

# Genomic Relationships and Speciation Times of Human, Chimpanzee and Gorilla Inferred from a Coalescent Hidden Markov Model

Asger Hobolth, Ole F. Christensen, Thomas Mailund, Mikkel H. Schierup.

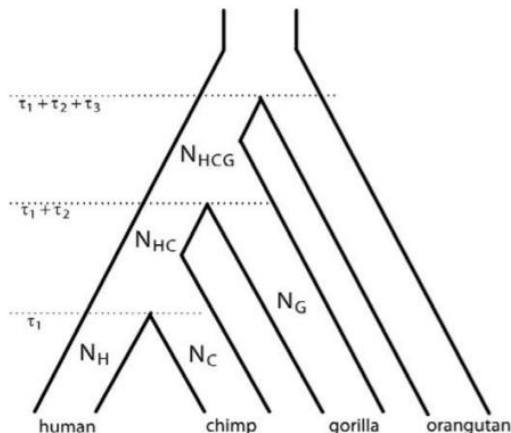
Fiona Grimson

16 April 2013

# INTRODUCTION

## Objective

Comparative analysis of Human, Chimpanzee, Gorilla and Orangutan genomes



# INTRODUCTION

Observed: Alignment of modern-day genome sequences

H	aactggtacg	tggtcg	tacgaactgg
C	aaccggttcg	cggtcg	ttcgaaccgg
G	aacctgttcg	ctgtcg	ttcgaacctg
O	atctggtacg	tggtcg	tacgatctgg

Not Observed: Evolutionary history/gene genealogy at each locus

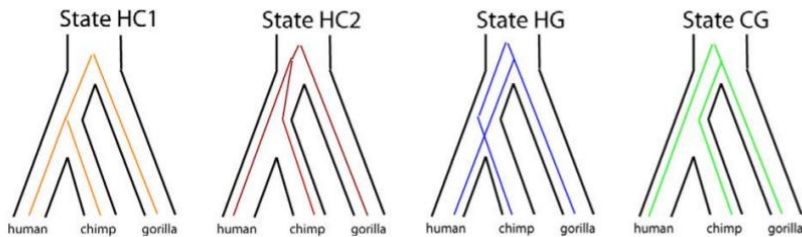
# INTRODUCTION

**Observed:** Alignment of modern-day genome sequences

H	aactggtacg	tggtcg	tacgaactgg
C	aacggttcg	cggtcg	ttogaaccgg
G	aacctgttcg	ctgtcg	ttogaacctg
O	atctggtacg	tggtcg	tacgatctgg

**Not Observed:** Evolutionary history/gene genealogy at each locus

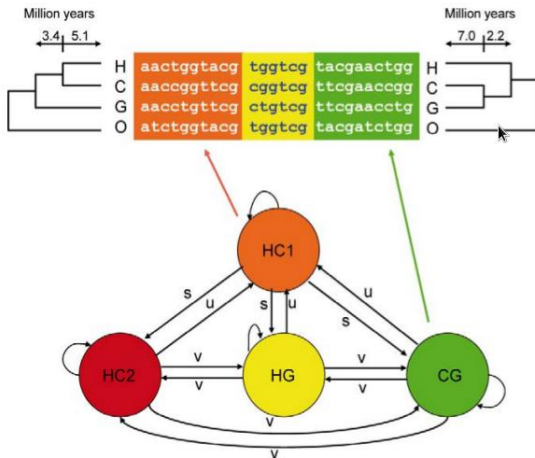
**Complication:**



- Many genealogies fit within the species tree
- Genealogy changes by locus - recombination

# COALESCENT HIDDEN MARKOV MODEL

- Hidden Markov Model
- Geanealogy is the hidden state at each locus.



# WHAT'S NEW?

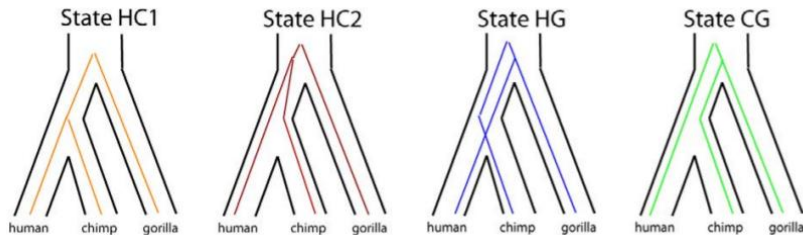
At the time of publication (2007):

- Information from fossil record, analysis of short chunks of DNA.
- Recent availability of LONG sequences
- Need to model dependence between loci

# HIDDEN STATES

Transition Probabilities between hidden states:

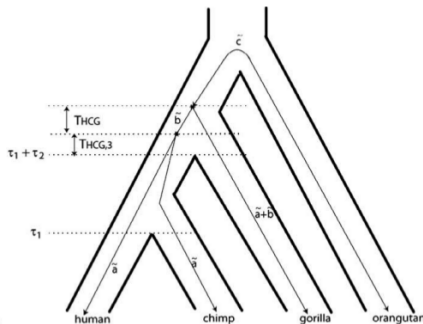
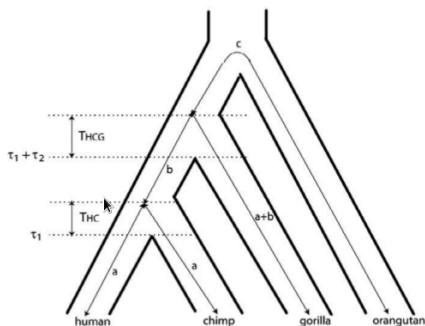
$$P = \begin{bmatrix} . & s & s & s \\ u & . & v & v \\ u & v & . & v \\ u & v & v & . \end{bmatrix}$$



# EMISSIONS

The gene genealogy is also known as the **coalescent**:

- Coalescents are CTMCs
- Model molecular substitutions as events in continuous evolutionary time. Different species “coalesce” into common ancestor.
- Restrictions on CTMC process from species tree and which of the four states.





# EMISSIONS

Emission Probabilities from hidden states depend on rate matrix  $Q$ , branch lengths and hidden state.

$$Q = \begin{bmatrix} . & \pi_G & \beta\pi_C & \gamma\pi_T \\ \pi_A & . & \delta\pi_C & \beta\pi_T \\ \beta\pi_A & \delta\pi_G & . & \pi_T \\ \gamma\pi_A & \beta\pi_G & \pi_C & . \end{bmatrix}$$

Assume  $Q$  has a reversible, strand symmetric form.

# EMISSIONS

Emission Probabilities from hidden states depend on rate matrix  $Q$ , branch lengths and hidden state.

$$Q = \begin{bmatrix} . & \pi_G & \beta\pi_C & \gamma\pi_T \\ \pi_A & . & \delta\pi_C & \beta\pi_T \\ \beta\pi_A & \delta\pi_G & . & \pi_T \\ \gamma\pi_A & \beta\pi_G & \pi_C & . \end{bmatrix}$$

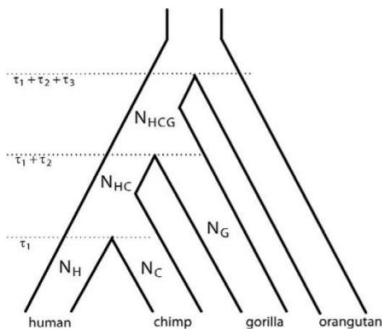
Assume  $Q$  has a reversible, strand symmetric form.

- The branch lengths are the free parameters in the HMM.
- $Q$  parameters are set by calibrating with branch lengths

## SOLVE THE HMM

Transitions + Emissions  $\rightarrow$  **likelihood** of alignment.

- Baum-Welch Algorithm  $\rightarrow$  MLEs
- Fisher Information (numerical derivative)  $\rightarrow$  SEs
- Speciation times and ancestral population sizes functions of these parameters.



Thank You!