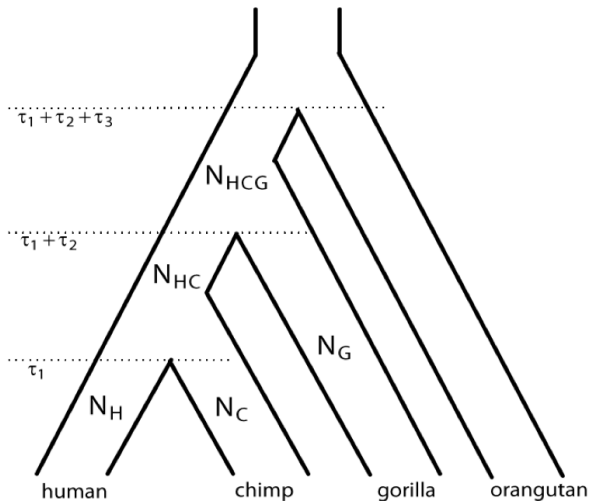


# Genomic Relationships and Speciation Times of the Great Apes from a Coal-HMM

Asger Hobolth, et al.

November 23, 2008

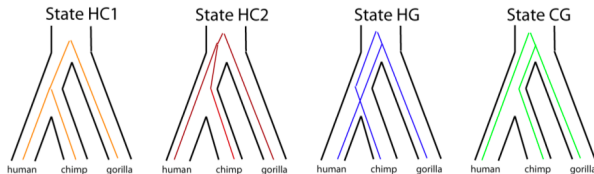
# An Evolutionary Relationship of Them



# Questions

1. The conflict between species and gene genealogy
2. Speciation times
3. Properties of ancestral states
4. How speciation occurred

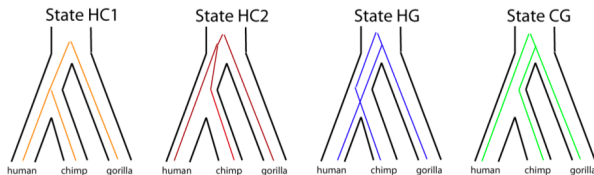
## Various genealogies



# Questions

1. The conflict between species and gene genealogy
2. Speciation times
3. Properties of ancestral states
4. How speciation occurred

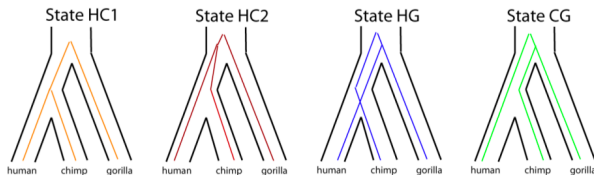
## Various genealogies



# Questions

1. The conflict between species and gene genealogy
2. Speciation times
3. Properties of ancestral states
4. How speciation occurred

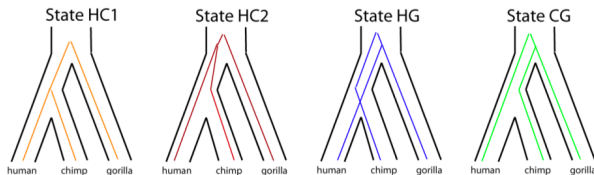
## Various genealogies



# Questions

1. The conflict between species and gene genealogy
2. Speciation times
3. Properties of ancestral states
4. How speciation occurred

## Various genealogies

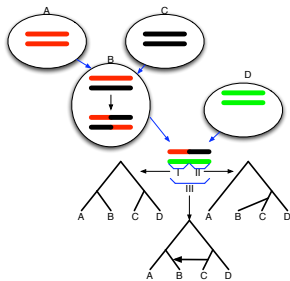


# Settings

## Data

1. Whole genome sequences of the four taxa available
2. They are so closely related that alignments of large contiguous parts of the genomes can be constructed
3. The analysis is challenging because different parts of the alignment will have different evolutionary histories

## Recombination

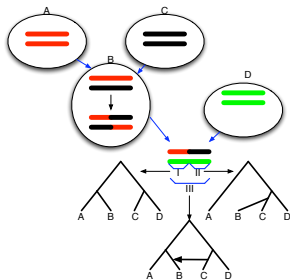


# Settings

## Data

1. Whole genome sequences of the four taxa available
2. They are so closely related that alignments of large contiguous parts of the genomes can be constructed
3. The analysis is challenging because different parts of the alignment will have different evolutionary histories

## Recombination

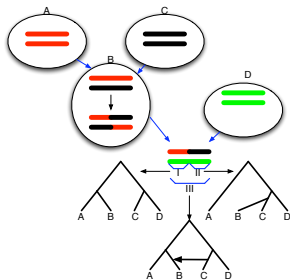


# Settings

## Data

1. Whole genome sequences of the four taxa available
2. They are so closely related that alignments of large contiguous parts of the genomes can be constructed
3. The analysis is challenging because different parts of the alignment will have different evolutionary histories

## Recombination



# Approaches

## Research Problem

How would the most probable phylogenetic look like incorporating different genealogies from recombinations?

## Approaches

1. A natural one: Moving along the alignment, simultaneously compute the probabilities of properties of interest
2. The spatial information along the alignment ignored
3. HMM:

# Approaches

## Research Problem

How would the most probable phylogenetic look like incorporating different genealogies from recombinations?

## Approaches

1. A natural one: Moving along the alignment, simultaneously compute the probabilities of properties of interest
2. The spatial information along the alignment ignored
3. HMM:

# Approaches

## Research Problem

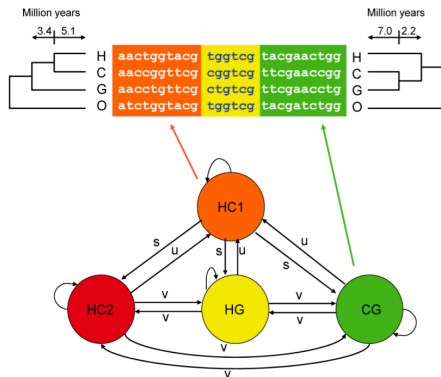
How would the most probable phylogenetic look like incorporating different genealogies from recombinations?

## Approaches

1. A natural one: Moving along the alignment, simultaneously compute the probabilities of properties of interest
2. The spatial information along the alignment ignored
3. HMM:

# coal-HMM

## Input Data



## States and transitions of the model

# Value of interest 1.

## Transition Probabilities

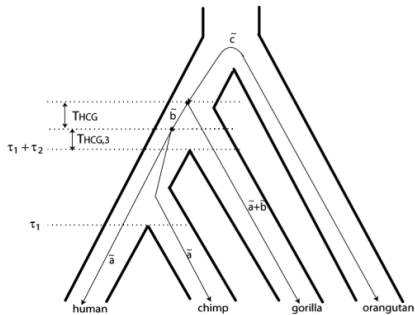
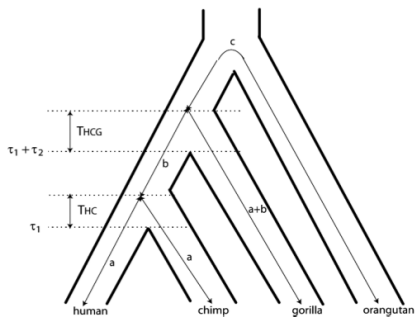
	HC1	HC2	HG	CG
HC1	$1 - 3s$	$s$	$s$	$s$
HC2	$u$	$1 - (u + 2v)$	$v$	$v$
HG	$u$	$v$	$1 - (u + 2v)$	$v$
CG	$u$	$v$	$v$	$1 - (u + 2v)$

## Emission Probabilities

1. Substitution rate matrix  $Q$
2. The phylogenetic tree corresponding to the hidden state
3. What does it mean to know the phylogenetic tree?

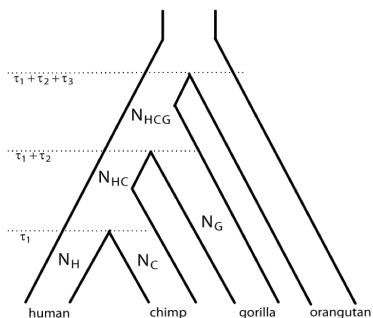
# Value of interest 2.

## Divergence Times



# Value of interest 3.

## Speciation Times



- ▶  $\tau_2 = (b - 2N_{HCG}\mu)/(1/\log(1-\psi) + 1/\psi)$ ,  
where  $\psi = 1 - \exp(-\tau_2/2N_{HC}\mu)$
- ▶  $\tau_1 = a + b - (\tau_2 + 2N_{HCG}\mu)$
- ▶  $\eta = (s, u, v, a, b, c, \tilde{a})$

# Experiment

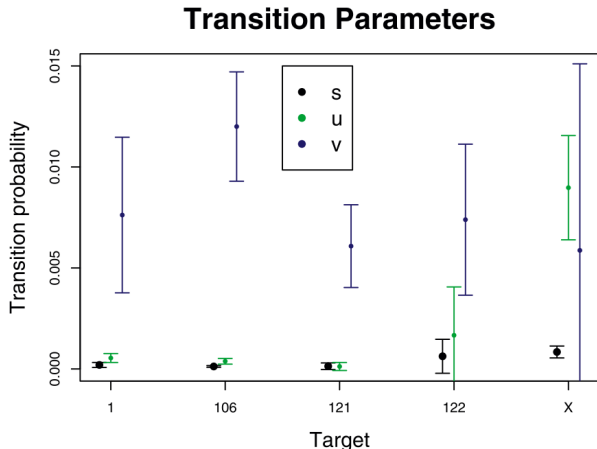
## Dataset

Five long contiguous human-chimp-gorilla-orangutan(HCGO) alignments from the NIH(Targets 1, 106, 121, and 122 on four different autosomes and Target 46 on the X chromosome)

## Probabilities estimation in HMM

Run Baum-Welch!

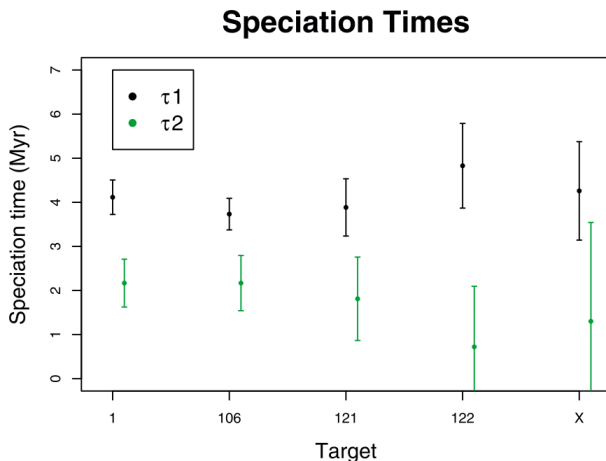
## Result 1.



## Interpretation

- ▶ Together with other results, HC1 is strongly supported and usually is being expanded kilo-basepairs, while alternatives are not.

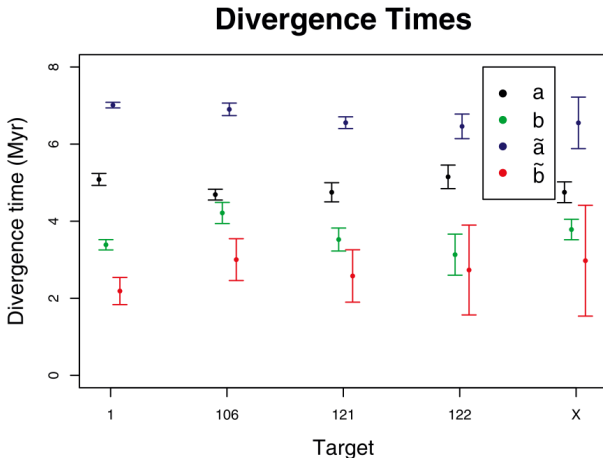
## Result 2.



### Interpretation

- ▶ This measure is consistently around 4Myrs even at X chromosome.

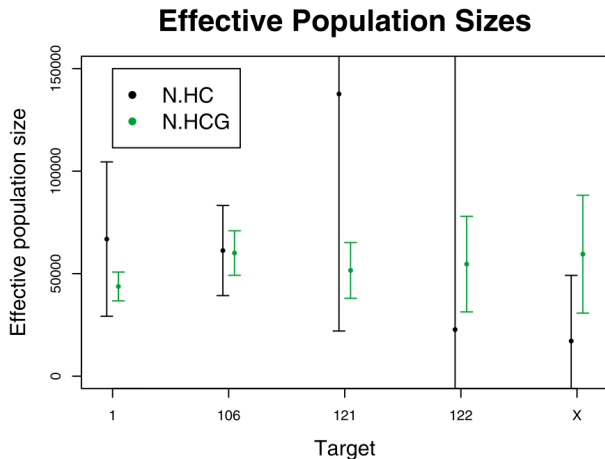
## Result 3.



### Interpretation

- ▶ The divergence times are much higher than speciation times
- ▶ It's because of large effective sizes in the ancestral species

## Result 4.



### Interpretation

- ▶ It looks counterintuitive, but it matches with the result of simulations

# Future Work

1. Handling much more complex coalescent-with-recombination process when more data become available
2. Introducing more hidden states in the HMM to provide a more detailed approximation of the different coalescent times.

Thank you and Questions?