

# Accounting for genetic recombination in phylogenetic analyses

Ariane Weber & Arthur Kocher

Pathogen phylogenetics workshop

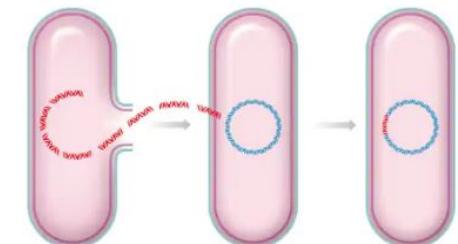
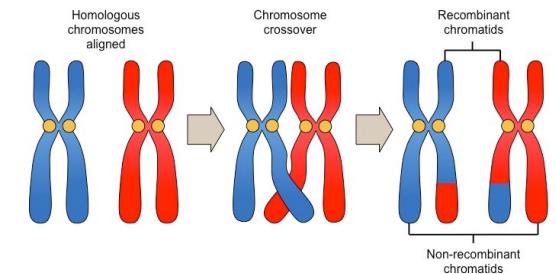
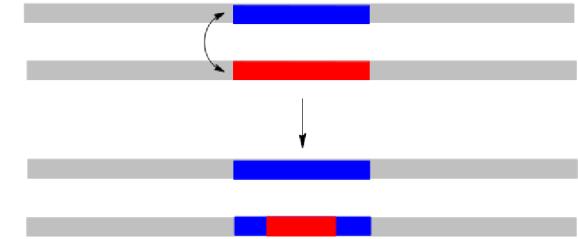
14-17 January 2020

# Genetic recombination

- exchange of genetic material between organisms/chromosomes

# Genetic recombination

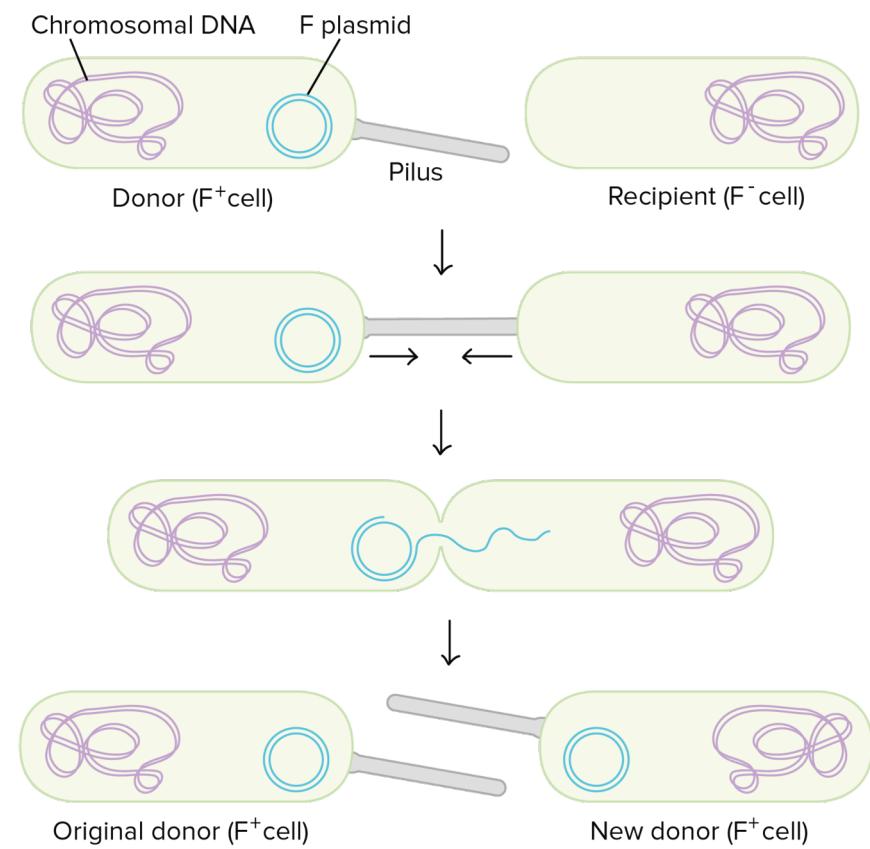
- exchange of genetic material between organisms/chromosomes
- exists in all forms of life
- allows DNA repair (homologous recombination)
- generates genetic diversity: driver of evolution/adaptation



# Genetic recombination

## Genetic recombination in bacteria

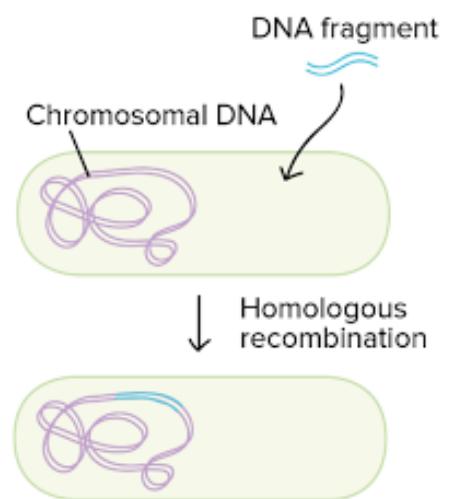
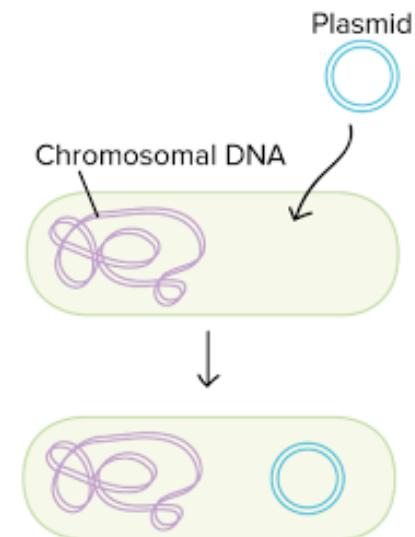
### - conjugation



# Genetic recombination

## Genetic recombination in bacteria

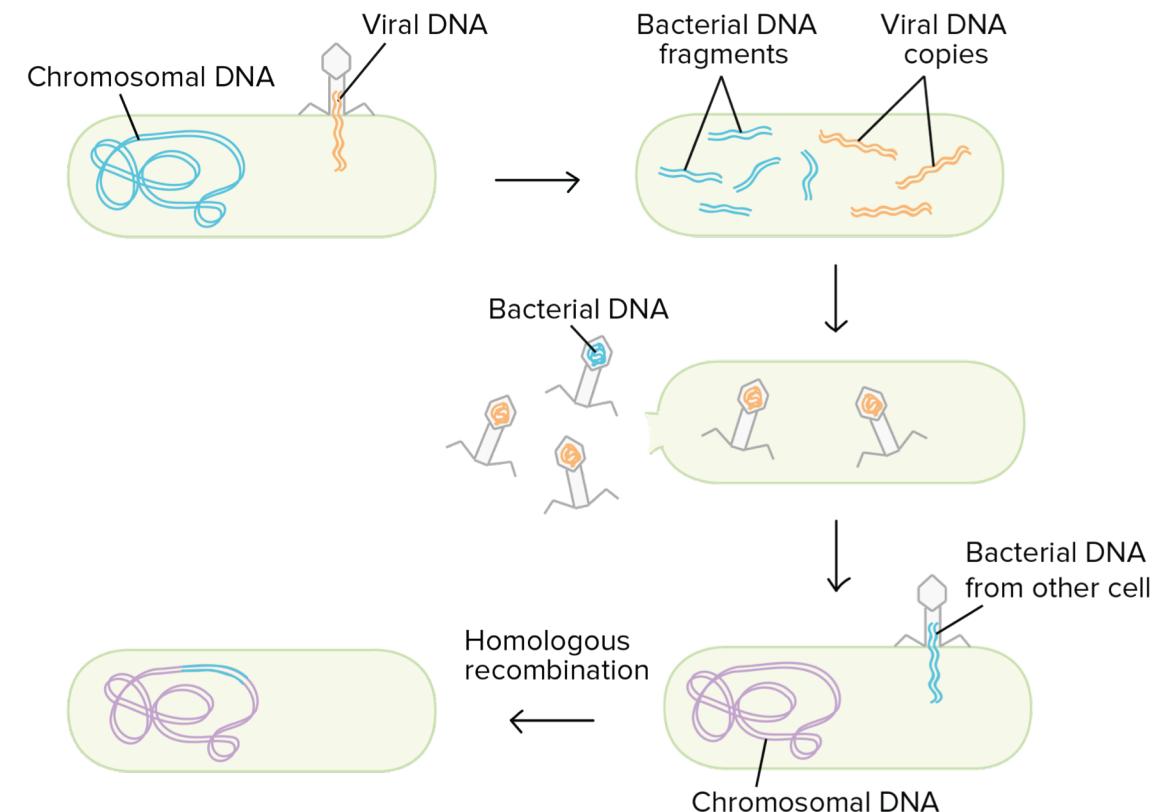
- conjugation
- transformation
- transduction



# Genetic recombination

## Genetic recombination in bacteria

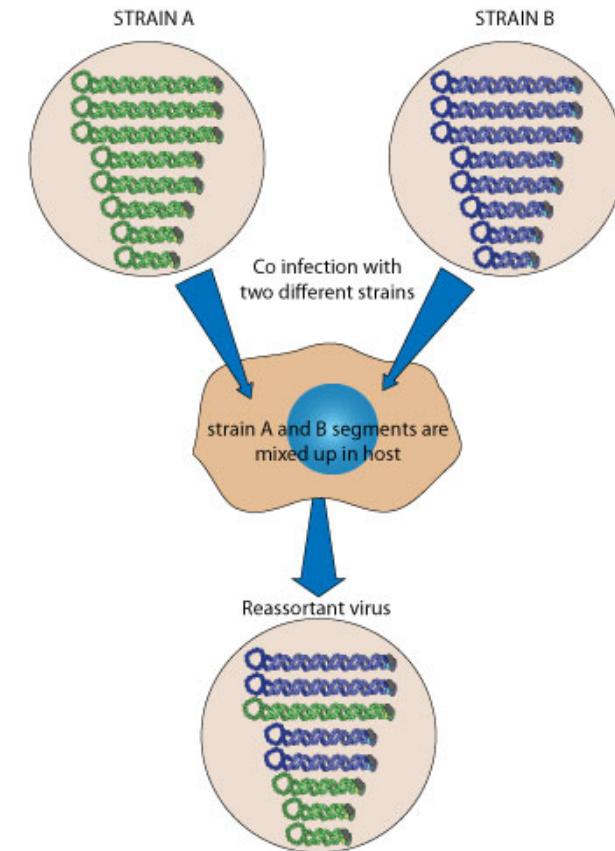
- conjugation
- transformation
- transduction



# Genetic recombination

## Genetic recombination in viruses

- **reassortment:** may occur for any segmented-genome viruses



# Genetic recombination

Genetic recombination in viruses: happens when several viral strains infect the same cell

- reassortment
  - **template-switching**
  - **breakage-rejoining**
- } May lead to homologous or non-homologous recombination in a variety of RNA and DNA viruses: HIV, Hepatitis C virus, Adenoviruses, Herpesviruses, Hepatitis B virus

# Genetic recombination

Genetic recombination in viruses: happens when several viral strains infect the same cell

- reassortment
  - **template-switching**
  - **breakage-rejoining**
- } May lead to homologous or non-homologous recombination in a variety of RNA and DNA viruses: HIV, Hepatitis C virus, Adenoviruses, Herpesviruses, **Hepatitis B virus**

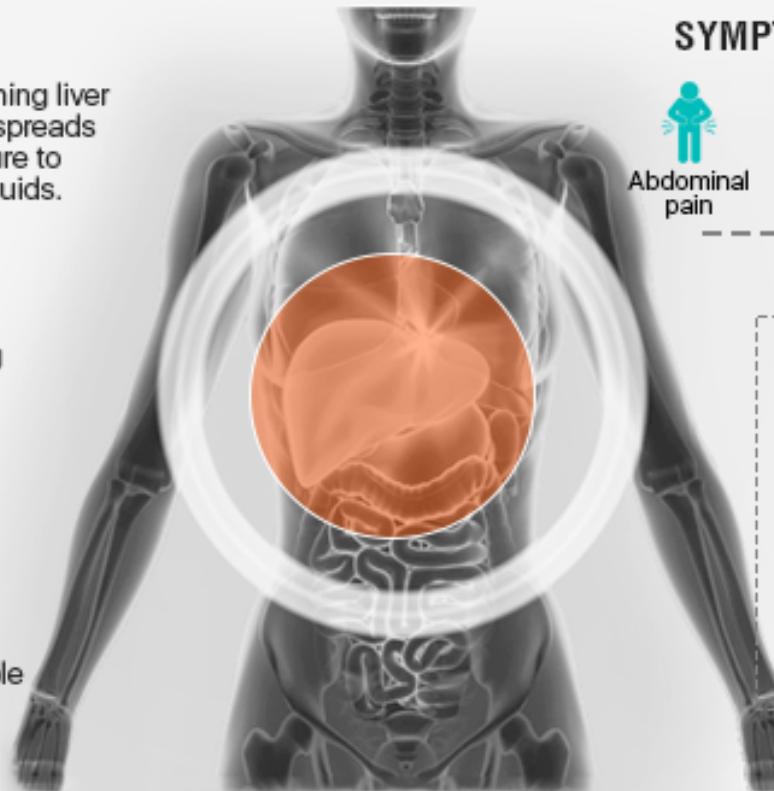
# HEPATITIS **B** FACTS

**1** Is a life-threatening liver disease which spreads through exposure to infected body fluids.

**2** Can be life-threatening

**3** Prevention:  
Hepatitis B vaccination as soon as possible after birth.

NAME



## SYMPTOMS INCLUDE



Abdominal pain



Vomiting



Yellowing of Skin & Eyes

## SPREAD BY



Mother to Child at birth



Exposure to infected blood



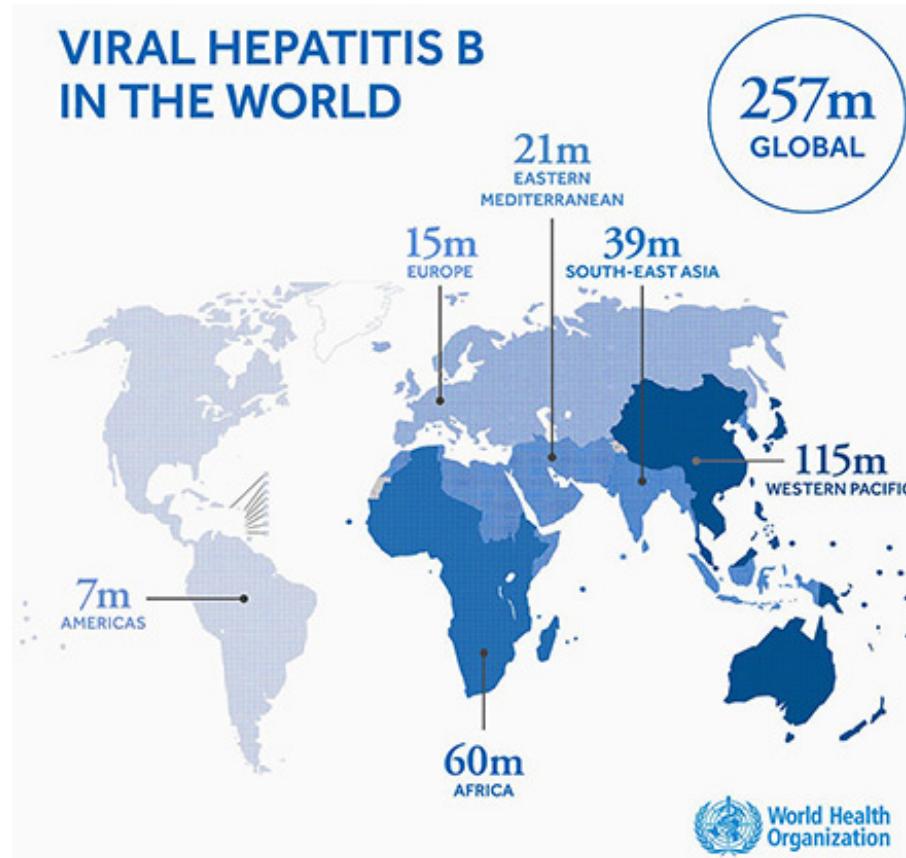
Sexual transmission

& other objects handled by infected person.

# Hepatitis B

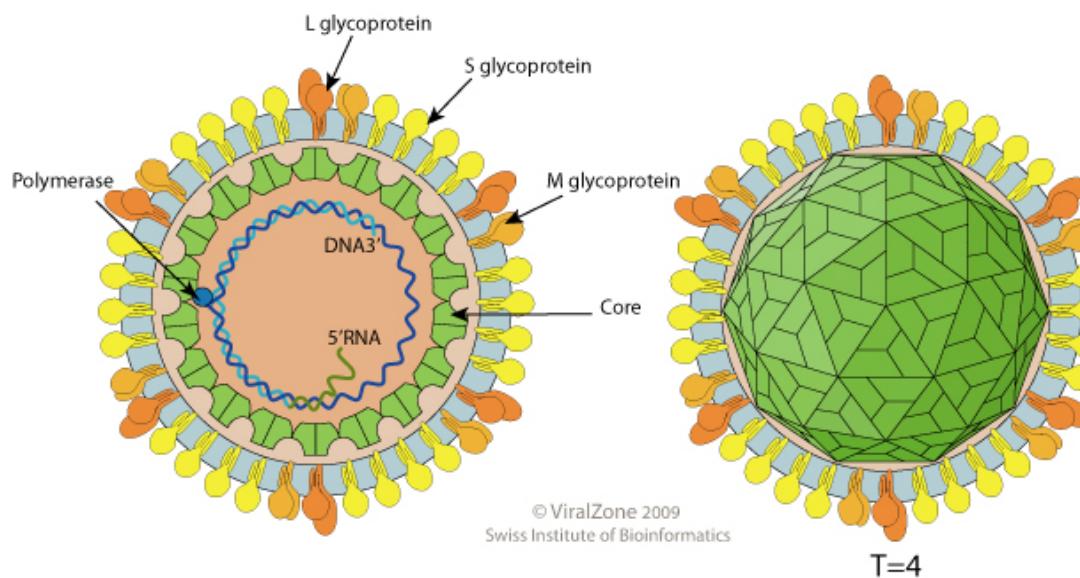
Today: efficient vaccine but:

- still 3.5% of world population infected
- close to 1 M deaths/year (chronic infection)

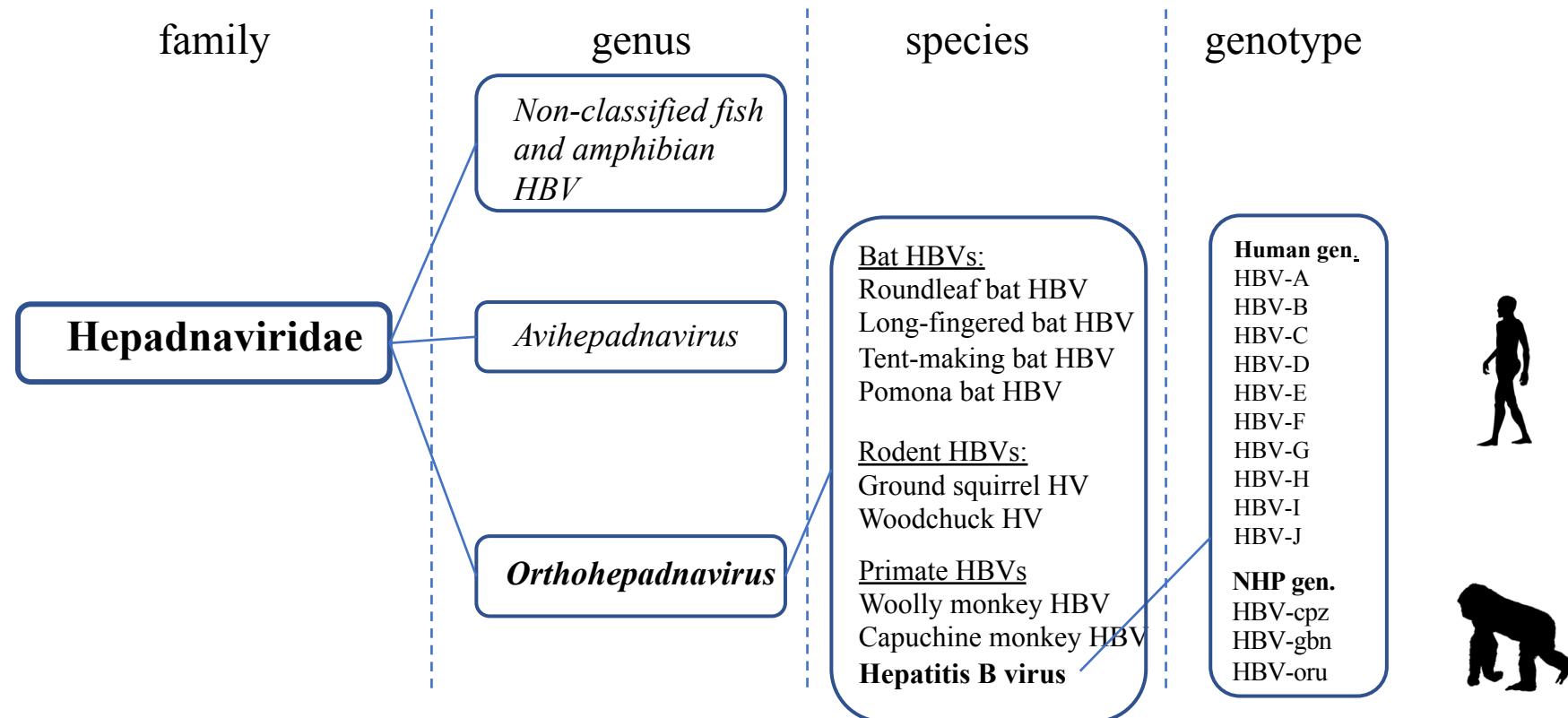


# Hepatitis B virus (HBV)

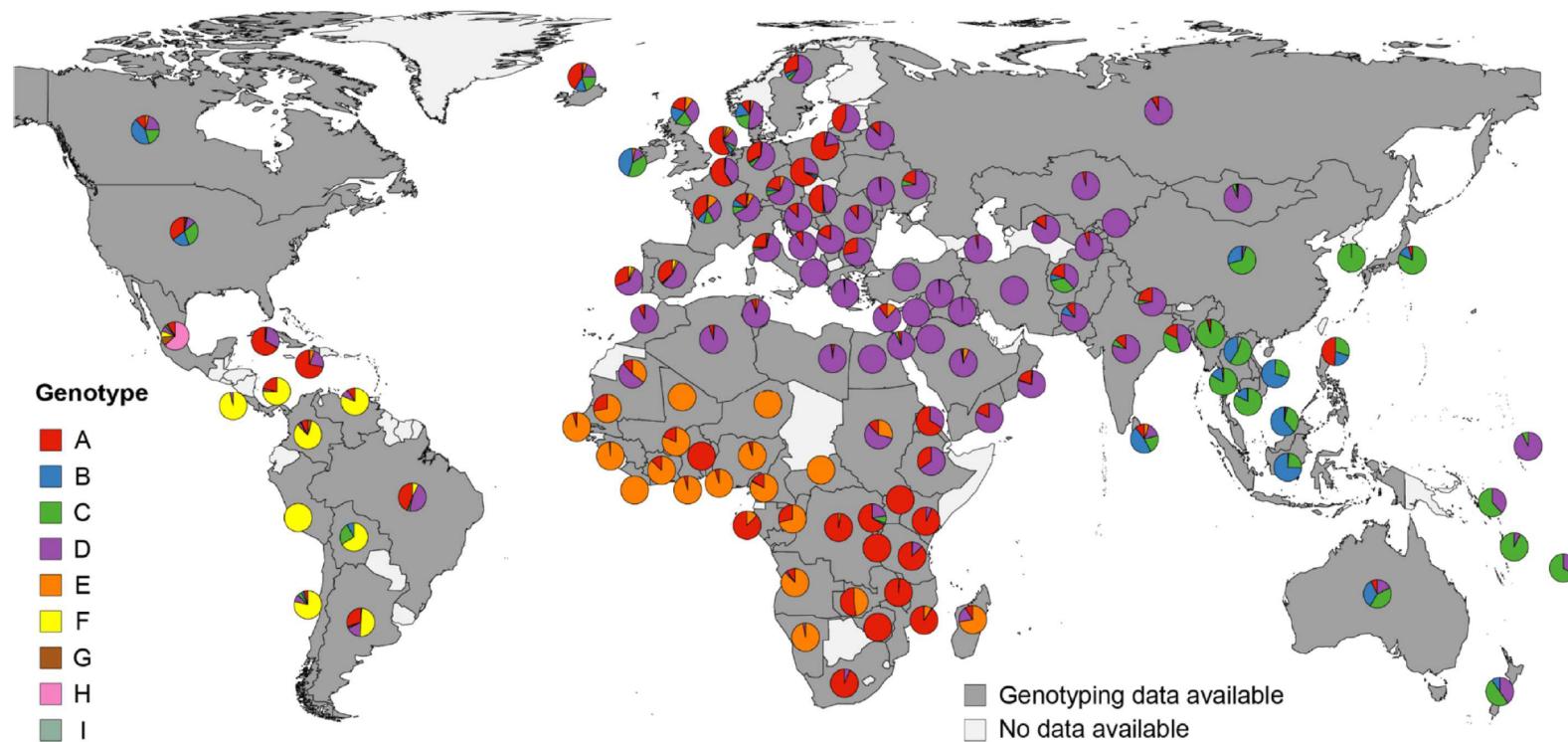
- small enveloped virus
- circular DNA genome:



# Hepatitis B virus: taxonomy (ICTV)

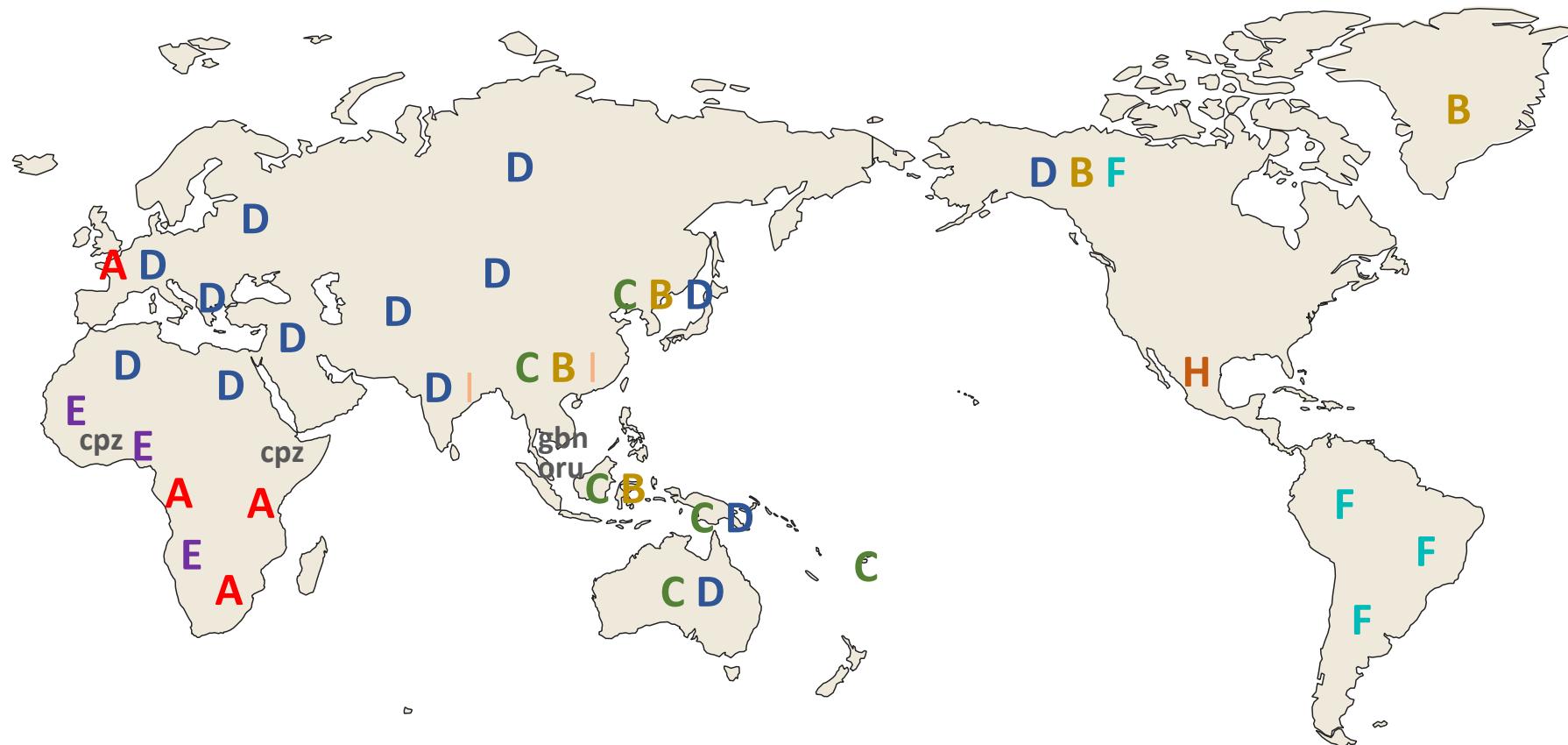


# Hepatitis B virus: ethno-geographic distribution



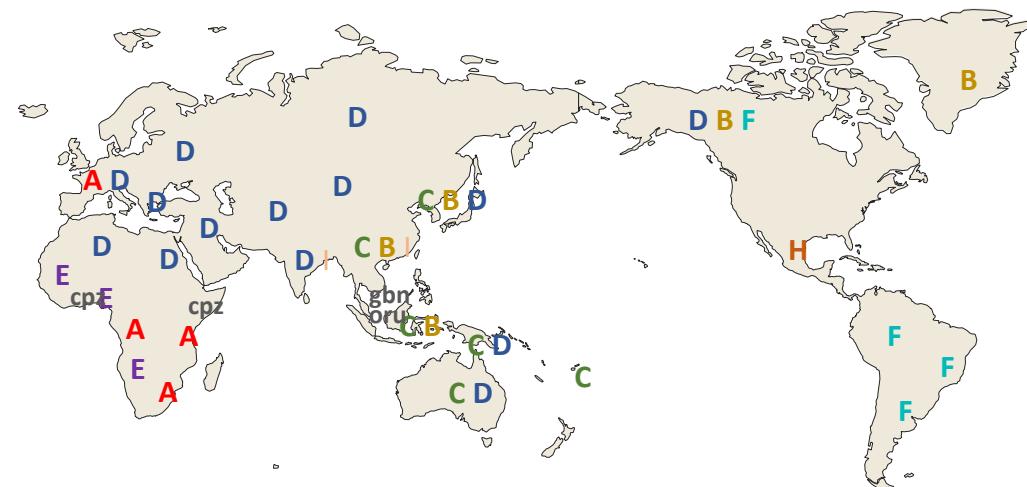
Velkov et al. 2018

# Hepatitis B virus: ethno-geographic distribution

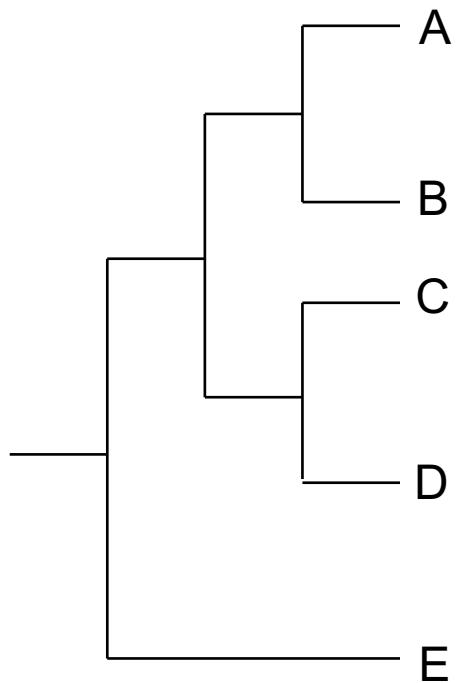


# HBV evolution: questions

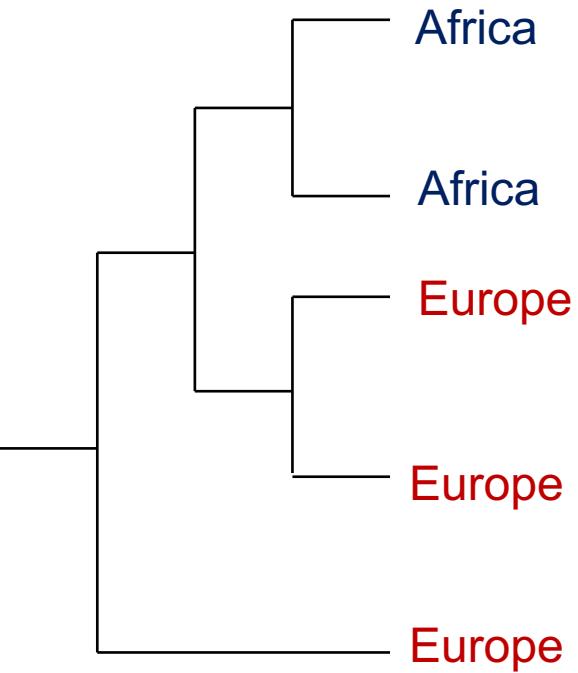
- where and when did HBV emerge in humans ?
- how did it spread worldwide ?
- ancient coevolution or single/multiple recent spillover ?



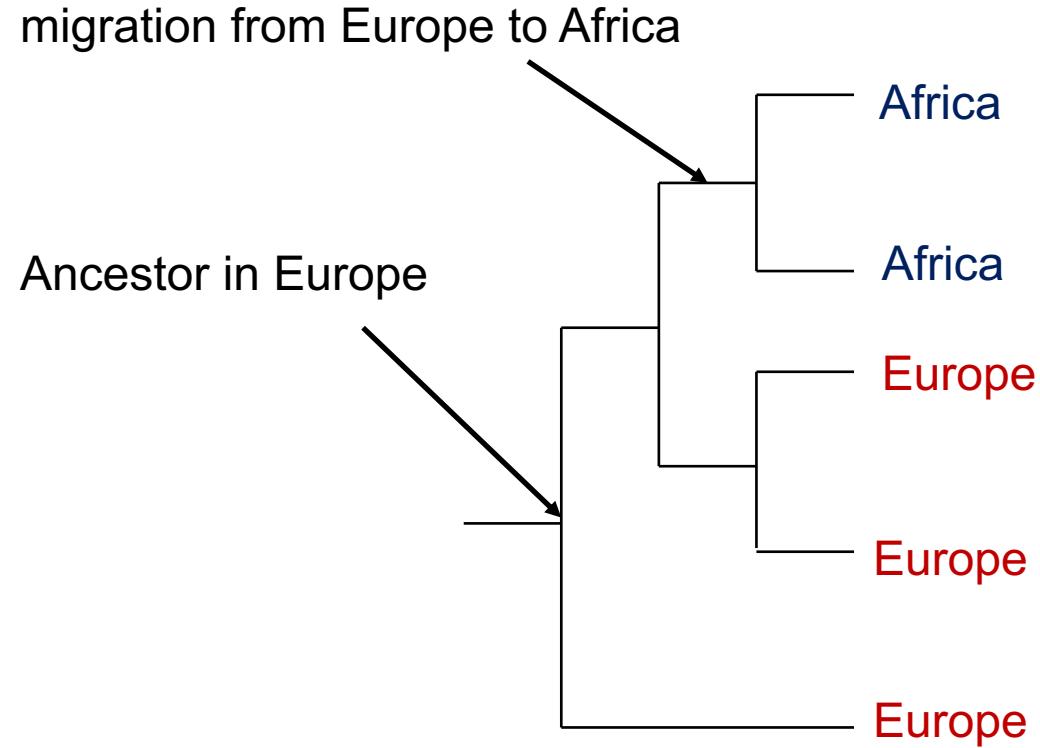
# Testing evolutionary hypothesis using phylogenetics



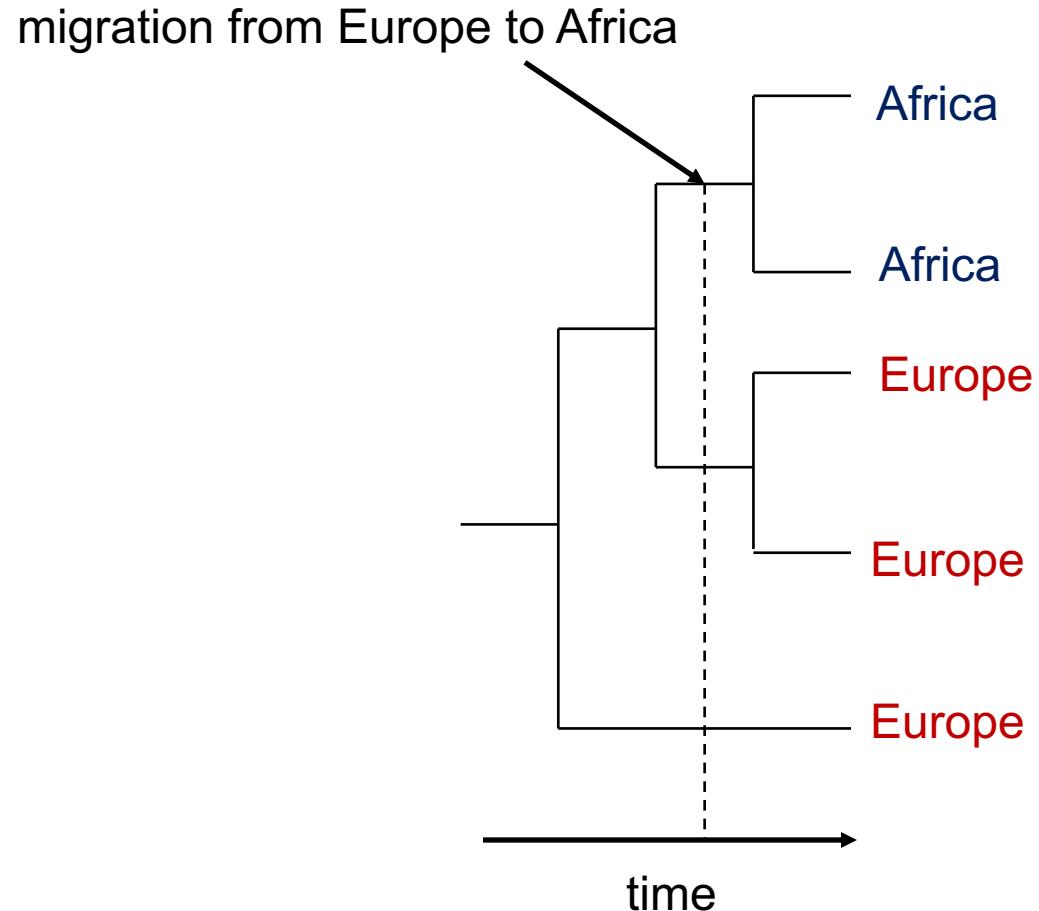
# Testing evolutionary hypothesis using phylogenetics



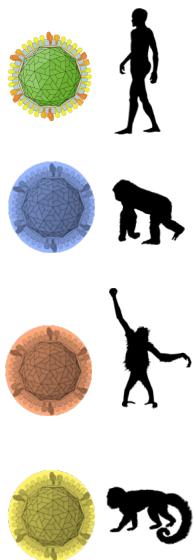
# Testing evolutionary hypothesis using phylogenetics



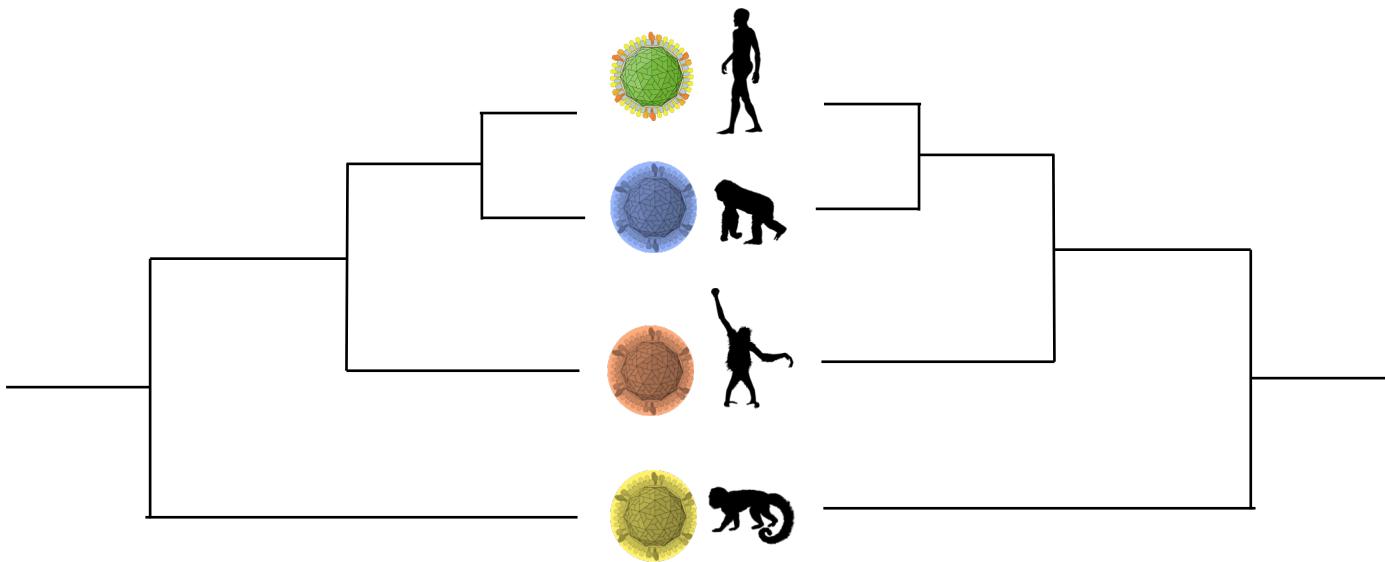
# Testing evolutionary hypothesis using phylogenetics



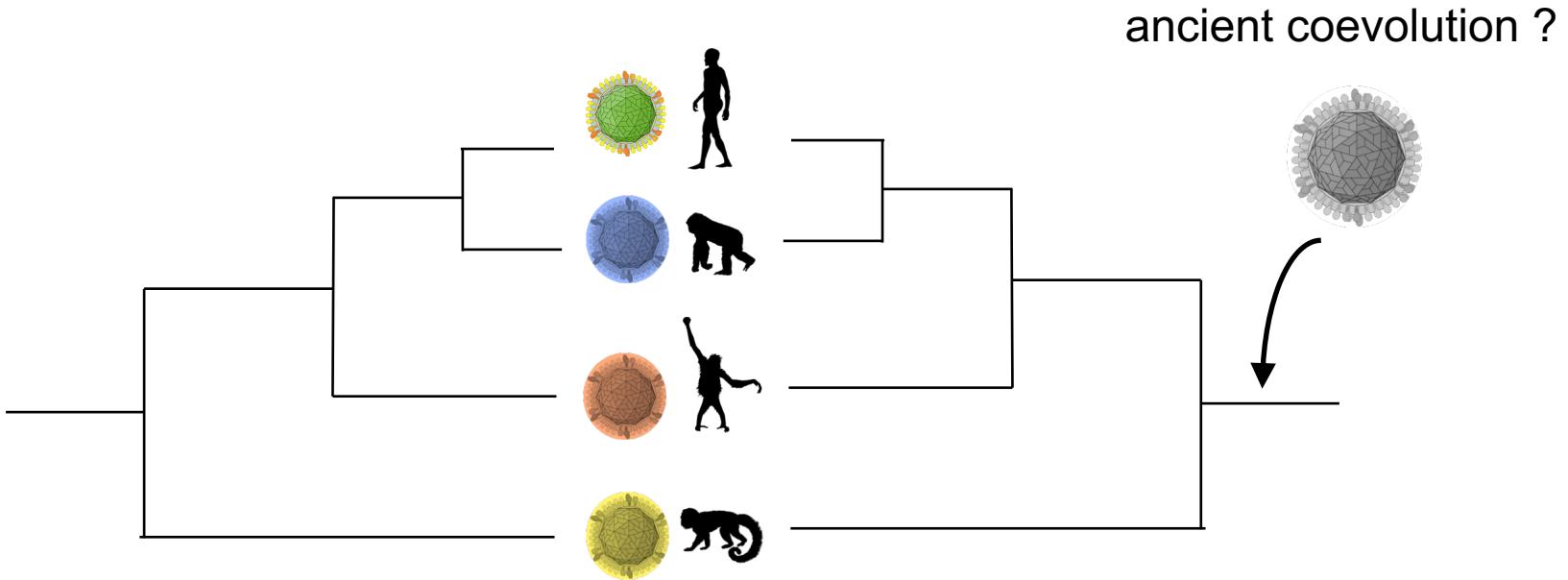
# Testing evolutionary hypothesis using phylogenetics



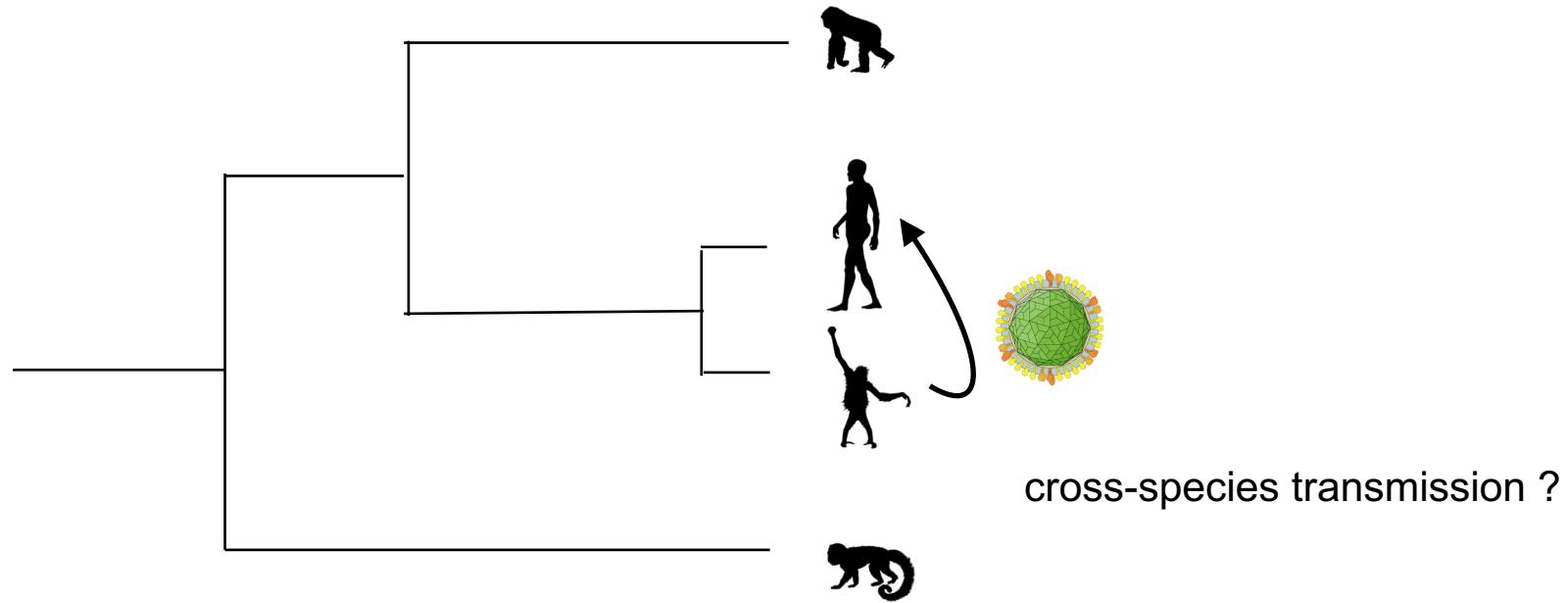
# Testing evolutionary hypothesis using phylogenetics



# Testing evolutionary hypothesis using phylogenetics



# Testing evolutionary hypothesis using phylogenetics



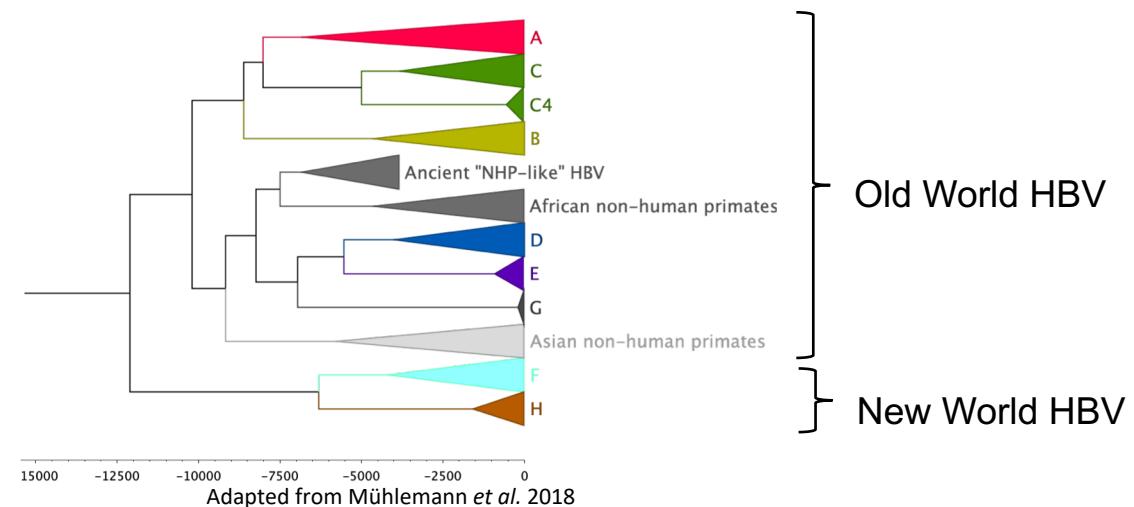
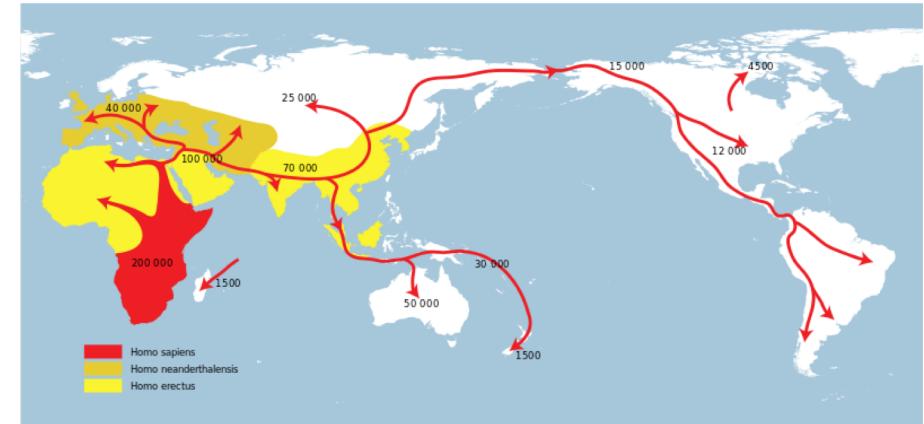
# HBV phylogeographic hypotheses

## Origin of HBV in humans ?

### The “out of Africa” hypothesis:

HBV has coevolved with *Homo sapiens* for a long time and followed his migration out of Africa. Could explain:

- worldwide prevalence
- geographical structure of genetic diversity
- highest diversity in the old world
- presence in other great apes



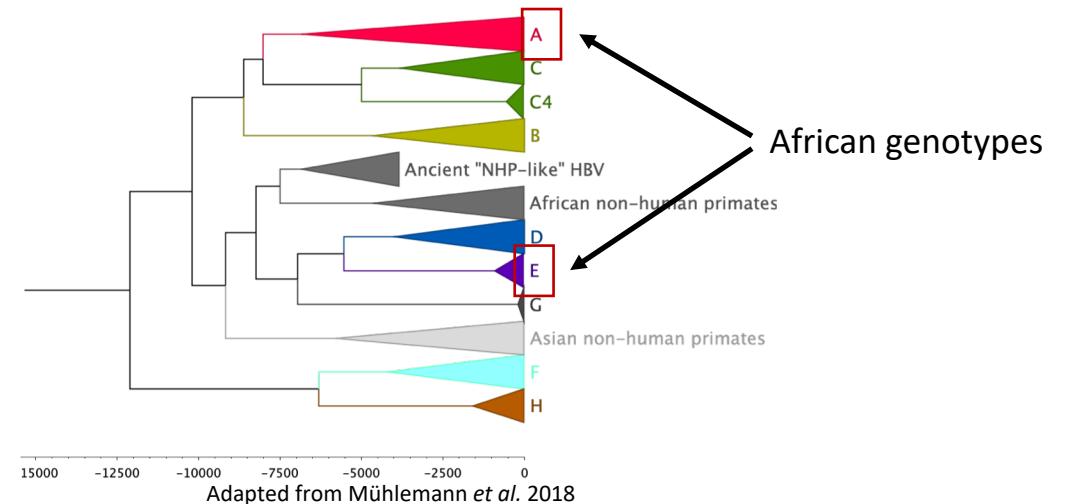
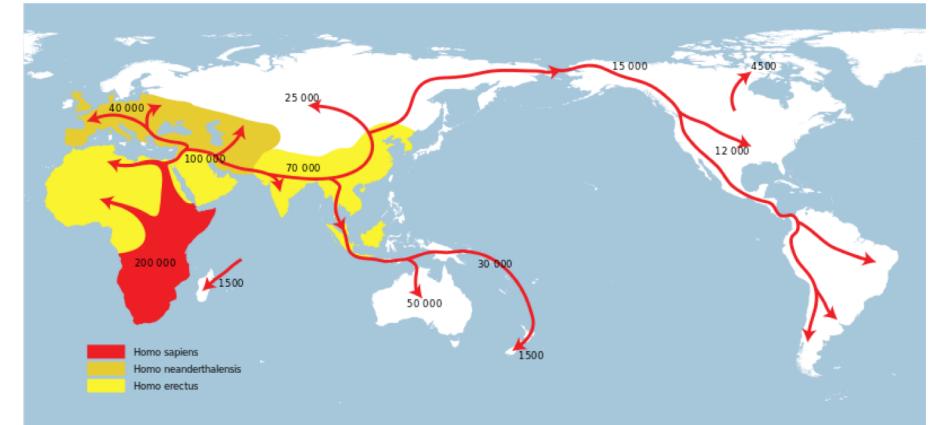
# HBV phylogeographic hypotheses

## Origin of HBV in humans ?

The “out of Africa” hypothesis:

BUT

- African genotypes rather terminal



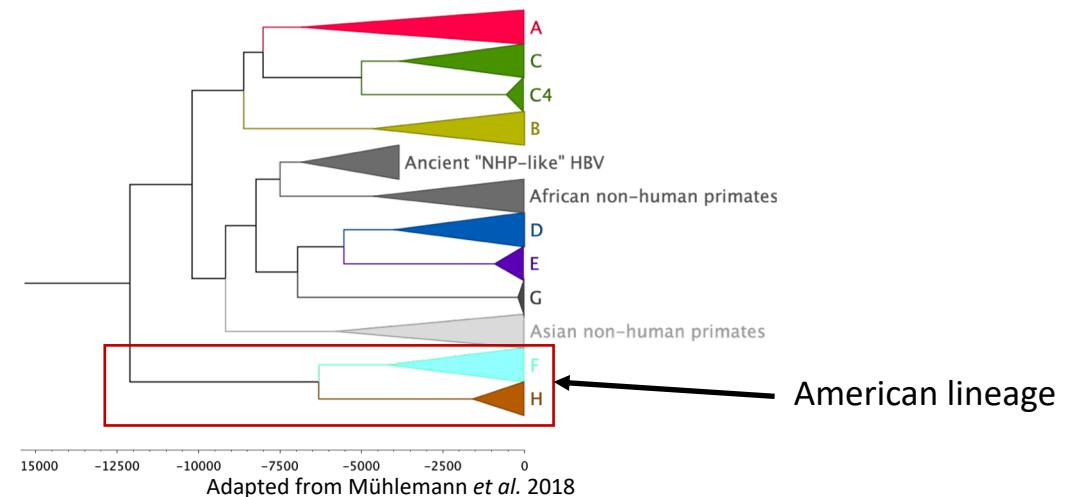
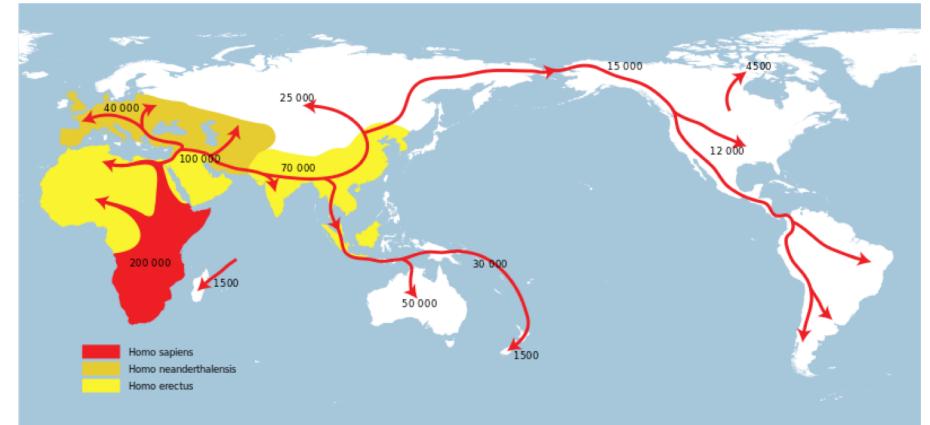
# HBV phylogeographic hypotheses

## Origin of HBV in humans ?

The “out of Africa” hypothesis:

BUT

- African genotypes rather terminal
- American genotypes basal in the phylogenetic tree



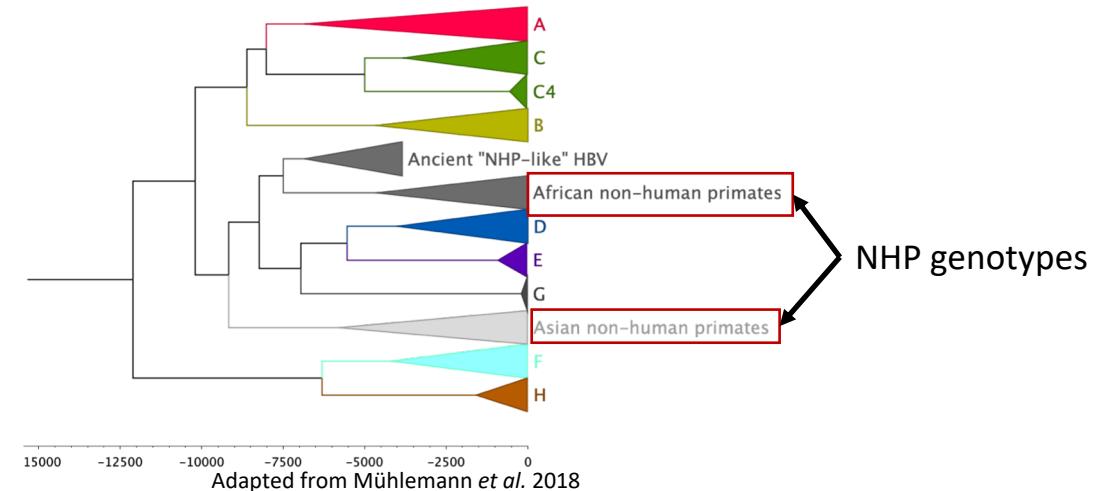
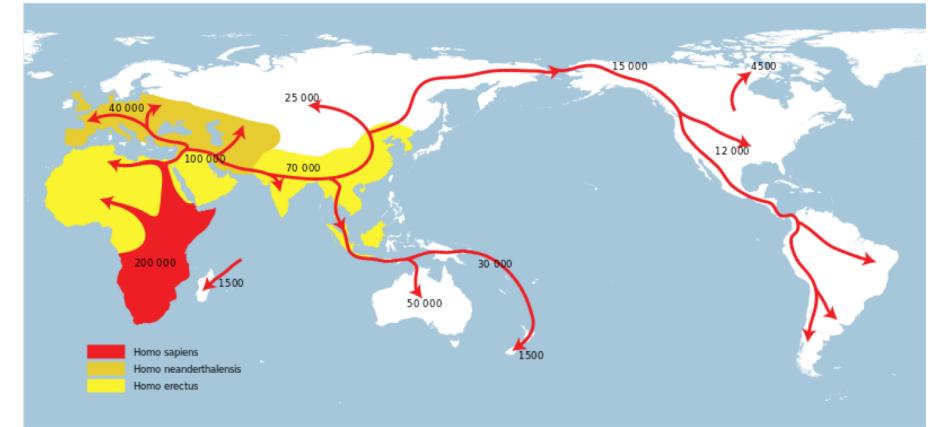
# HBV phylogeographic hypotheses

## Origin of HBV in humans ?

The “out of Africa” hypothesis:

BUT

- African genotypes rather terminal
- American genotypes basal in the phylogenetic tree
- Non-human primate HBV niched within human HBV



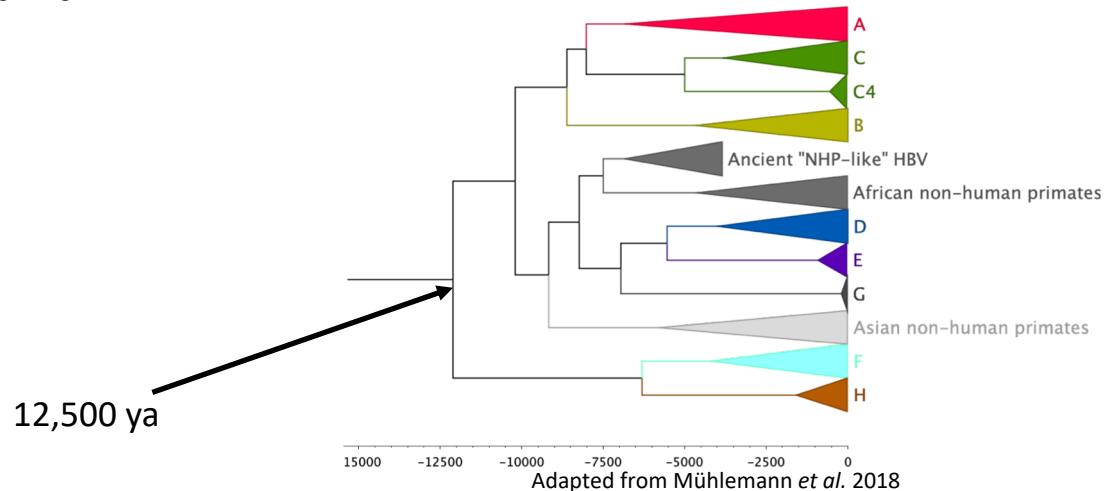
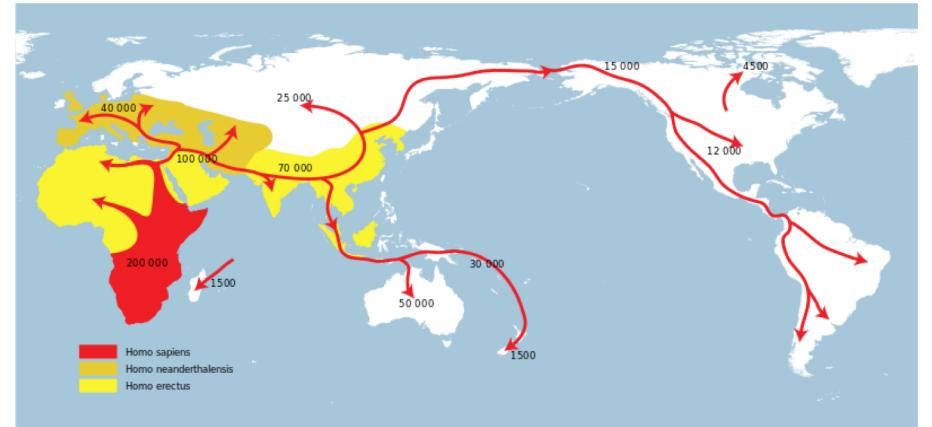
# HBV phylogeographic hypotheses

## Origin of HBV in humans ?

The “out of Africa” hypothesis:

BUT

- African genotypes rather terminal
- American genotypes basal in the phylogenetic tree
- Non-human primate HBV niched within human HBV
- TMRCA = 12,500 ya



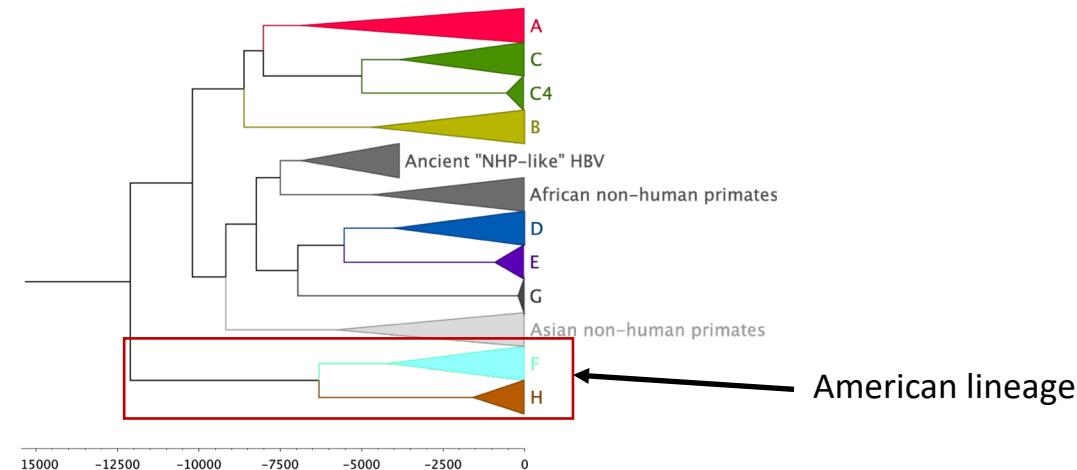
# HBV phylogeographic hypotheses

## Origin of HBV in humans ?

### The “New World” hypothesis:

HBV originated in the Americas and was brought into other continents after contact ~500 ya.

- basal position of American Lineages



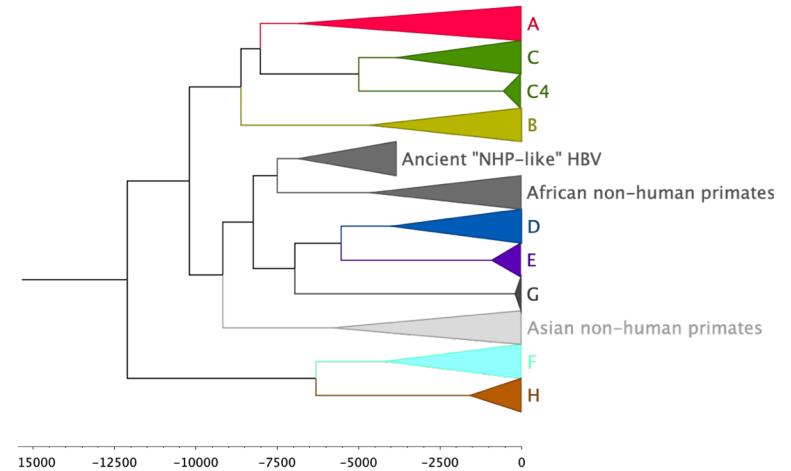
# HBV phylogeographic hypotheses

## Origin of HBV in humans ?

### The “New World” hypothesis:

HBV originated in the Americas and was brought into other continents after contact ~500 ya.

- basal position of American Lineages
- highest diversity of hepadnaviruses in the Americas



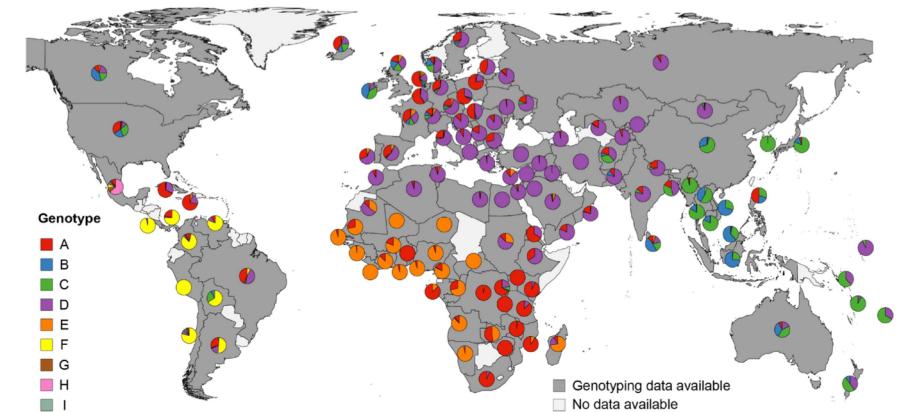
# HBV phylogeographic hypotheses

## Origin of HBV in humans ?

The “New World” hypothesis:

BUT:

- hardly explains worldwide prevalence
- lineages specific to indigenous populations in Eurasia and to NHP



Velkov et al. 2018

# HBV phylogeographic hypotheses

## Origin of HBV in humans ?

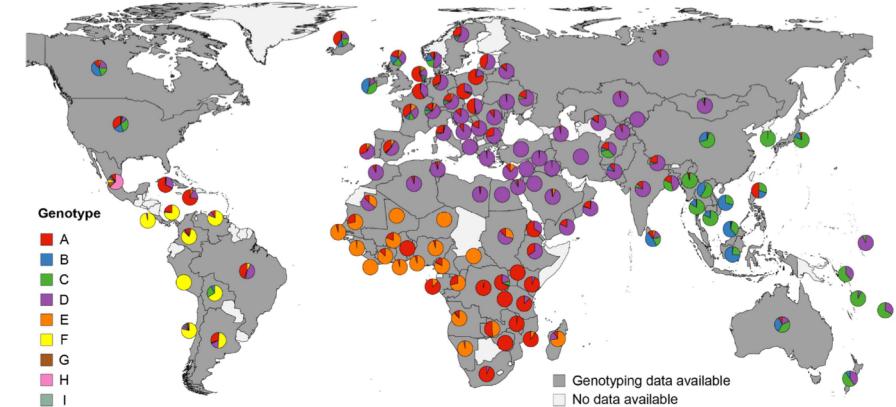
The “New World” hypothesis:

BUT:

- hardly explains worldwide prevalence
- lineages specific to indigenous populations in Eurasia and to

NHP

- ancient genome: HBV already in Europe during Neolithic



Velkov et al. 2018

### Neolithic and medieval virus genomes reveal complex evolution of hepatitis B

Ben Krause-Kyora<sup>1,2†\*</sup>, Julian Susat<sup>1†</sup>, Felix M Key<sup>2</sup>, Denise Kühnert<sup>2,3</sup>, Esther Bosse<sup>1,4</sup>, Alexander Immel<sup>1,2</sup>, Christoph Rinne<sup>5</sup>, Sabin-Chrissin Kornell<sup>1</sup>, Diego Yepes<sup>4</sup>, Sören Franzenburg<sup>1</sup>, Henrike O Heyne<sup>6,7,8</sup>, Thomas Meier<sup>9,10</sup>, Sandra Lösch<sup>11</sup>, Harald Meller<sup>12</sup>, Susanne Friederich<sup>12</sup>, Nicole Nicklisch<sup>12,13</sup>, Kurt W Alt<sup>12,13,14,15</sup>, Stefan Schreiber<sup>1,16</sup>, Andreas Tholey<sup>4</sup>, Alexander Herbig<sup>2</sup>, Almut Nebel<sup>1</sup>, Johannes Krause<sup>2\*</sup>

### Ancient hepatitis B viruses from the Bronze Age to the Medieval period

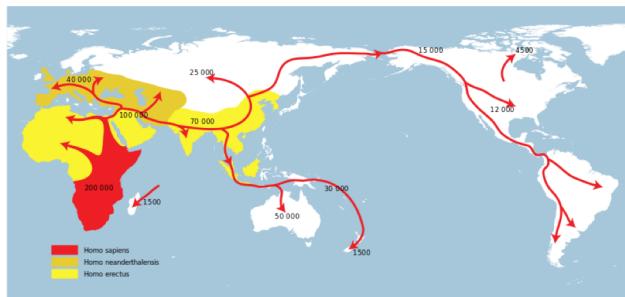
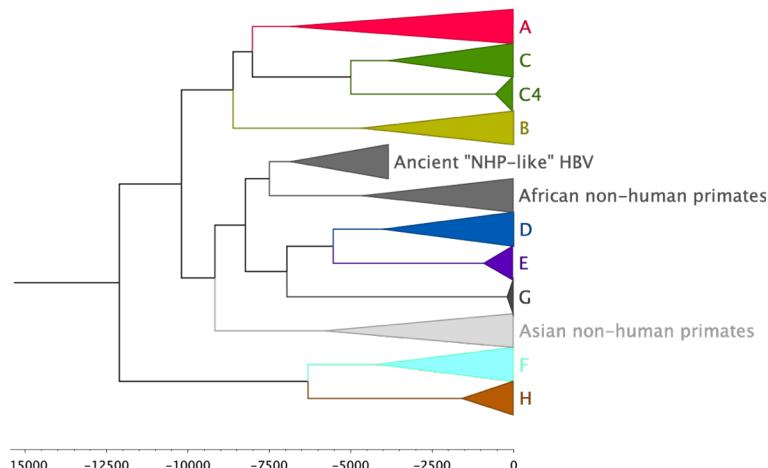
Barbara Mühlmann<sup>1,29</sup>, Terry C. Jones<sup>1,2,29</sup>, Peter de Barros Damgaard<sup>3,29</sup>, Morten E. Allentoft<sup>3,29</sup>, Irina Shevchenko<sup>4</sup>, Andrey Logvin<sup>4</sup>, Emma Usmanova<sup>5</sup>, Irina P. Panyushkina<sup>6</sup>, Bazarsuren Boldgiv<sup>7</sup>, Tsevel Bazarsuren<sup>8</sup>, Kadicha Tashbaeva<sup>9</sup>, Victor Merz<sup>10</sup>, Nina Lau<sup>11</sup>, Václav Smrká<sup>12</sup>, Dmitry Voyakin<sup>13</sup>, Egor Kitov<sup>14</sup>, Andrey Epimakhov<sup>15</sup>, Dalia Pokutta<sup>16</sup>, Magdalna Vicze<sup>17</sup>, T. Douglas Price<sup>18</sup>, Vyacheslav Moiseyev<sup>19</sup>, Anders J. Hansen<sup>3</sup>, Ludovic Orlando<sup>3,20</sup>, Simon Rasmussen<sup>21</sup>, Martin Sikora<sup>3</sup>, Lasse Vinner<sup>3</sup>, Albert D. M. E. Osterhaus<sup>22</sup>, Derek J. Smith<sup>1</sup>, Dieter Glebe<sup>23,24</sup>, Ron A. M. Fouchier<sup>25</sup>, Christian Drosten<sup>2,26</sup>, Karl-Göran Sjögren<sup>18</sup>, Kristian Kristiansen<sup>18</sup> & Eske Willerslev<sup>3,27,28\*</sup>

# HBV phylogeographic hypotheses

Complex phylogeographic history ?

OR

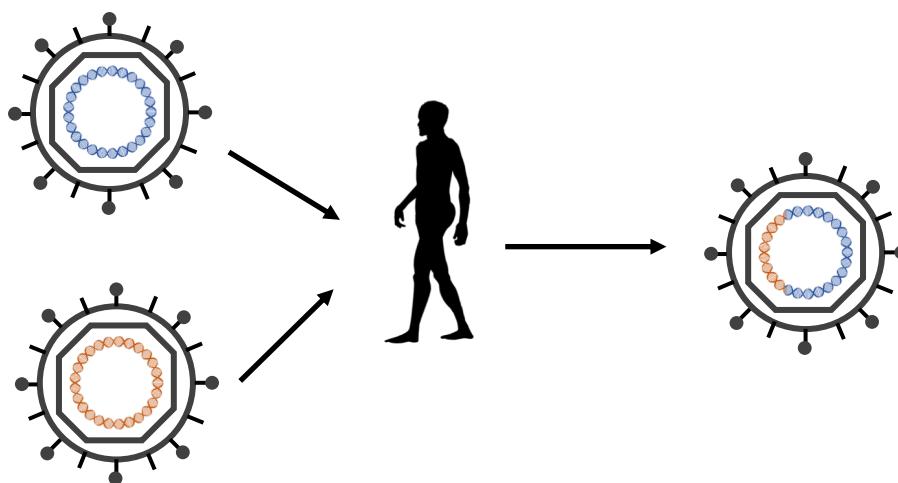
Biased phylogenetic inference ?



Adapted from Mühlmann *et al.* 2018

# Genetic recombination in HBV

- HBV is able to recombine (mechanism ?)
- chronicity + high prevalence = opportunities for multiple infection
- entire genotypes have been identified as recombinants



# Genetic recombination and phylogenetic reconstruction :



Parent 1



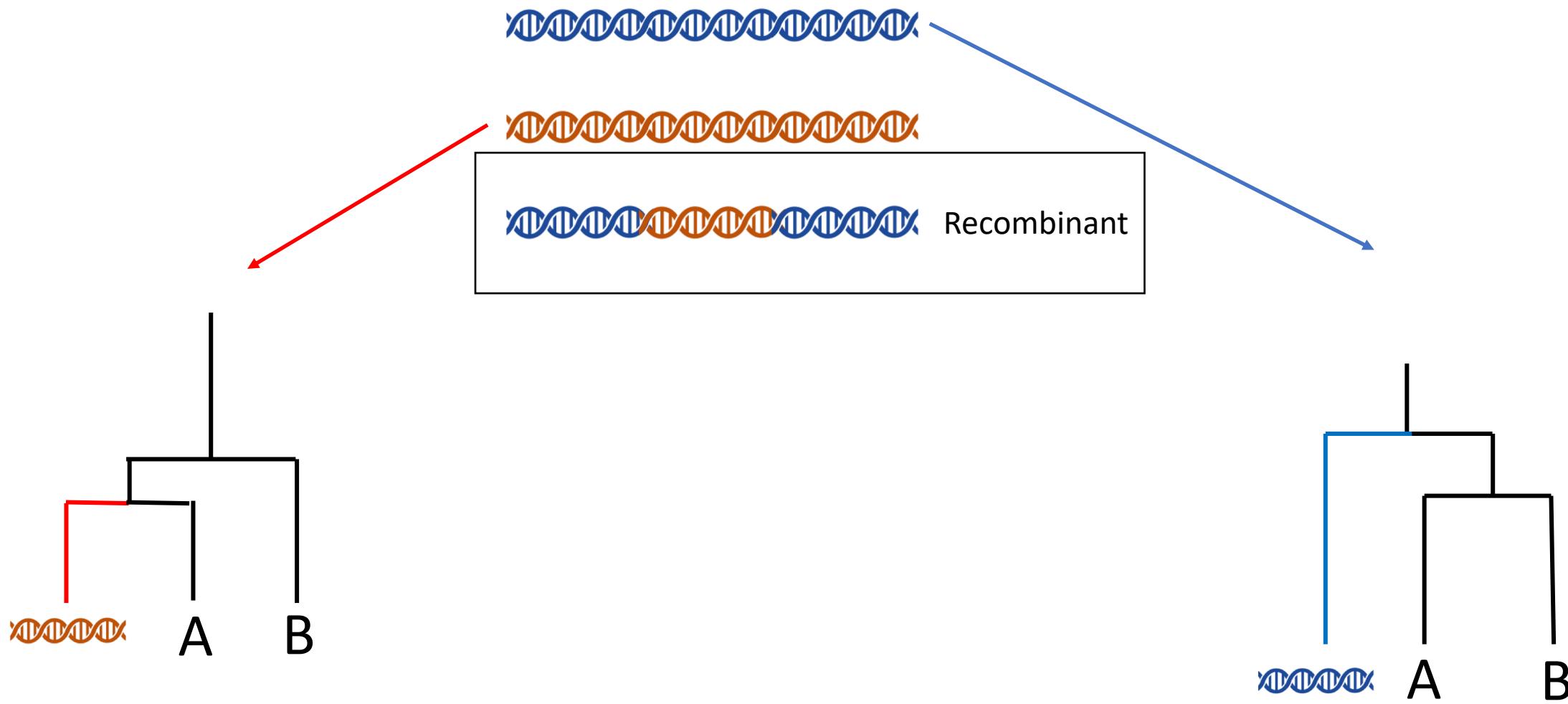
Parent 2



Recombinant

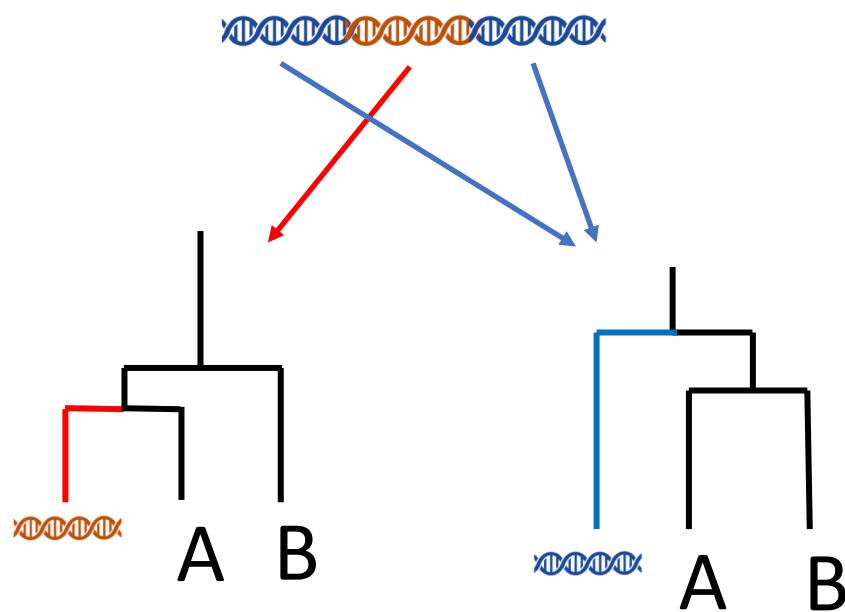
phylogenetic placement of this recombinant genome ?

# Genetic recombination and phylogenetic reconstruction :



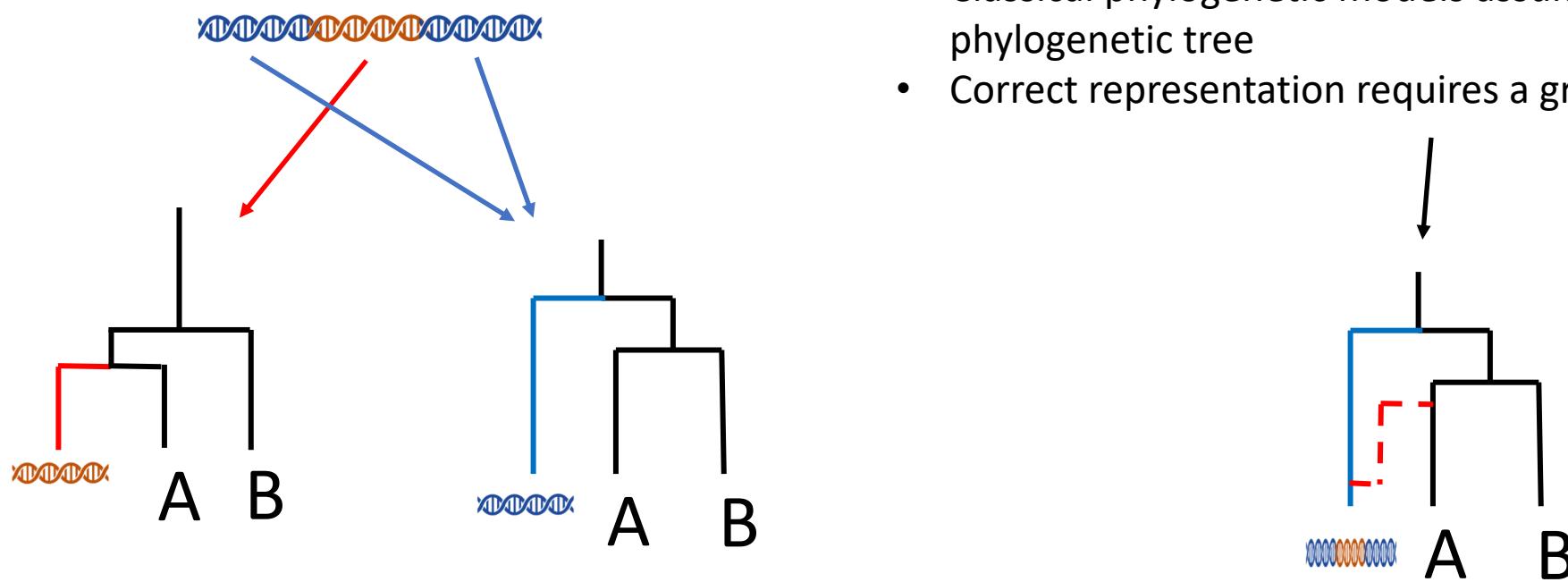
# Genetic recombination and phylogenetic reconstruction :

- In recombinant genomes, different regions have different phylogenetic histories
- Classical phylogenetic models assume a single bifurcating phylogenetic tree



# Genetic recombination and phylogenetic reconstruction :

- In recombinant genomes, different regions have different phylogenetic histories
- Classical phylogenetic models assume a single bifurcating phylogenetic tree
- Correct representation requires a graph structure



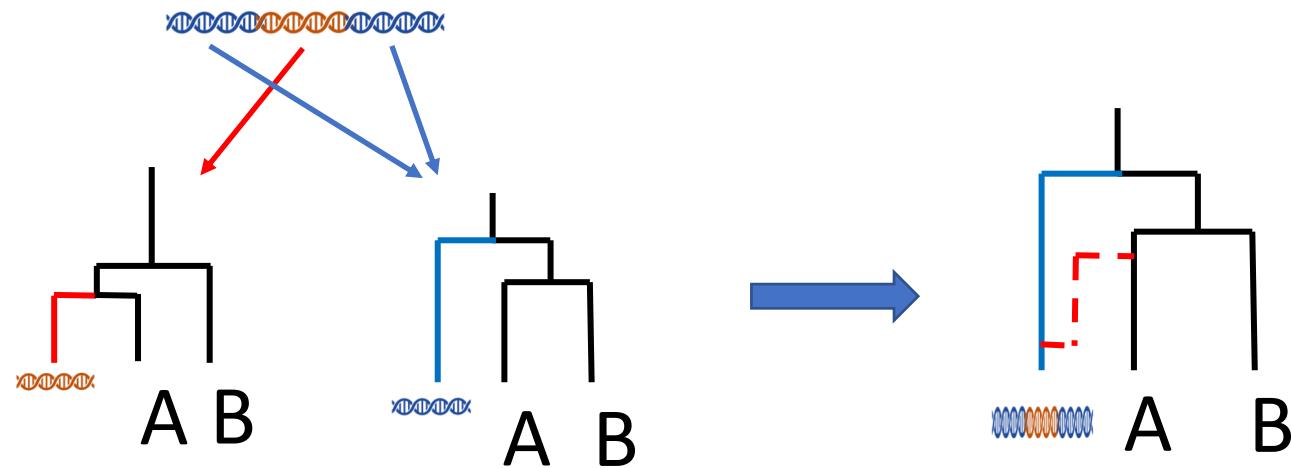
Genetic recombination and phylogenetic reconstruction :

Effect of not accounting for recombination in a phylogenetic analysis ?

# Genetic recombination and phylogenetic reconstruction :

Effect of not accounting for recombination in a phylogenetic analysis ?

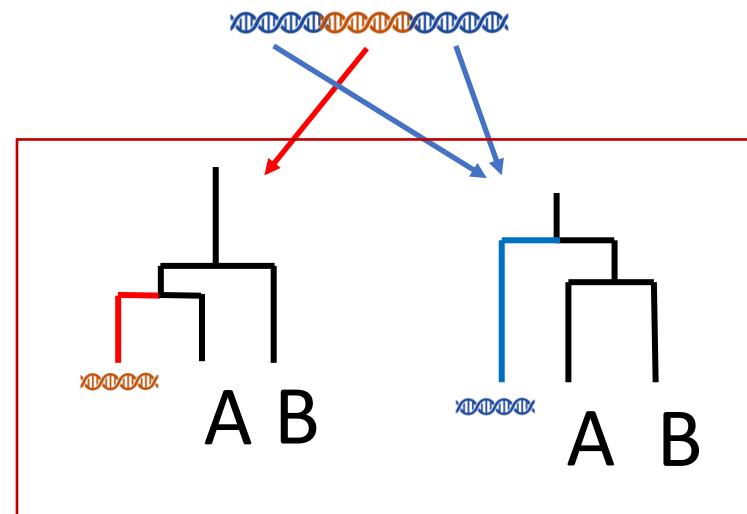
- what is the correct genealogy ? Correct representation requires a graph structure



# Genetic recombination and phylogenetic reconstruction :

Effect of not accounting for recombination in a phylogenetic analysis ?

- what is the correct genealogy ? Correct representation requires a graph structure

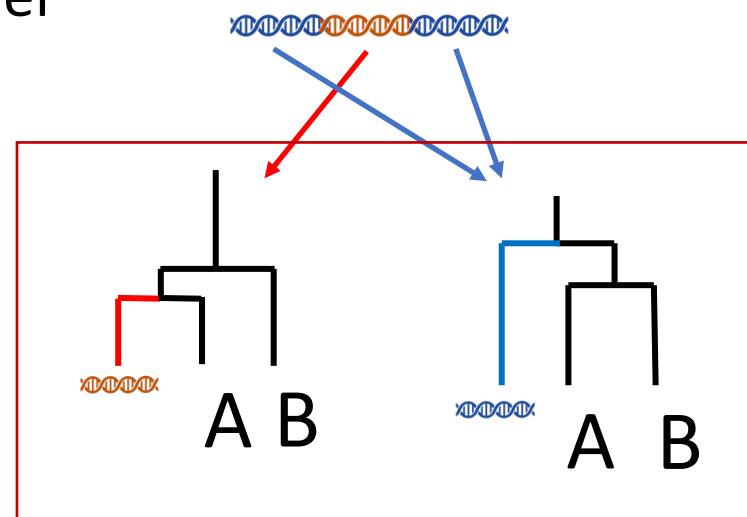


can we at least recover one of these ?

# Genetic recombination and phylogenetic reconstruction :

Effect of not accounting for recombination in a phylogenetic analysis ?

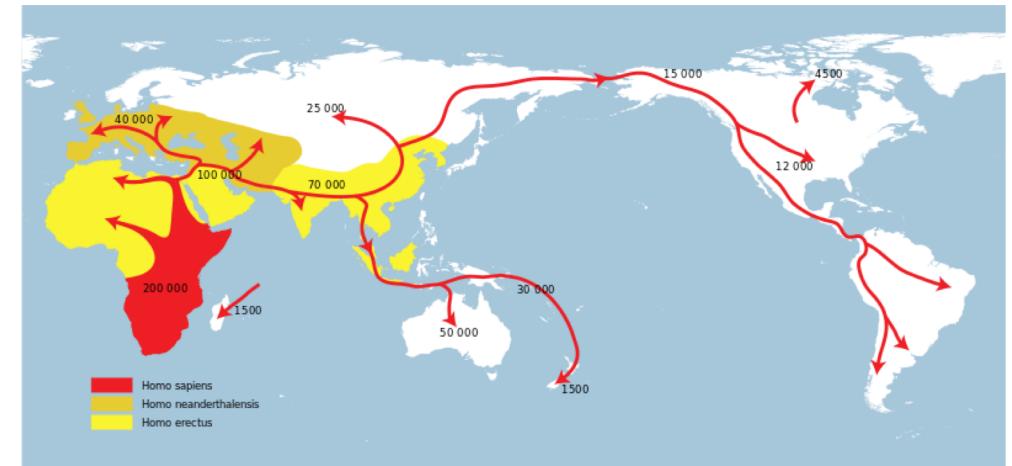
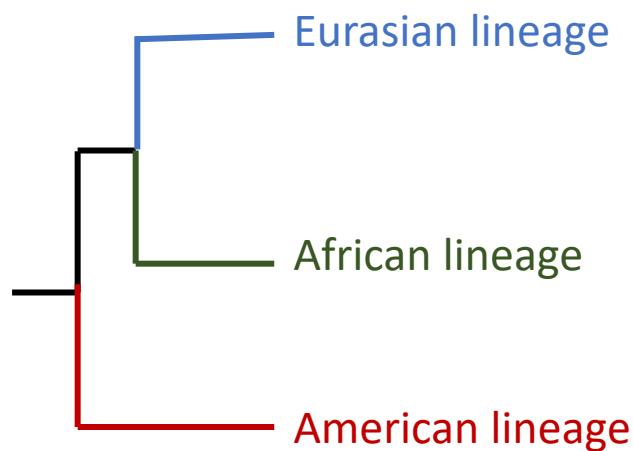
- what is the correct genealogy ? Correct representation requires a graph structure
- biased estimation
- loss of statistical power



# Genetic recombination and phylogenetic reconstruction :

Effect of not accounting for recombination in a phylogenetic analysis ?

- what is the correct genealogy ?
- biased estimation
- loss of statistical power

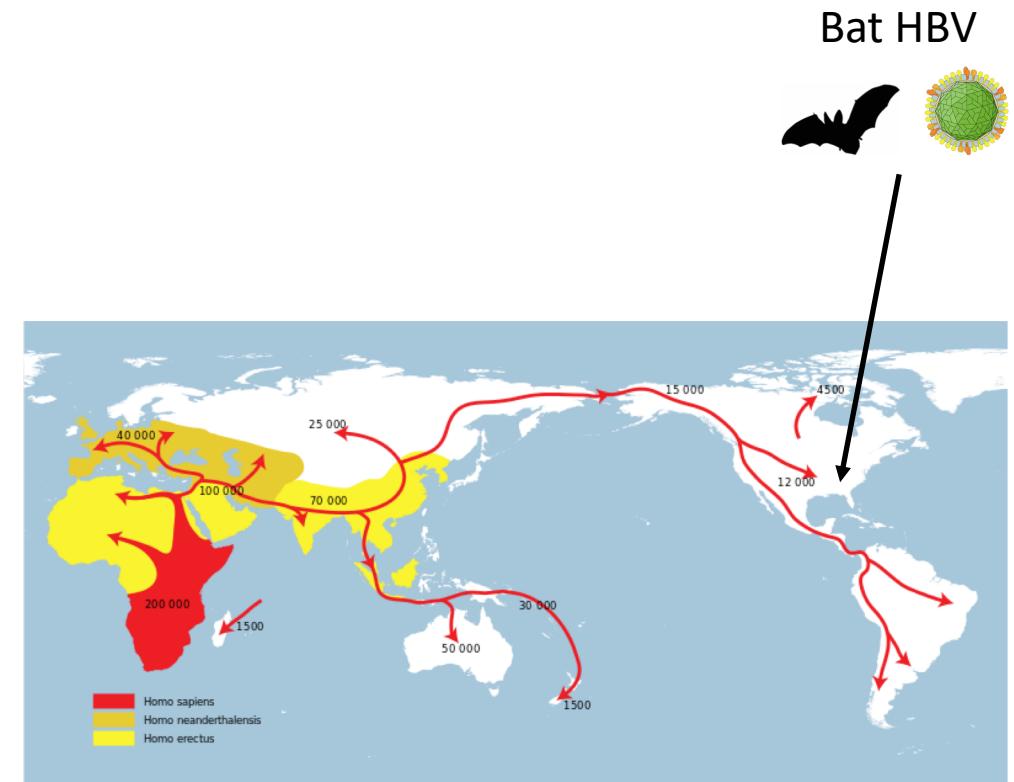
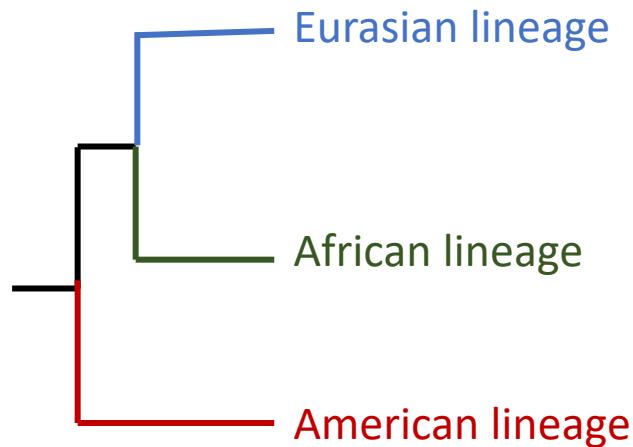


HBV phylogeny (hypothetic)

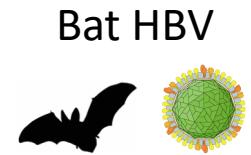
# Genetic recombination and phylogenetic reconstruction :

Effect of not accounting for recombination in a phylogenetic analysis ?

- what is the correct genealogy ?
- biased estimation
- loss of statistical power



HBV phylogeny (hypothetic)

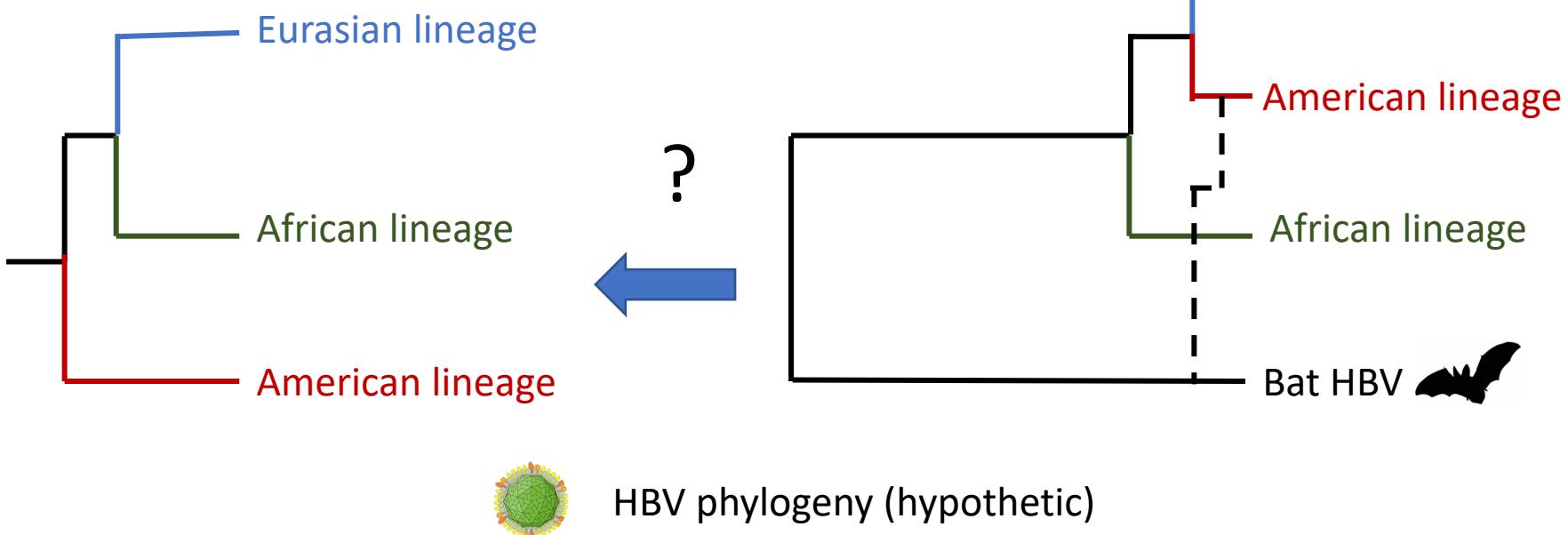


# Genetic recombination and phylogenetic reconstruction :

Effect of not accounting for recombination in a phylogenetic analysis ?

- what is the correct genealogy ?
- biased estimation
- loss of statistical power

What if ?



Genetic recombination and phylogenetic reconstruction :

Effect of not accounting for recombination in a phylogenetic analysis ?

# Genetic recombination and phylogenetic reconstruction :

Effect of not accounting for recombination in a phylogenetic analysis ?

- recombination is modelled as mutation
- inflation of mutation rate/molecular clock
- biased estimation of divergence times

# Genetic recombination and phylogenetic reconstruction :

Effect of not accounting for recombination in a phylogenetic analysis ?

- recombination is modelled as mutation
- inflation of mutation rate/molecular clock
- underestimation of divergence times



Adapted from Mühlmann et al. 2018

# Genetic recombination and phylogenetic reconstruction :

## Estimating genetic recombination

- understanding of genome dynamics
- evaluation of selective pressure
- better understanding of pathogen evolution
- ...

# Accounting for genetic recombination in phylogenetic analyses:

two-step approach:

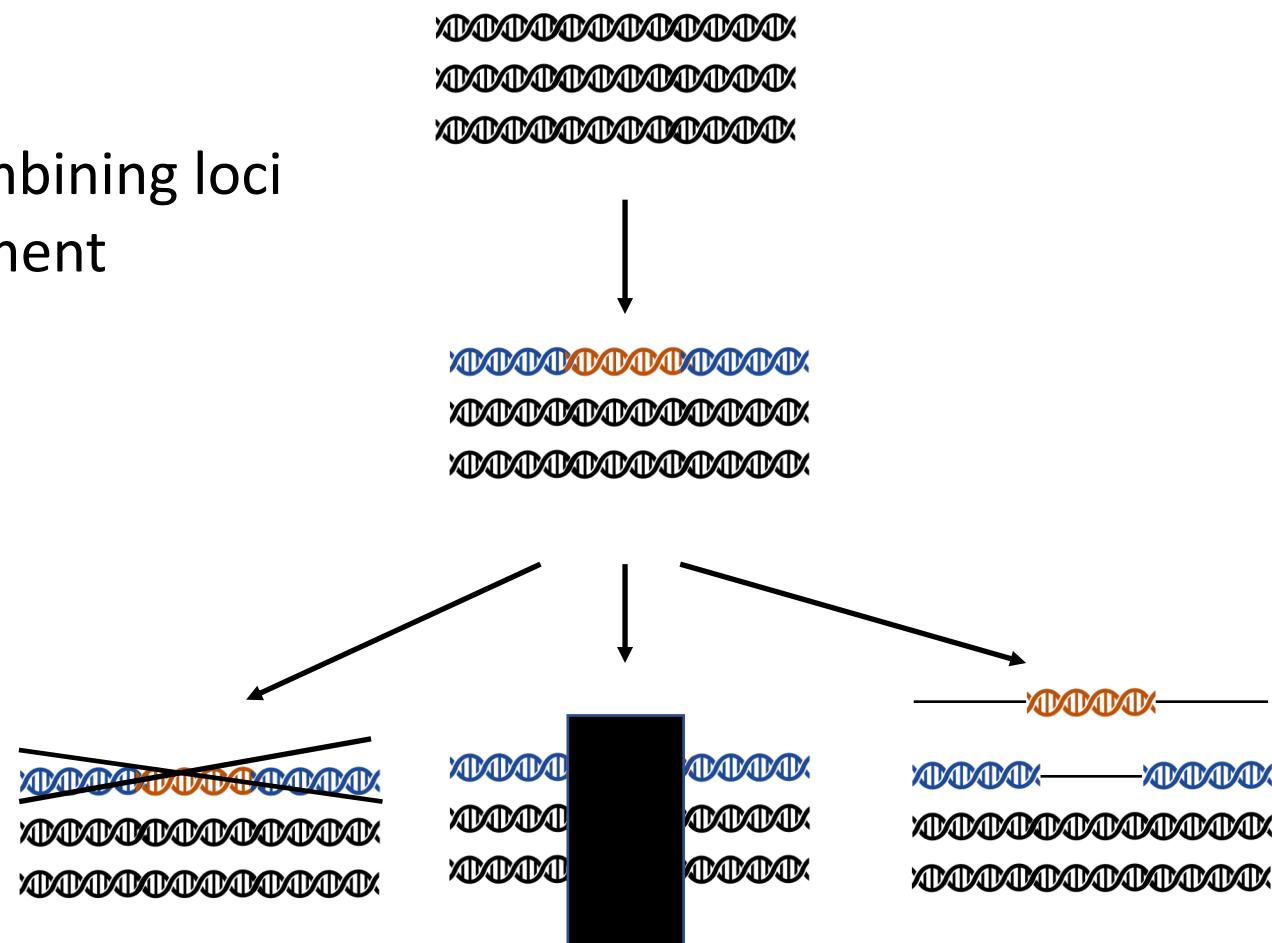
1. Detecting recombinant strains / recombining loci



# Accounting for genetic recombination in phylogenetic analyses:

two-step approach:

1. Detecting recombinant strains / recombining loci
2. produce an recombination-free alignment



# Detecting recombination



Parent 1

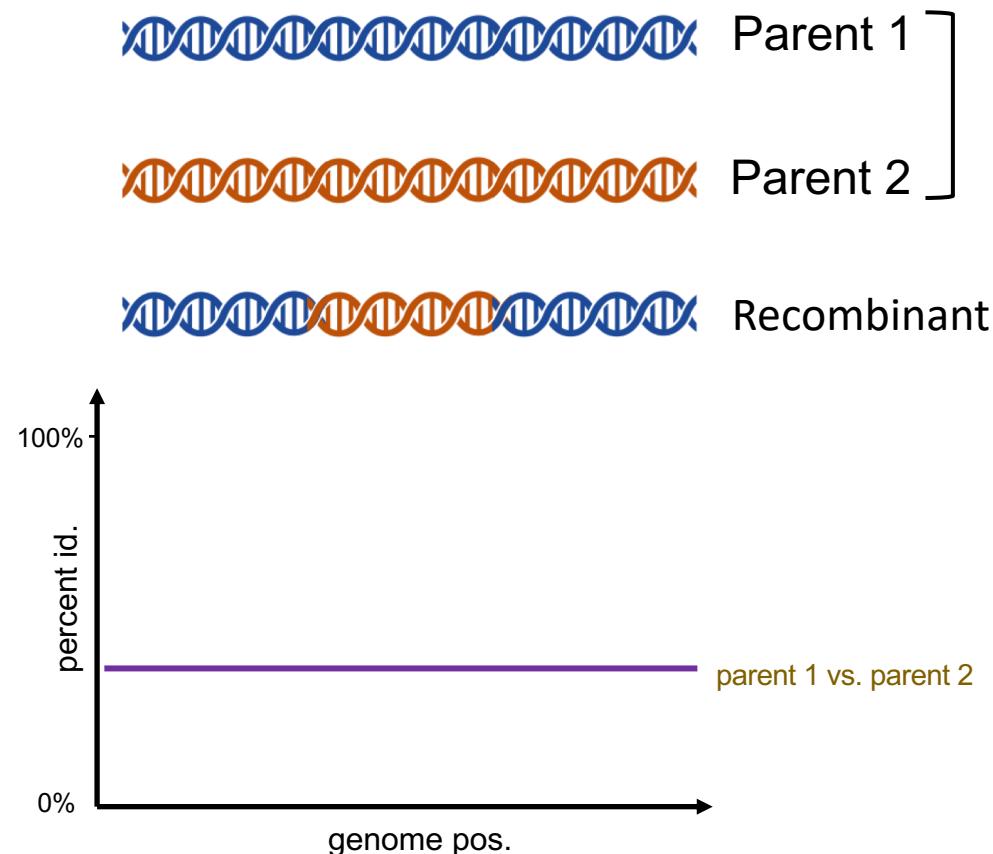


Parent 2

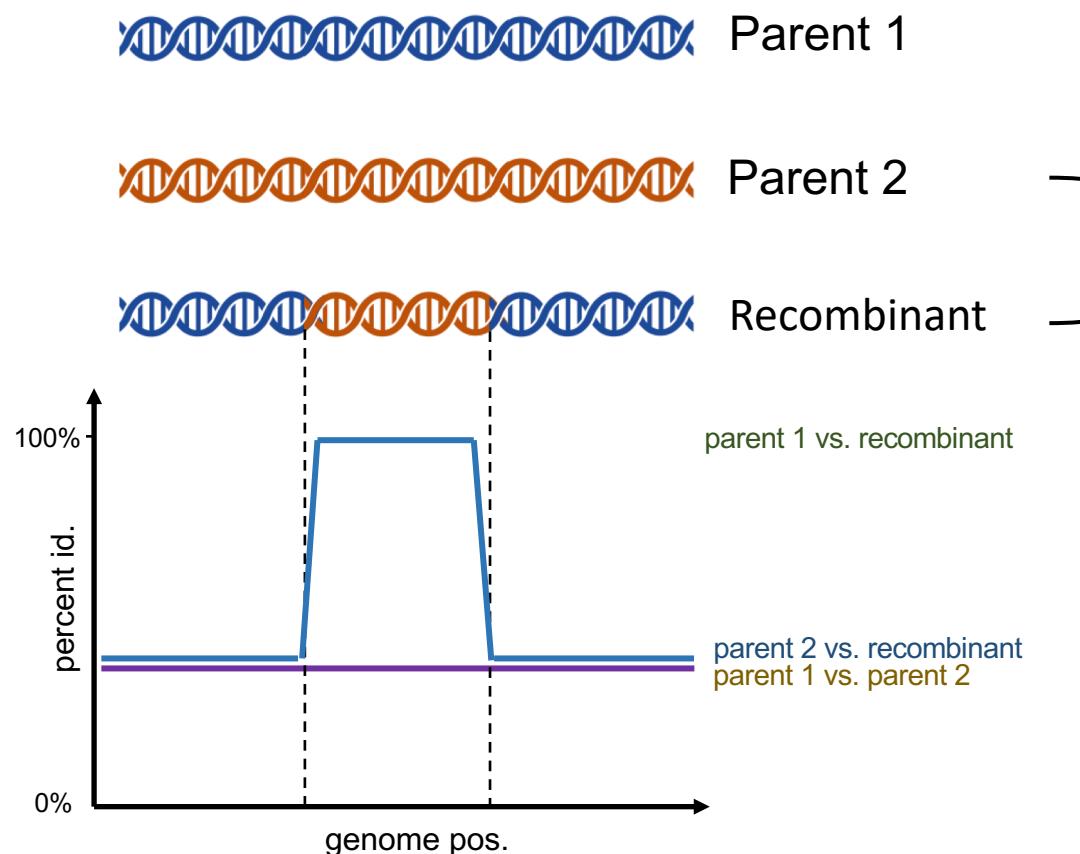


Recombinant

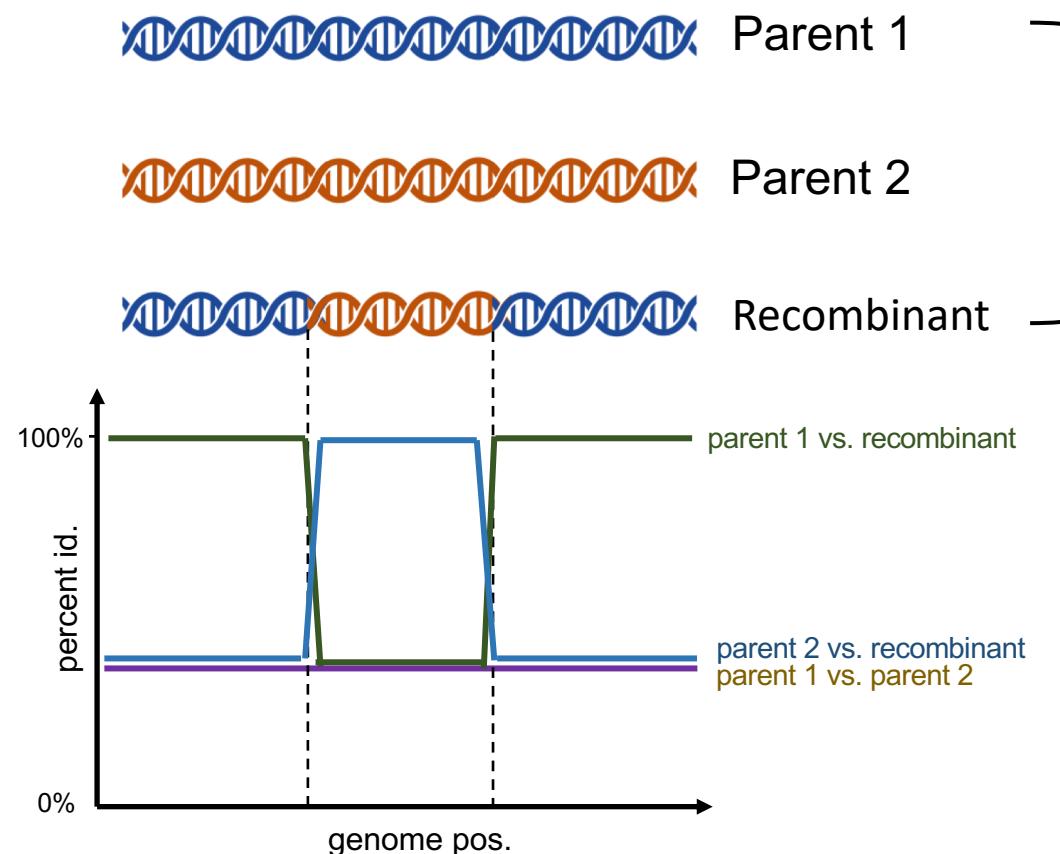
# Detecting recombination



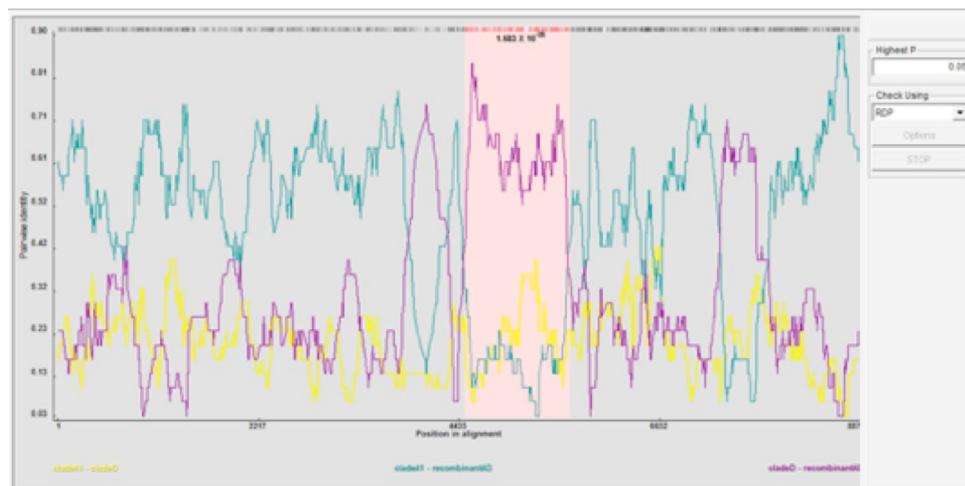
# Detecting recombination



# Detecting recombination



# Detecting recombination

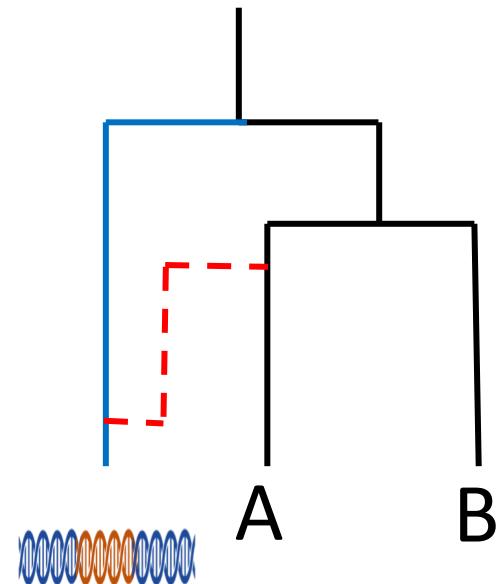


RDP4 (Martin *et al* 2014)

Accounting for genetic recombination in phylogenetic analyses:

1-step approach: recombination-aware phylogenetic models:

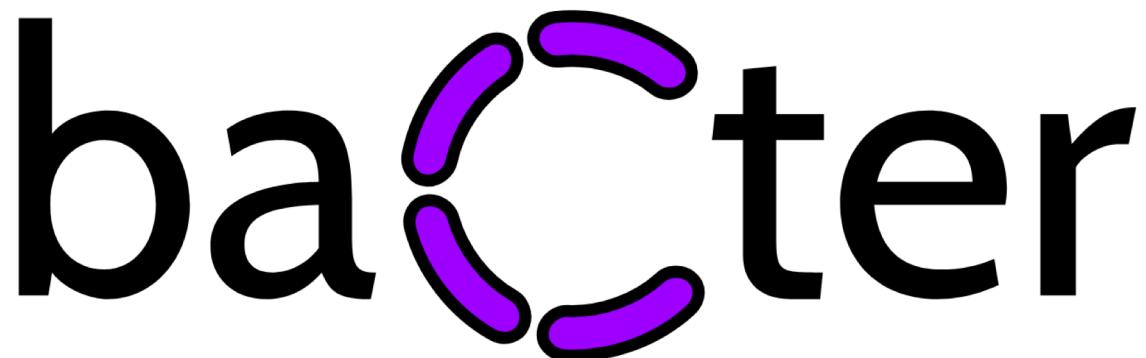
- SPLITSTREES
- SMARTIE
- CLONAL FRAME
- BACTER



Accounting for genetic recombination in phylogenetic analyses:

1-step approach: recombination-aware phylogenetic models:

- SPLITSTREES
- SMARTIE
- CLONAL FRAME
- BACTER



Bacterial phylogenetics in BEAST 2

(BEAST2 package; Vaughan *et al.* 2017)

# Bayesian inference with BEAST

$$f(\theta|D) = \frac{f(\theta)Pr(D|\theta)}{Pr(D)}$$

$f(\theta|D)$ : Posterior distribution

$f(\theta)$ : Prior distribution

$Pr(D|\theta)$ : Likelihood

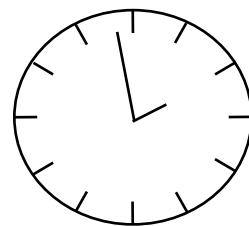
$Pr(D)$ : Marginal likelihood (constant with respect to  $\theta$ )

$D$ : data,  $\theta$ : set of parameters

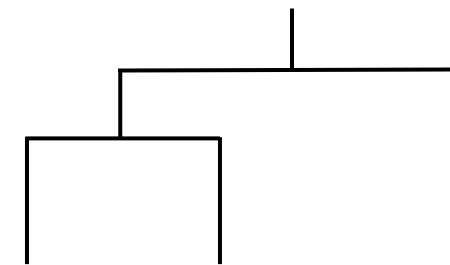
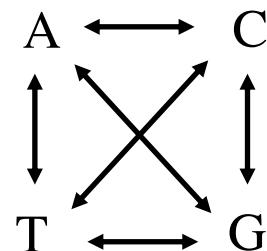
# Bayesian inference with BEAST

$$f(\theta|D) \propto f(\theta) Pr(D|\theta)$$

Prior

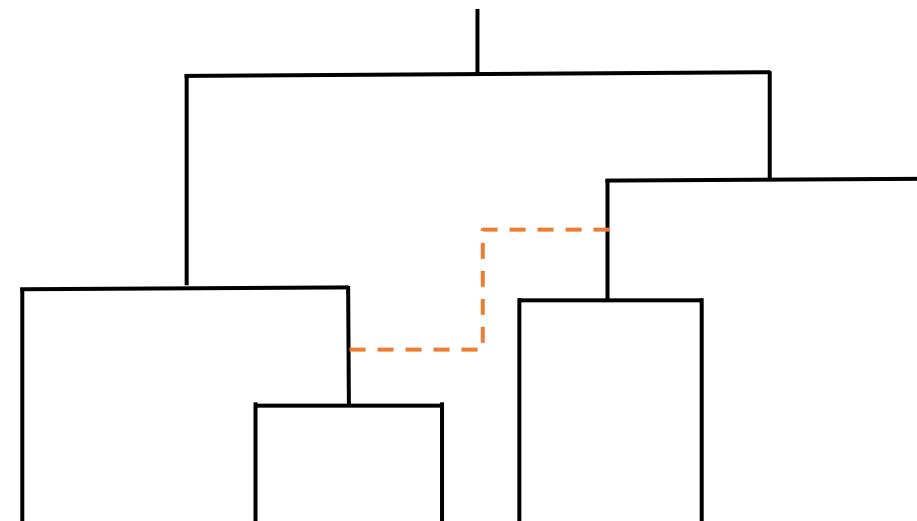


Likelihood

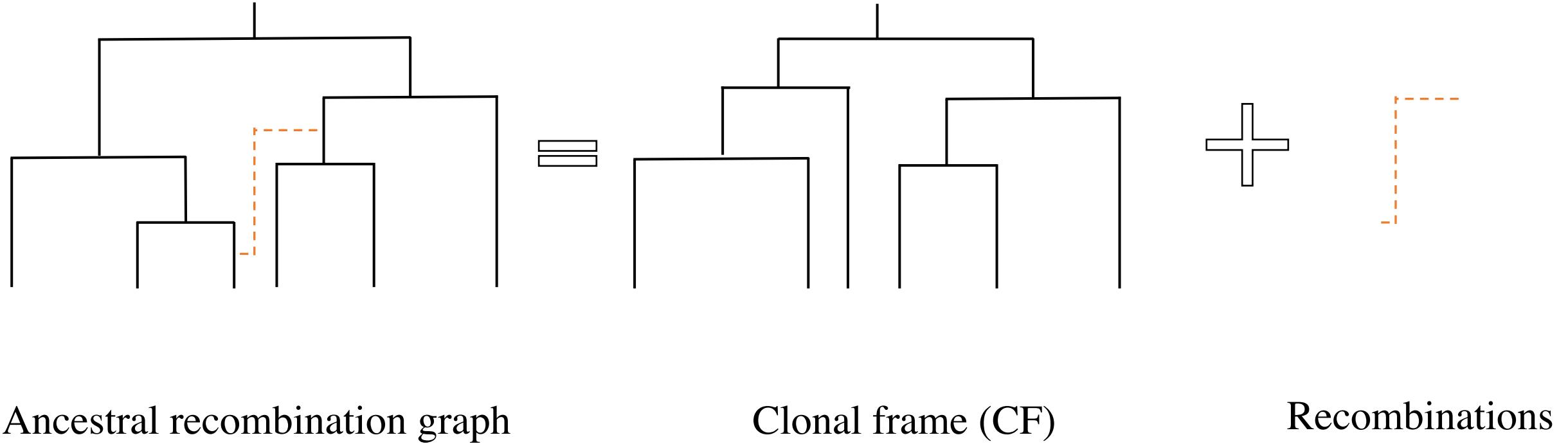


# Recombination analysis with bacter

- Aim:  
**Ancestral recombination graph** (ARG) integrating recombination events into the phylogeny



# Recombination analysis with bacter

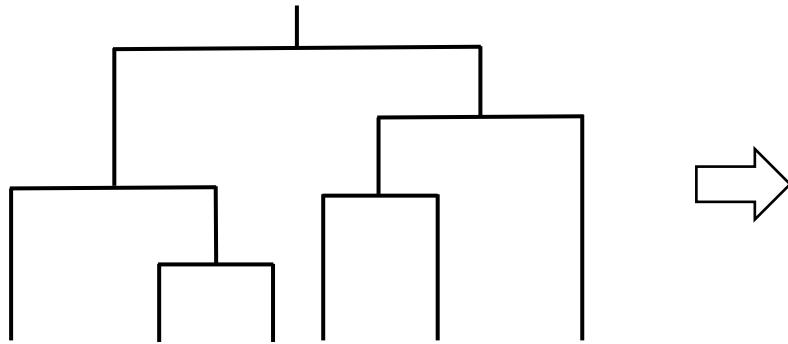


Ancestral recombination graph

Clonal frame (CF)

Recombinations

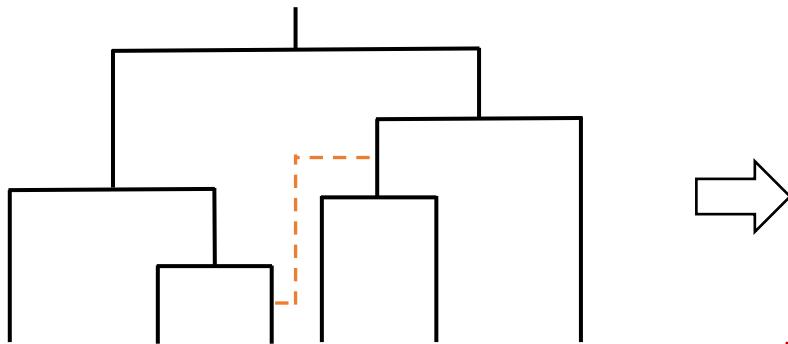
# Recombination analysis with bacter



$$f(\theta|D) \propto f(\theta) Pr(D|\theta)$$

=

$$f(T, M, \theta|D) \propto Pr(D|T, M) f(T|\theta) f(M, \theta)$$



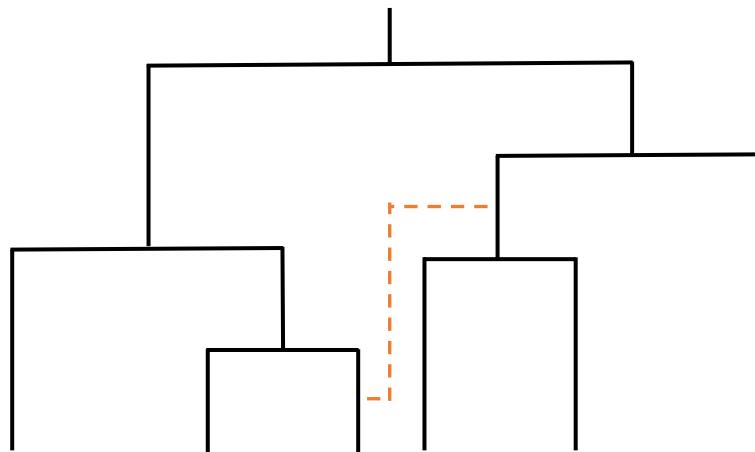
$$f(\theta_{\text{bact}}|D) \propto f(\theta_{\text{bact}}) Pr_{\text{ACG}}(D|\theta_{\text{bact}})$$

=

$$f(G, M, \theta_{\text{bact}}|D) \propto Pr_{\text{ACG}}(D|G, M) f(G|\theta_{\text{bact}}) f(M, \theta_{\text{bact}})$$

# Recombination analysis with bacter

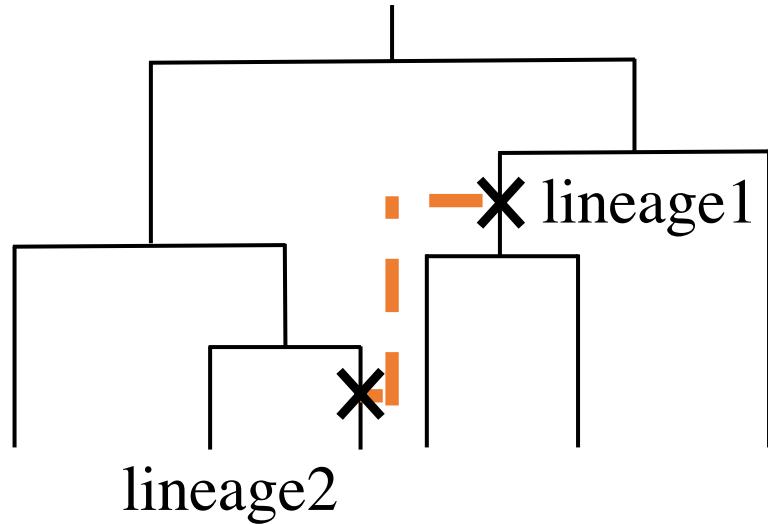
## Prior Distribution



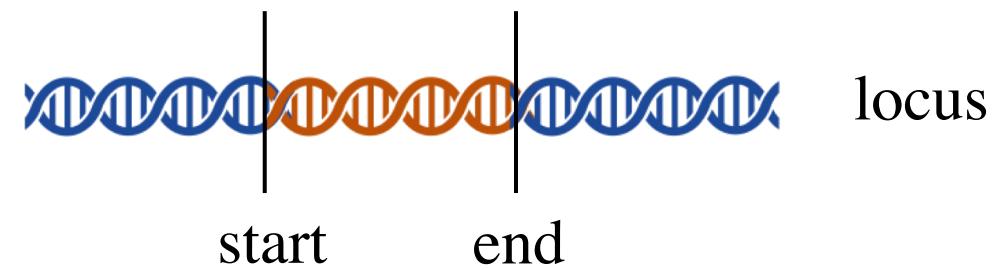
Prior for ACG

# Recombination analysis with bacter

## Tree Prior



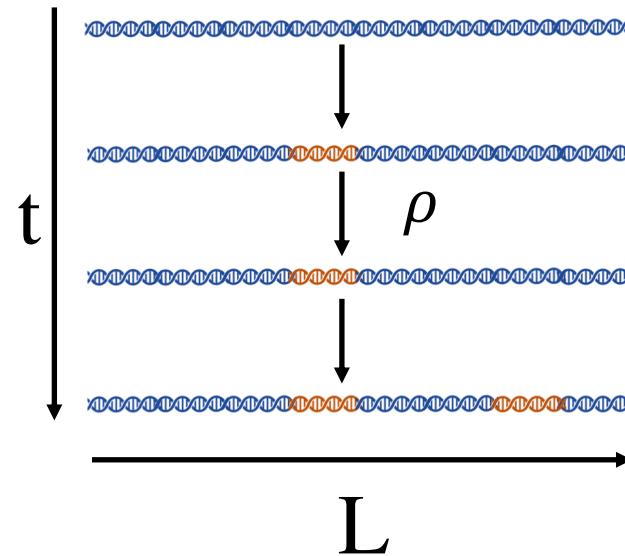
Attachment on CF  
Coalescent with gene conversion



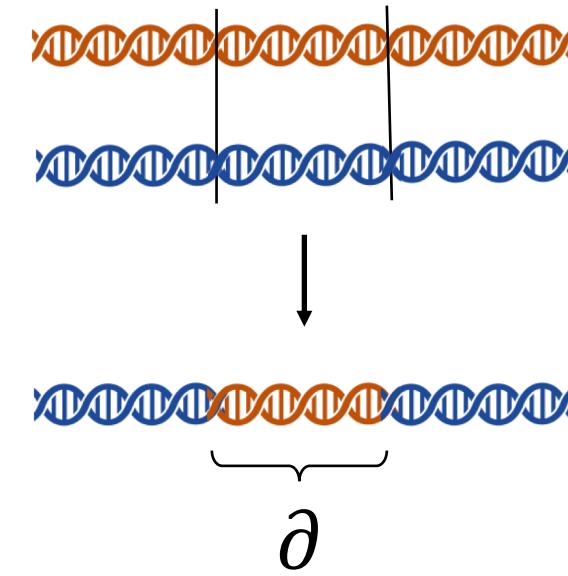
Position on genome  
Uniform distribution

# Recombination analysis with bacter

## Tree Prior



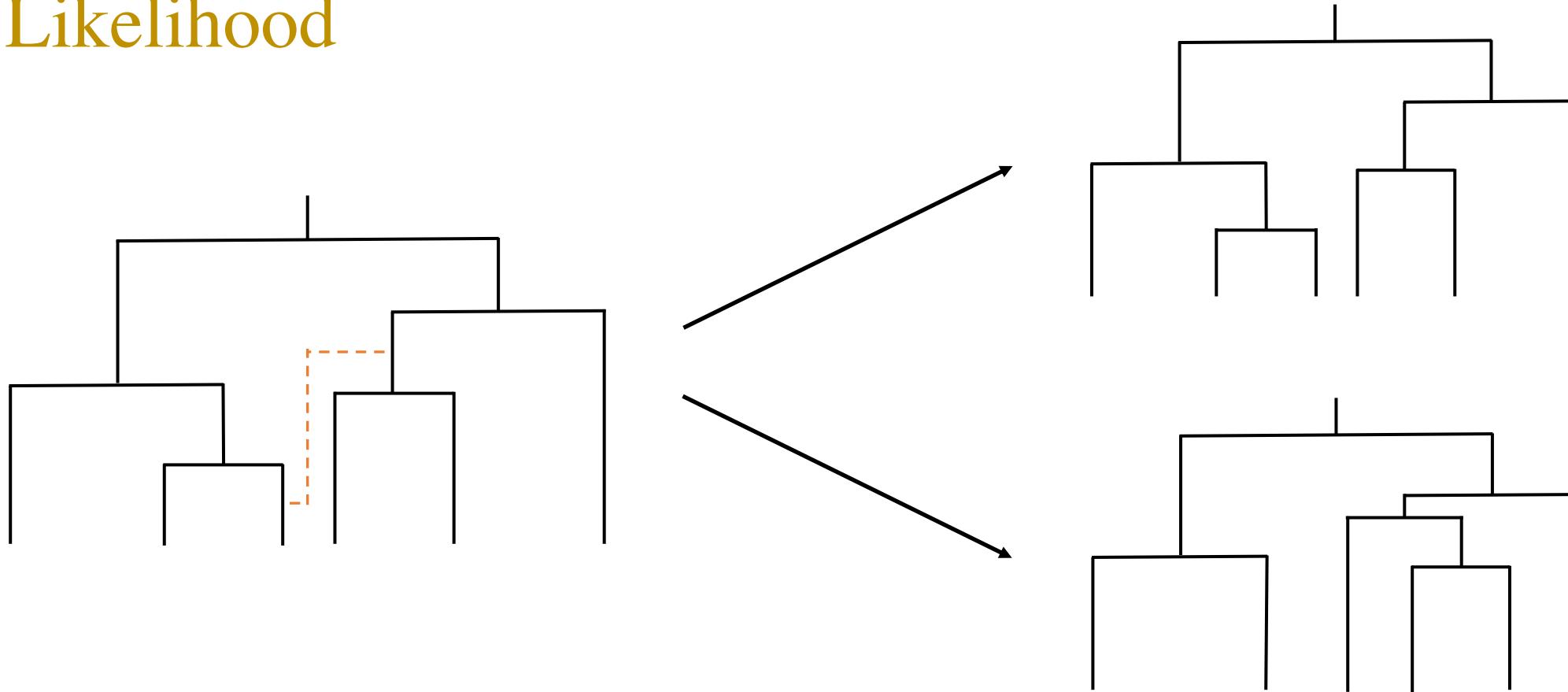
$\rho$ : recombination rate per site per unit time



$\delta$ : average recombining tract length

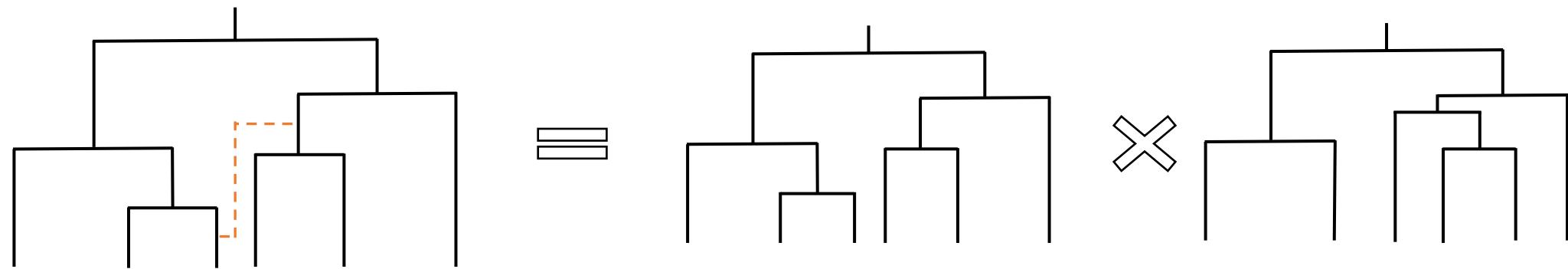
# Recombination analysis with bacter

Likelihood



# Recombination analysis with bacter

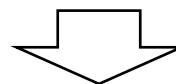
## Likelihood



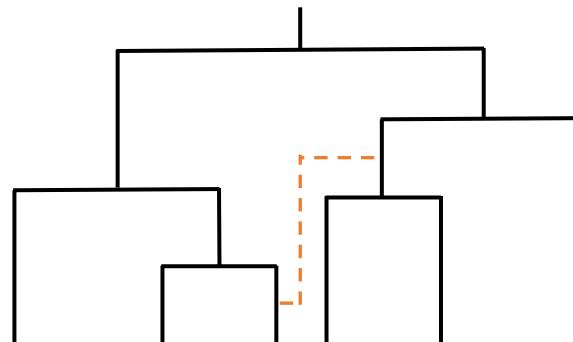
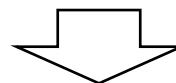
$$Pr_{ACG}(D|\theta_{\text{bact}}) = \prod_i Pr(D_i|\theta)$$

# Recombination analysis with bacter

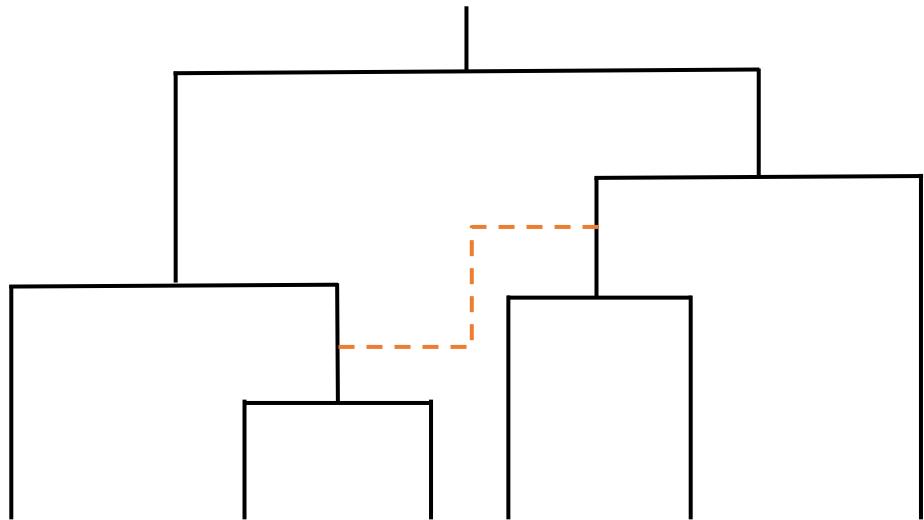
$$f(\theta_{\text{bact}}) \Pr_{\text{ACG}}(D | \theta_{\text{bact}})$$



$$f(\theta_{\text{bact}} | D)$$



# Questions?



ba<sup>C</sup>ter

Bacterial phylogenetics in BEAST 2

# Time to practice

Taming the BEAST



[news](#) [workshops](#) [tutorials](#) [contribute](#)

## Bacter Tutorial

Inferring ARGs from bacterial sequence data.  
by Tim Vaughan