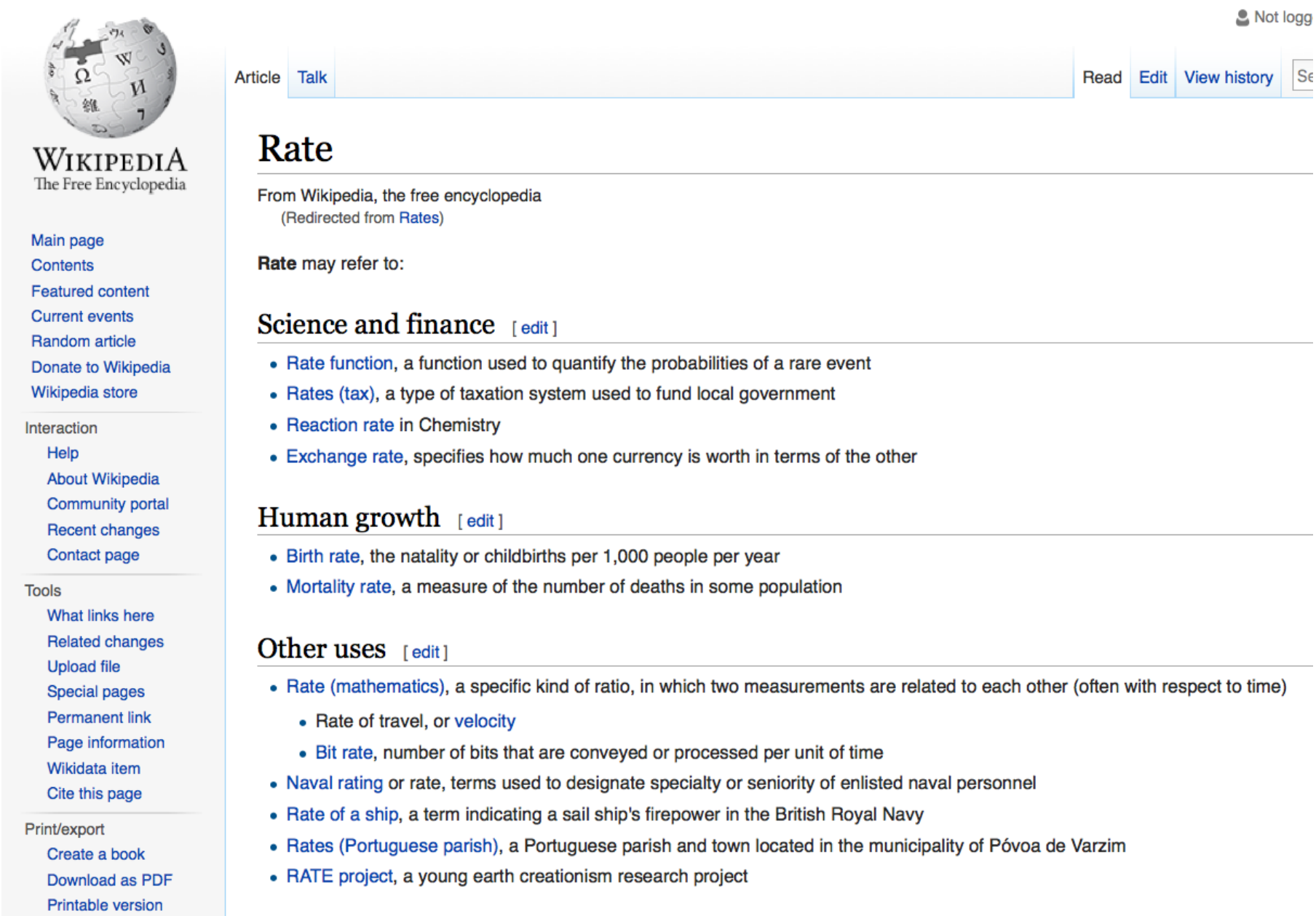
Which rate are you on?

Remco R. Bouckaert
remco@cs.auckland.ac.nz

Centre of Computational Evolution, University of Auckland
Max Planck Institute for the Science of Human History
Computer Science Department, Waikato University

Jena, 2106

What is a "rate"



Not logg

Article Talk Read Edit View history Se

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export
Create a book
Download as PDF
Printable version

Rate

From Wikipedia, the free encyclopedia
(Redirected from Rates)

Rate may refer to:

Science and finance [edit]

- Rate function, a function used to quantify the probabilities of a rare event
- Rates (tax), a type of taxation system used to fund local government
- Reaction rate in Chemistry
- Exchange rate, specifies how much one currency is worth in terms of the other

Human growth [edit]

- Birth rate, the natality or childbirths per 1,000 people per year
- Mortality rate, a measure of the number of deaths in some population

Other uses [edit]

- Rate (mathematics), a specific kind of ratio, in which two measurements are related to each other (often with respect to time)
  - Rate of travel, or velocity
  - Bit rate, number of bits that are conveyed or processed per unit of time
- Naval rating or rate, terms used to designate specialty or seniority of enlisted naval personnel
- Rate of a ship, a term indicating a sail ship's firepower in the British Royal Navy
- Rates (Portuguese parish), a Portuguese parish and town located in the municipality of Póvoa de Varzim
- RATE project, a young earth creationism research project

Rates commonly occurring in BEAST

- Substitution model rates
  - CTMC
  - Covarion
  - Stochastic dollo
- Gamma rate heterogeneity
  - with
  - without
- Relative substitution rates
  - fixed to 1
  - estimated
- Clock rates
  - Strict
  - Relaxed
  - Random local

Which one to choose? Which combinations make sense?

Database of cognates

word list

language	hand	mother	father	...
English	hand	mother	father	...
Dutch	hand	moeder	vader	...
German	hand	mutter	vater	...
French	main	mère	père	...
Spanish	mano	madre	padre	...
Dhudhuroa	?	papa	mama	...

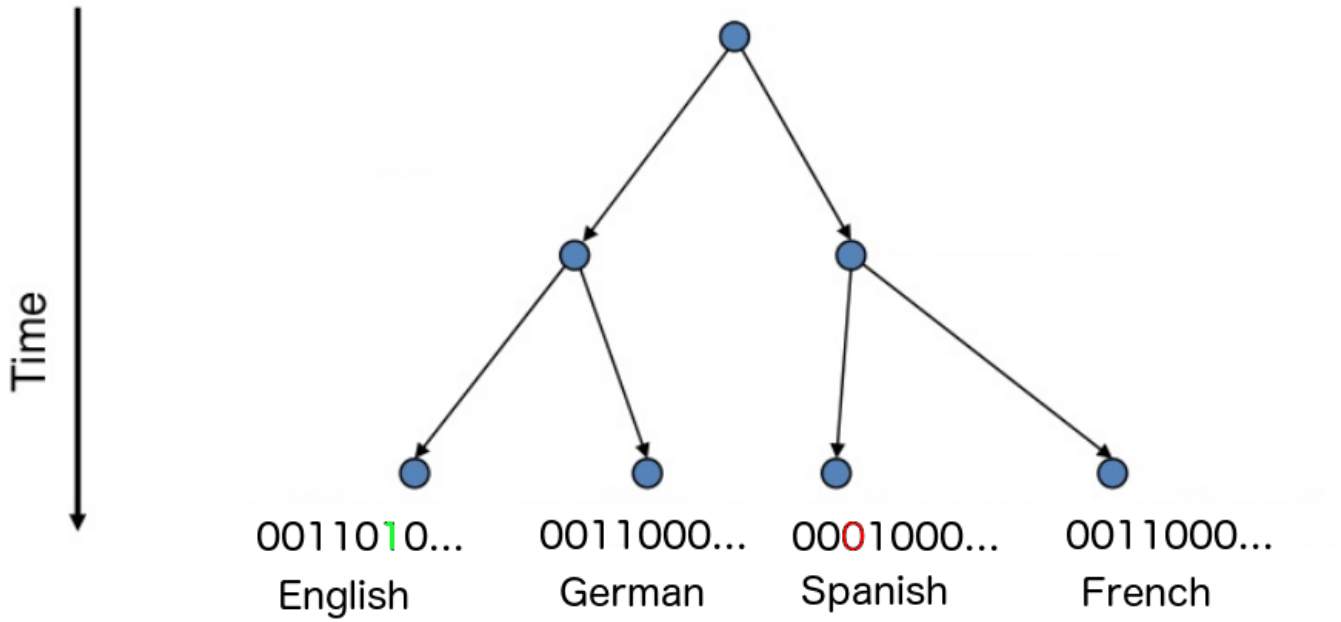
cognate list

language	ascetainment	hand	mano	mother	papa	father	mama	...
English	0	1	0	1	0	1	0	...
Dutch	0	1	0	1	0	1	0	...
German	0	1	0	1	0	1	0	...
French	0	0	1	1	0	1	0	...
Spanish	0	0	1	1	0	1	0	...
Dhudhuroa	0	?	?	0	1	0	1	...

Sites: columns in the table

Tree-likelihood

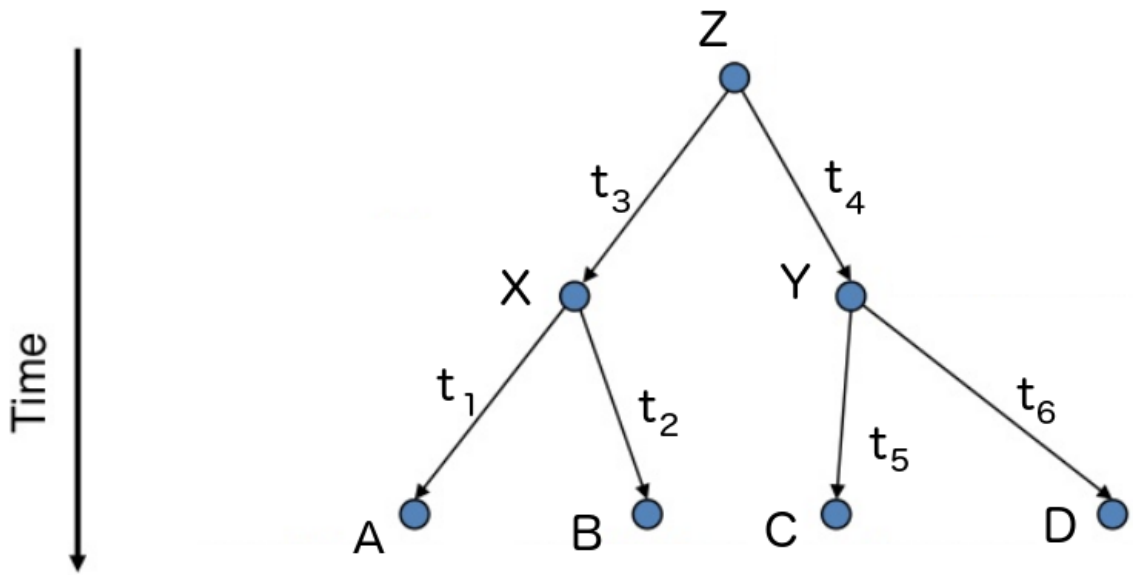
We want to calculate $P(data|tree, rates)$



Assume sites are independent:
$P(data|tree, rates) = \prod_{i \in sites} P(site_i|tree, rates)$

Substitution model

$P(site_i|tree, rates) = P(A|X, t_1, \theta)P(B|X, t_2, \theta)P(X|Z, t_3, \theta)P(C|Y, t_5, \theta)$
$P(D|Y, t_6, \theta)P(Y|Z, t_4, \theta)P(Z)$
$=$ product over all branches $\times$ root distribution.



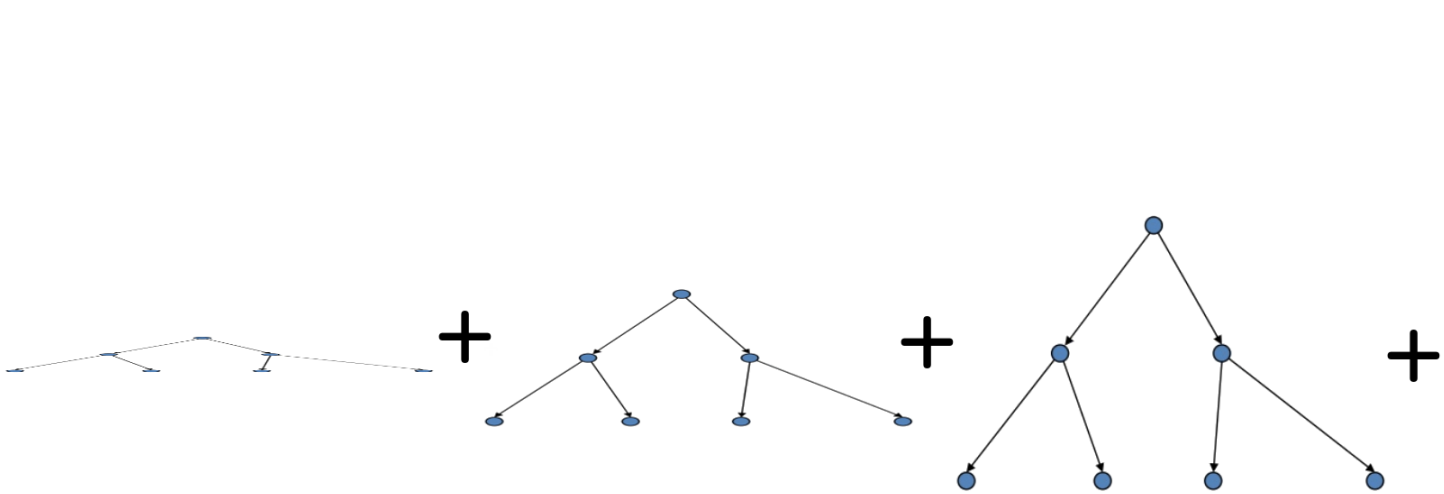
θ the substitution model rates

Gamma rate heterogeneity

Gamma rate heterogeneity

Gamma rate heterogeneity

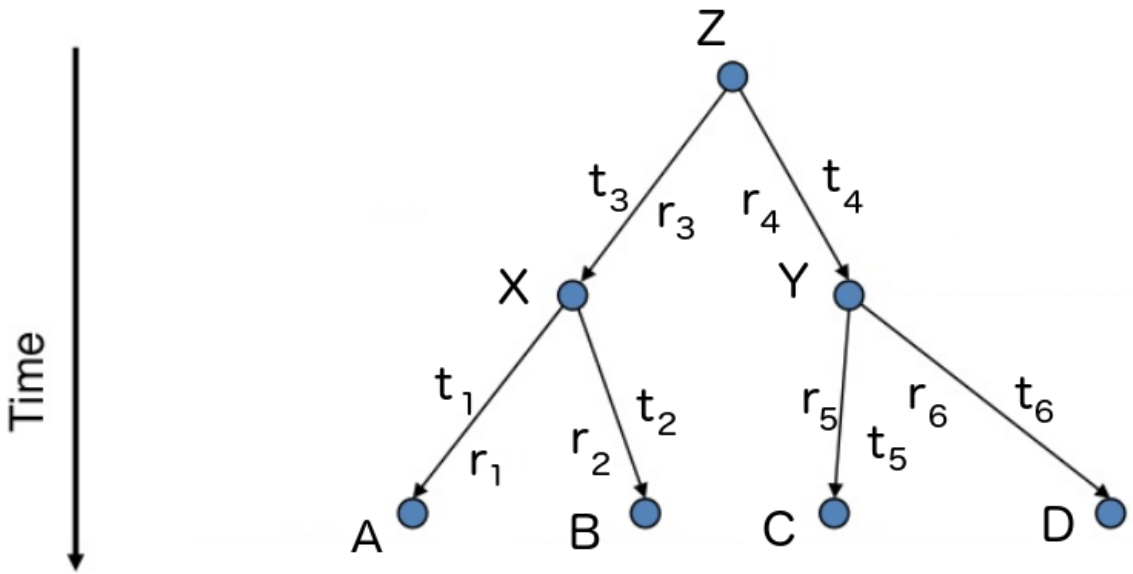
Mixture of tree likelihoods, trees scaled by gamma category rate



Clock rates

Associate a rate r_j with each branch

Use $r_i \times t_i$ instead of t_i for $P(site_i|tree, rates)$.



- Strict clock: all rates equal
- Uncorrelated relaxed clock: all rates independent, drawn from common distribution
- Random local clock: some rates same as parent branch rates

Summary so far

- Tree likelihood is product of likelihoods over sites
- Likelihood over sites is product of $P(A|B,t,\theta)$ over branches
- Substitution model rates determine $P(A|B,t,\theta)$
- Gamma rate heterogeneity: rate variation across sites
- Clock models: rate variation across branches

What about substitution rates?

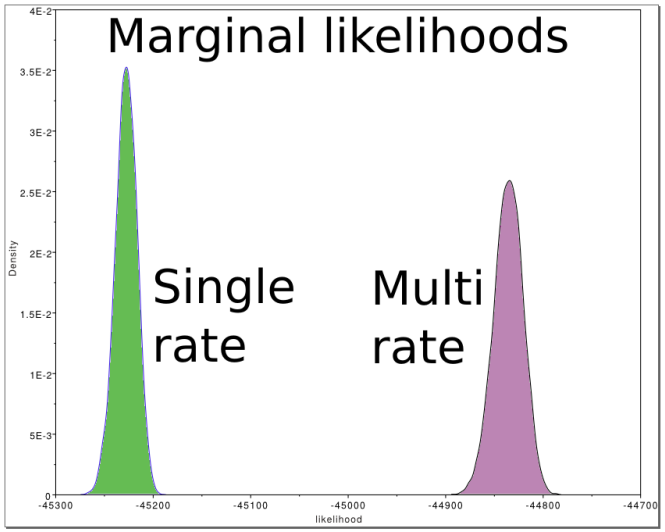
Change et al's model

word list

language	hand	mother	father	...
English	hand	mother	father	...
Dutch	hand	moeder	vader	...
German	hand	mutter	vater	...
French	main	mère	père	...
Spanish	mano	madre	padre	...
Dhudhuroa	?	papa	mama	...

cognate list

language	ascetainment	hand	mano	mother	papa	father	mama	...
English	0	1	0	1	0	1	0	...
Dutch	0	1	0	1	0	1	0	...
German	0	1	0	1	0	1	0	...
French	0	0	1	1	0	1	0	...
Spanish	0	0	1	1	0	1	0	...
Dhudhuroa	0	?	?	0	1	0	1	...

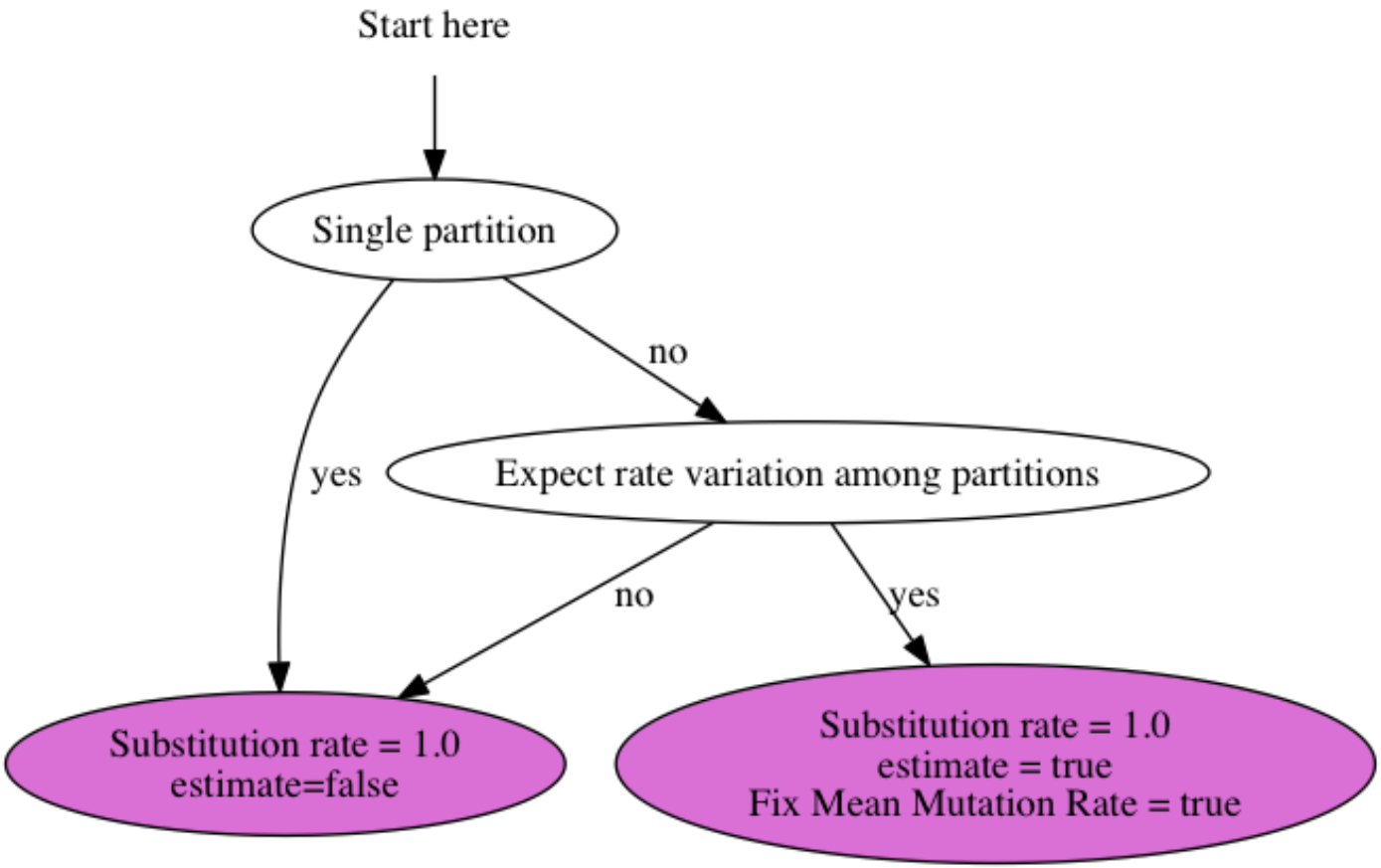


cognate list

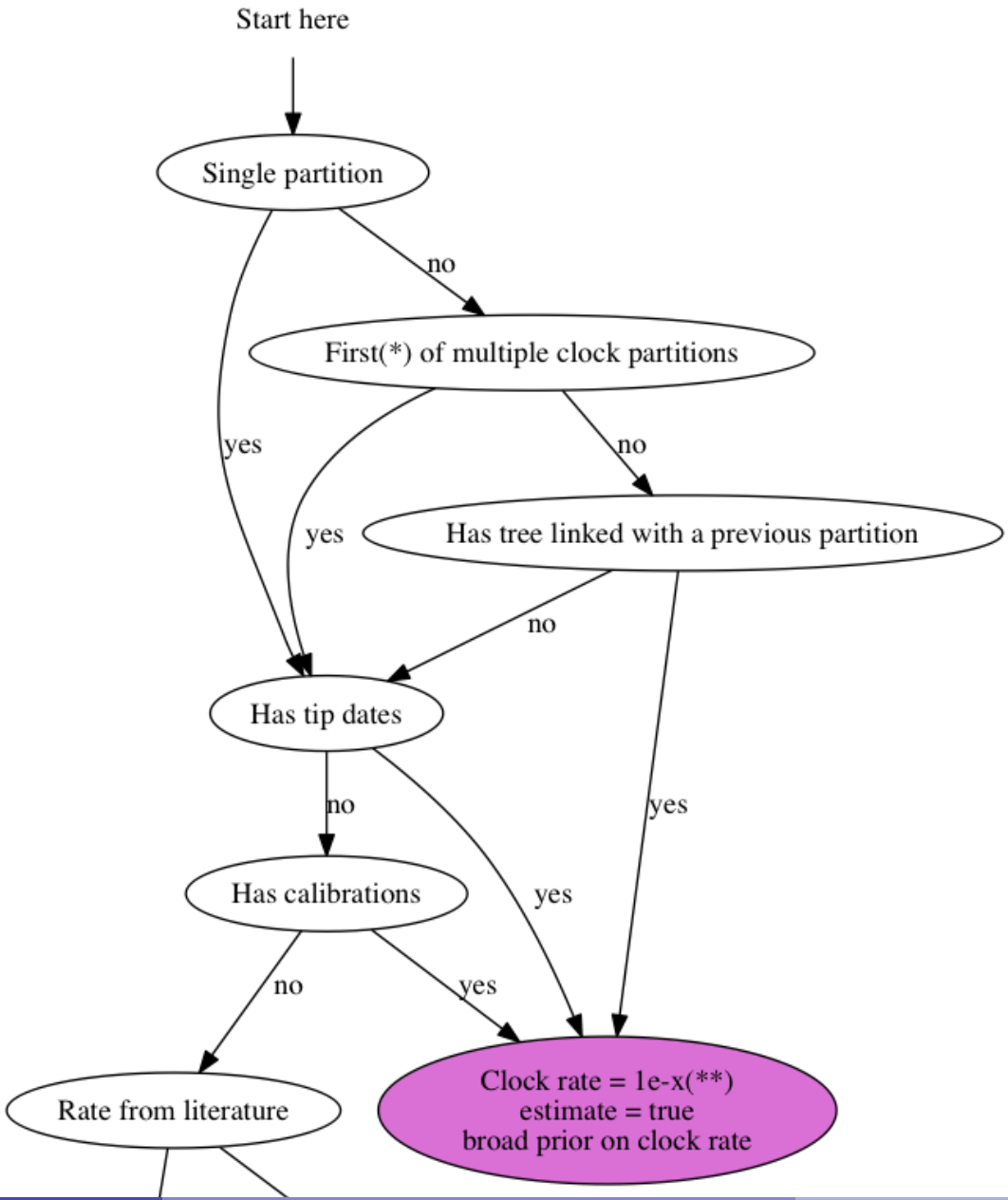
language	ascetainment	hand	mano	ascetainment	mother	papa	ascetainment	father	mama	...
English	0	1	0	0	1	0	0	1	0	...
Dutch	0	1	0	0	1	0	0	1	0	...
German	0	1	0	0	1	0	0	1	0	...
French	0	0	1	0	1	0	0	1	0	...
Spanish	0	0	1	0	1	0	0	1	0	...
Dhudhuroa	?	?	?	0	0	1	0	0	1	...

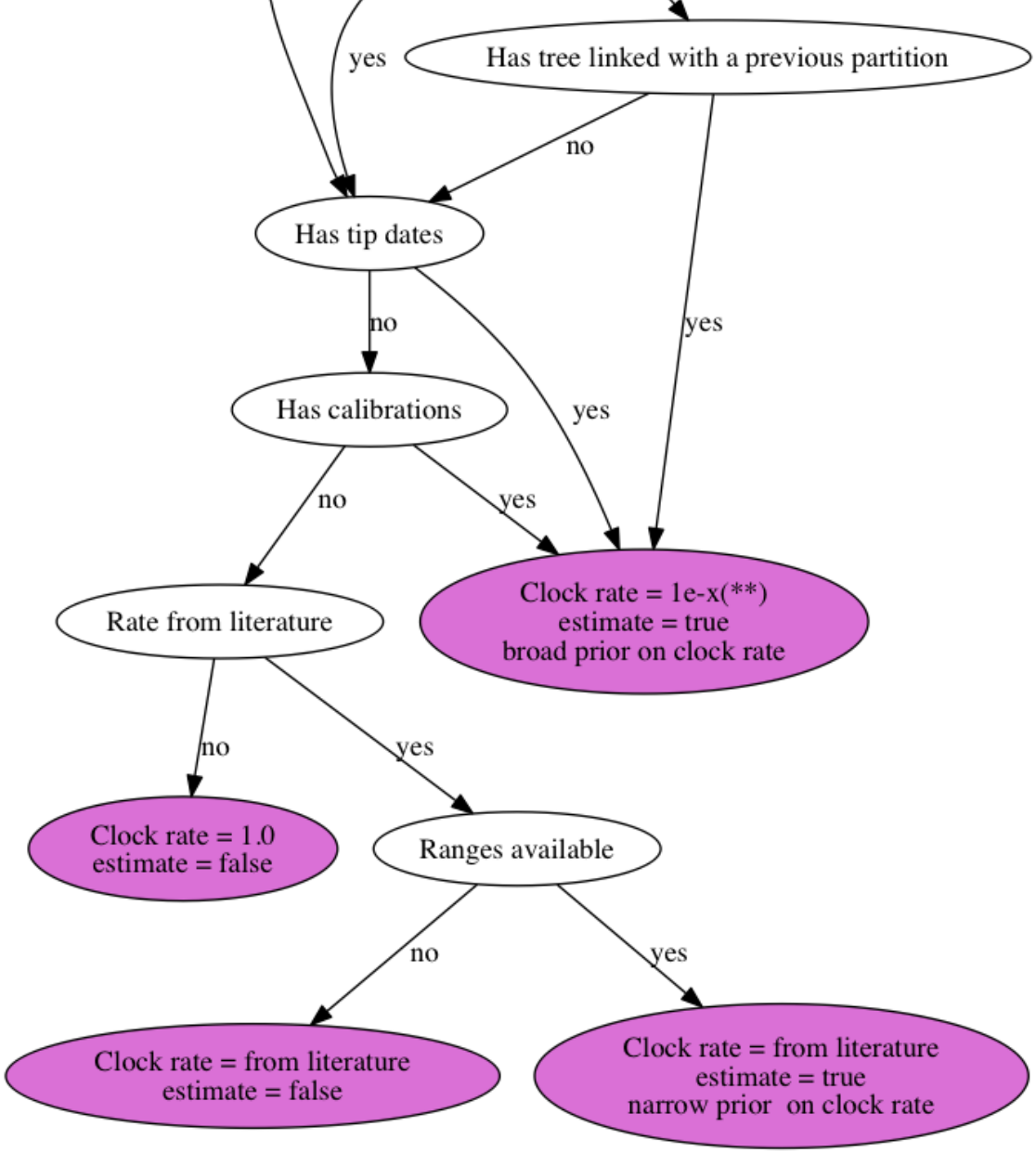
+ 1 relative rate for every meaning class = 207-1 extra parameters

Setting up substitution rates



Setting up clock rates





**choose appropriate starting value

Summary

- Tree likelihood is product of likelihoods over sites
- Likelihood over sites is product of $P(A|B,t,\theta)$ over branches
- Substitution model rates determine $P(A|B,t,\theta)$
- Gamma rate heterogeneity: rate variation across sites
- Clock models: rate variation across branches
- Substitution rates: rate variation across partitions