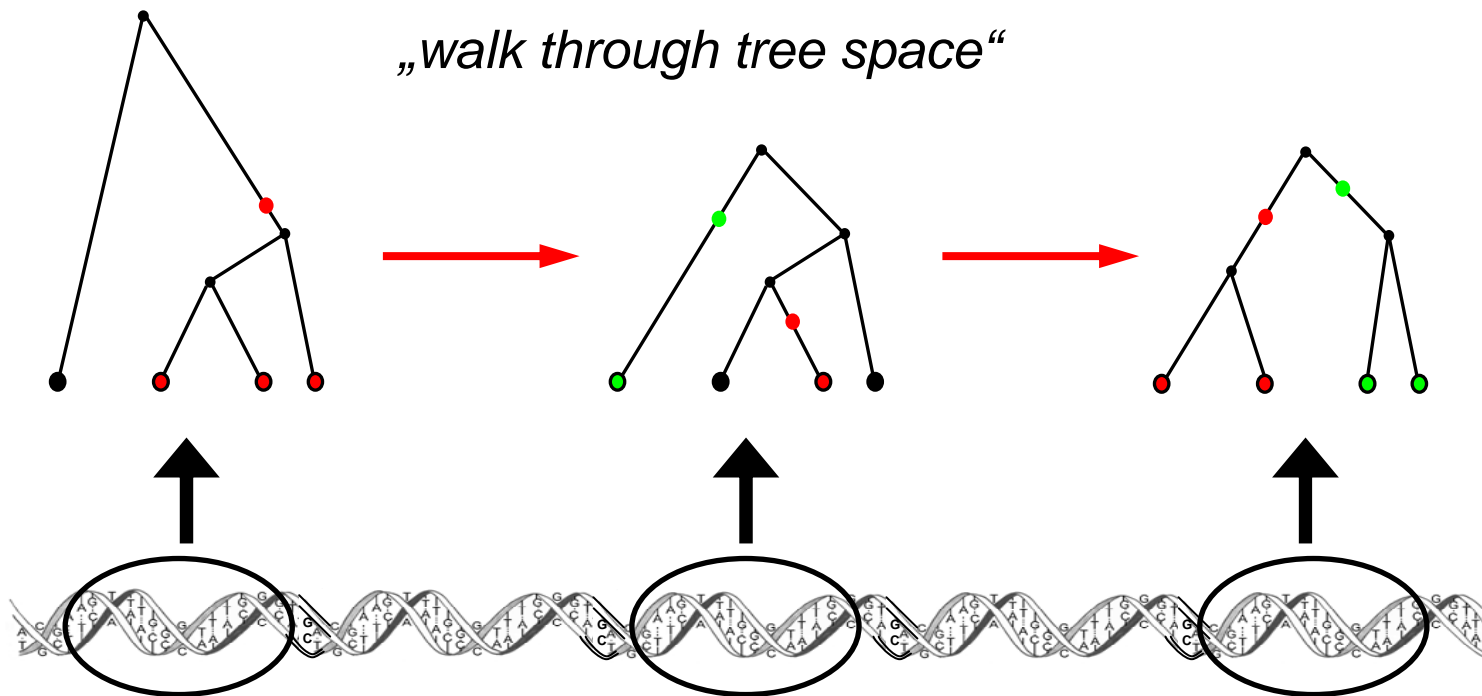
1. tree topology
2. tree length
3. mutation



# Patterns of Evolution coalescence process

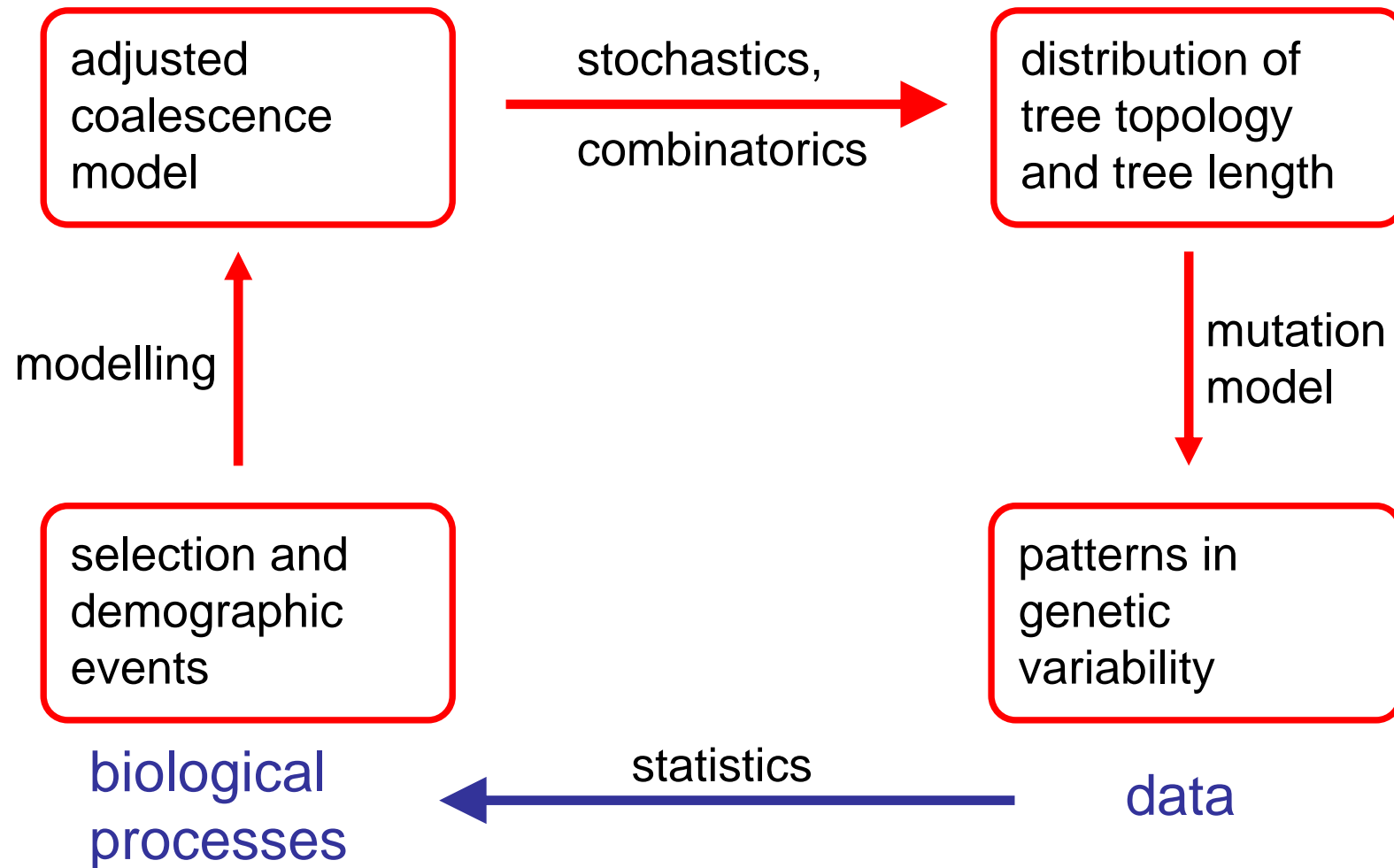
**Recombination** changes the genealogy along the chromosome

- stochastic process in the space of coalescence trees



# Patterns of Evolution

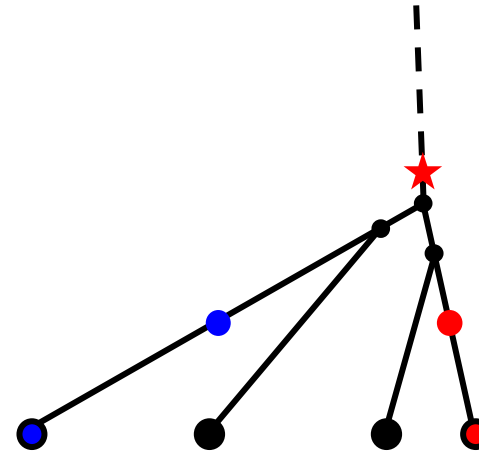
## coalescence and statistical inference



# Patterns of Evolution coalescence with selection

## For example: positive selection

- small trees
  - little variation (*selective sweep*)
- „star-like“ genealogy
  - mutations in the sample visible as „*singletons*“



## „Genome scans“ for evolutionary recent adaptative events:

- Look for regions with low levels of variation and significant surplus of *singletons* relative to the genome average
  - events up to  $\sim 0.1 N_e$  generations can be detected (humans:  $\sim$  homo sapiens – neandertal split)
  - in some cases estimates of age and strength of the adaptive event possible

# Patterns of Evolution

## results and open questions

### Adaptation in humans

- Very **high rates** estimated:
  - some 1000 adaptations during the past 80.000 years
  - More than 7% of all genes in only 3000 generations
- **Some cases well-understood:**
  - adaptation to new diseases (partial malaria resistance)
  - adaptation to changes in life style (agriculture):  
e.g. adult lactose tolerance

### But important **open questions:**

- What is the impact of **spatial expansion** and **population structure** on the statistics?
- In particular: can we construct a tractable model for **coalescence with selection in a spatial model** ?