



Language Phylogenies

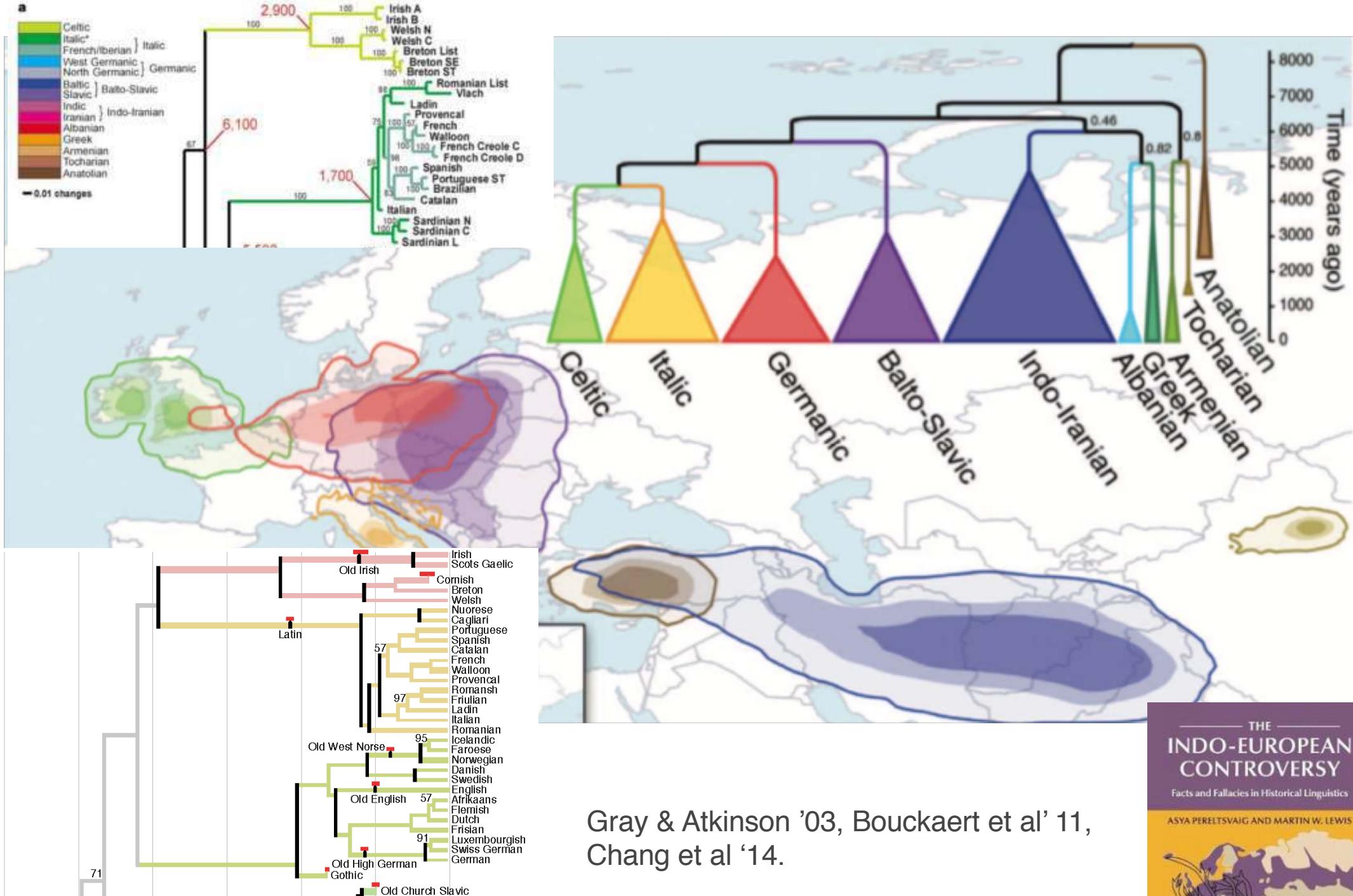
Simon J. Greenhill

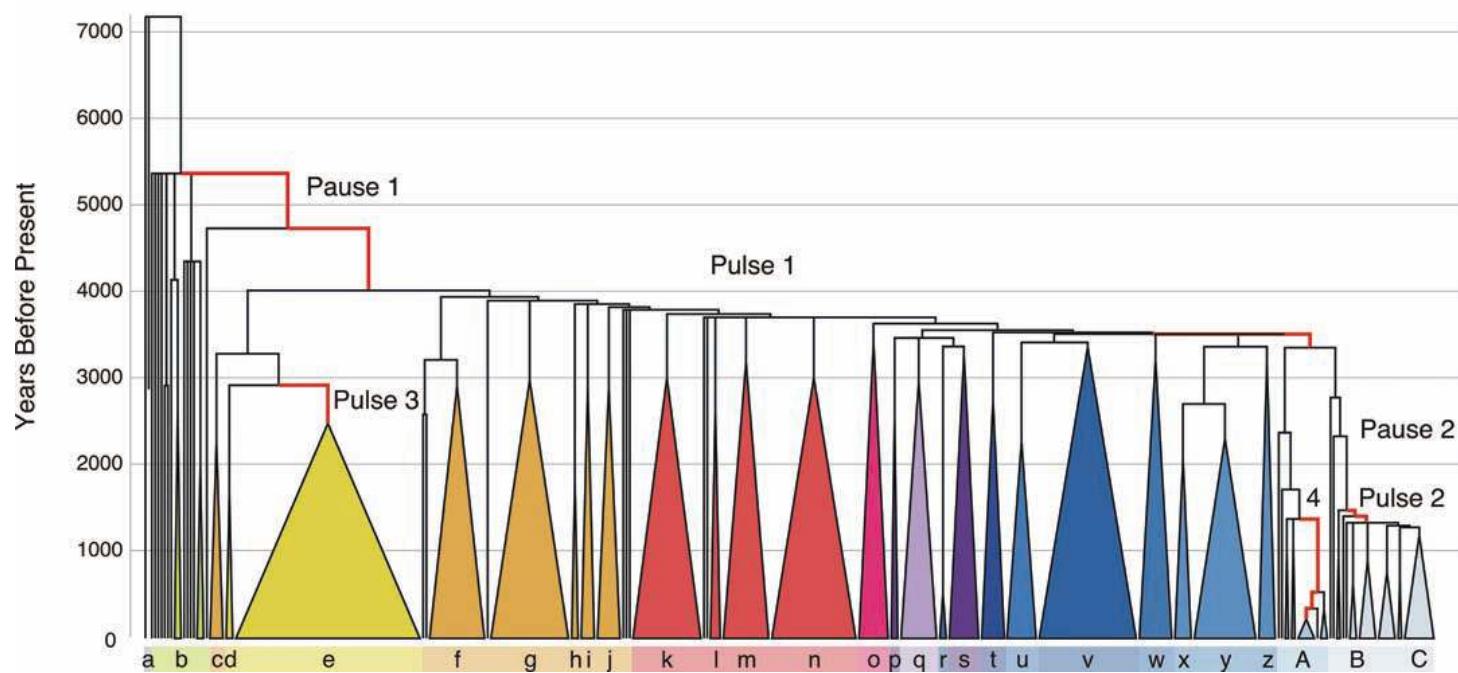
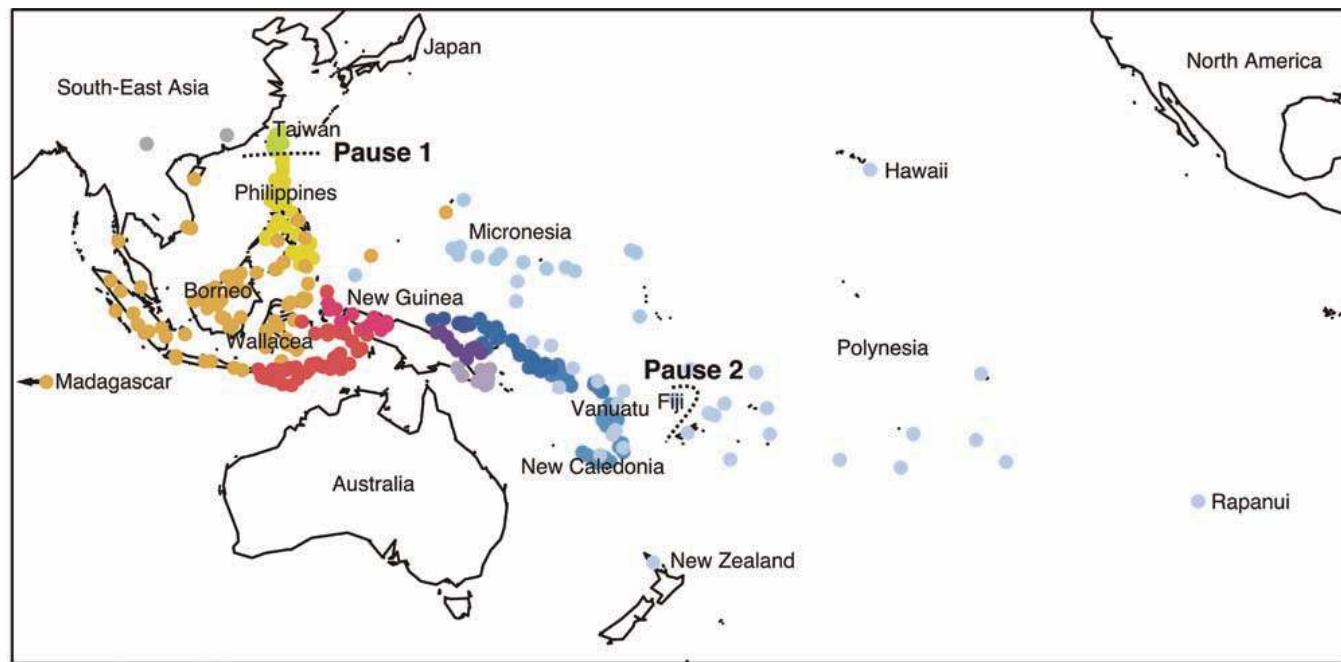


ARC CENTRE OF EXCELLENCE FOR
THE DYNAMICS OF LANGUAGE

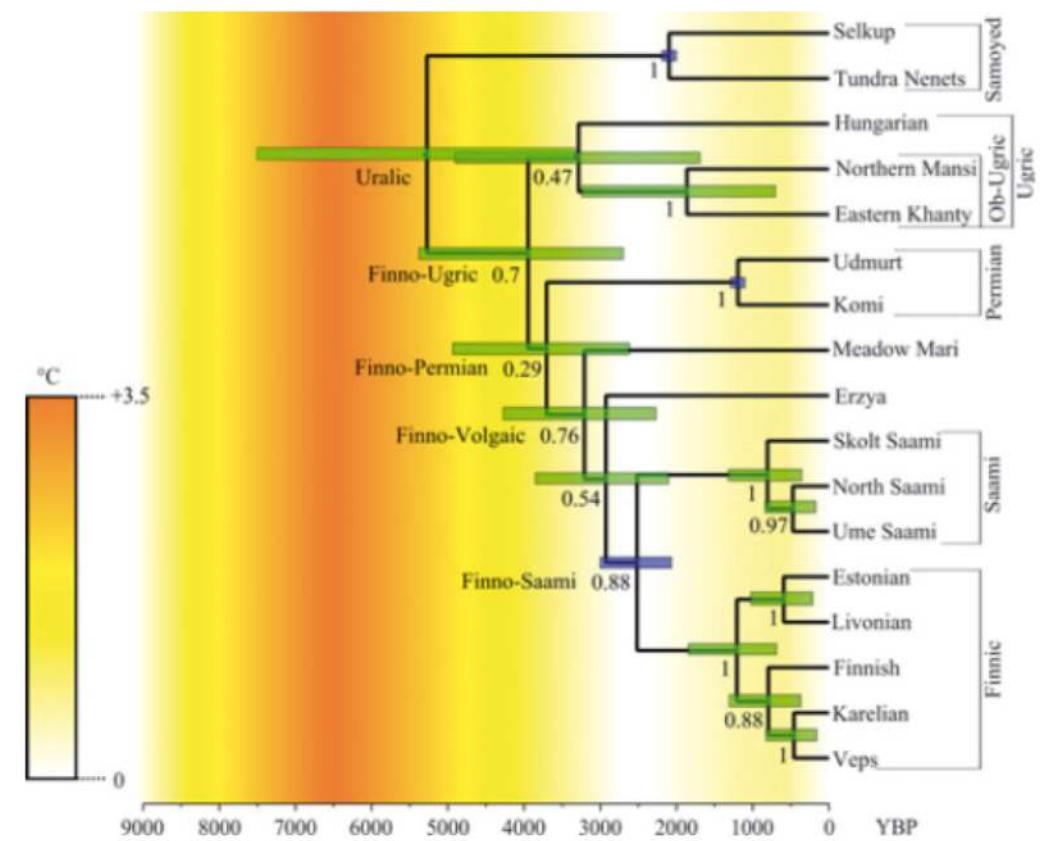
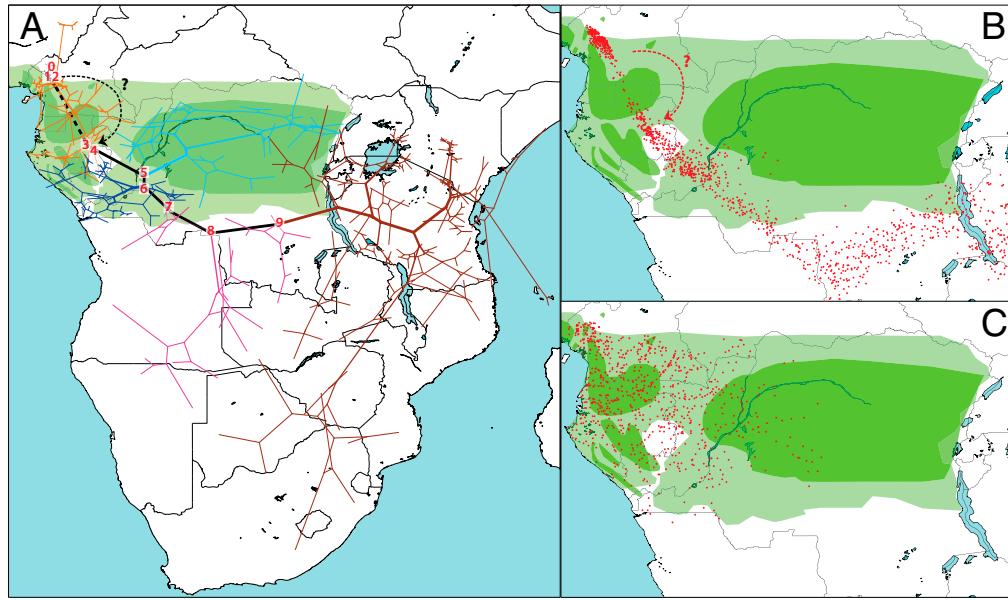


Max Planck Institute for the
Science of Human History

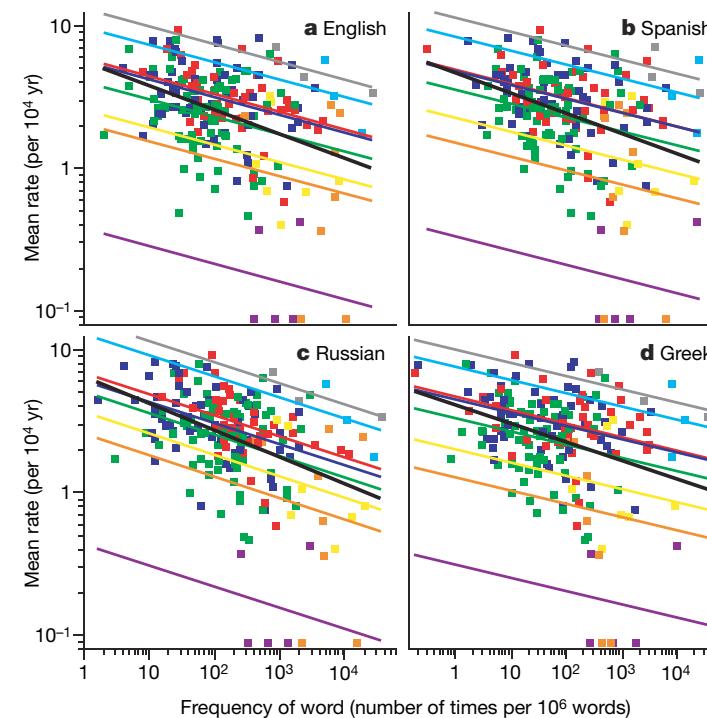
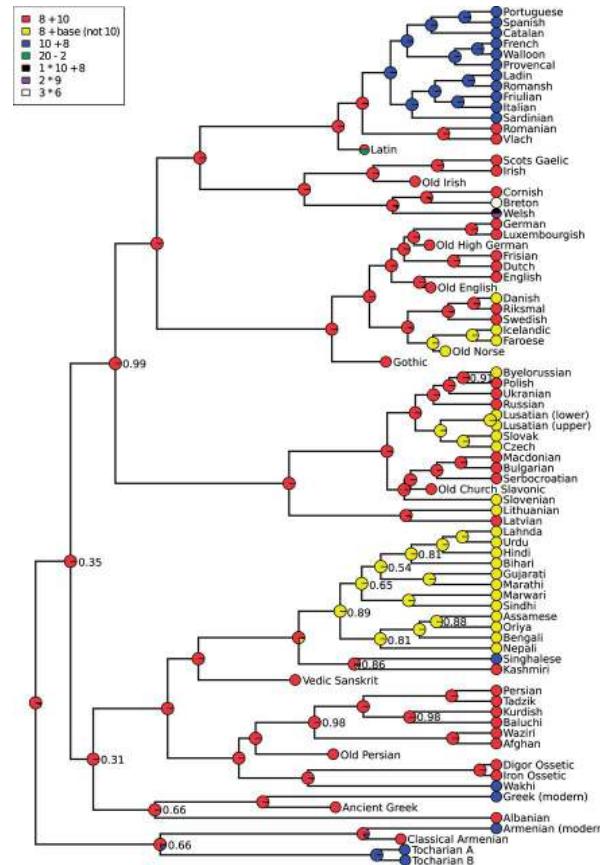
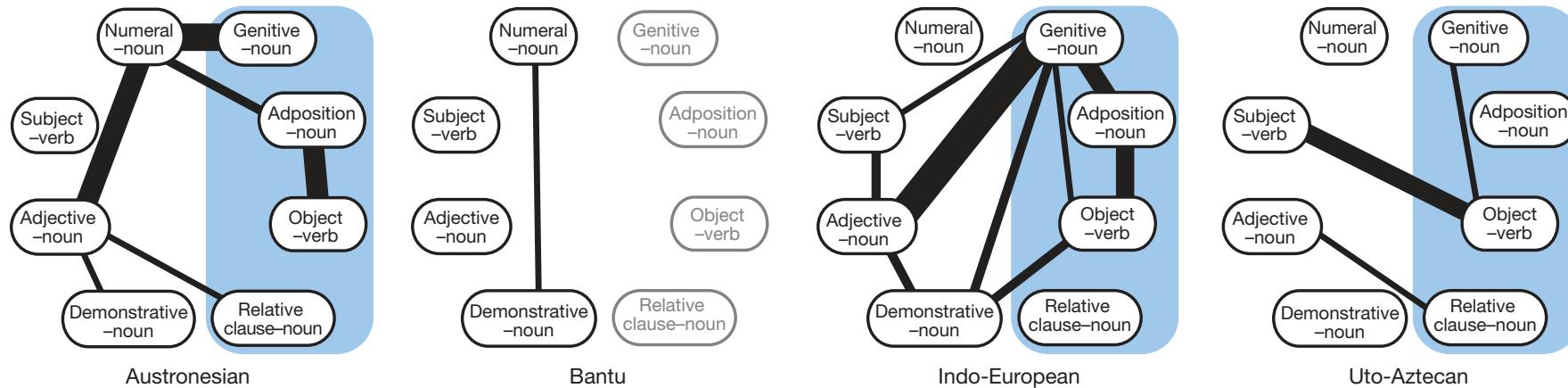




Gray et al '09

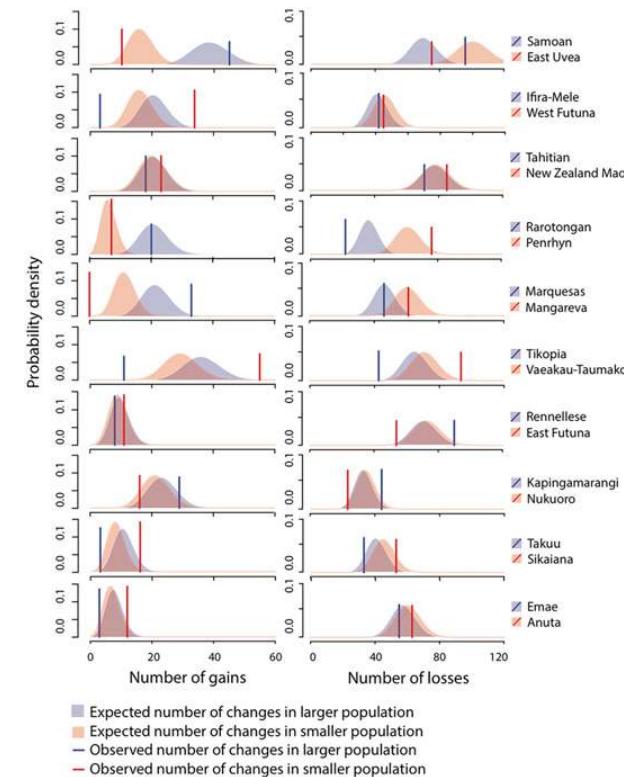


Dunn et al. '15



Calude and Verkerk '16

Pagel et al '07



Bromham et al. '15

Controversial

“most vibrant stream
of contemporary
linguistics”

“Computational methodologies
of this kind can only be helpful
for historical linguistics”

“languages and biomolecular
sequences evolve in very
different ways”

“more questions than
answers”

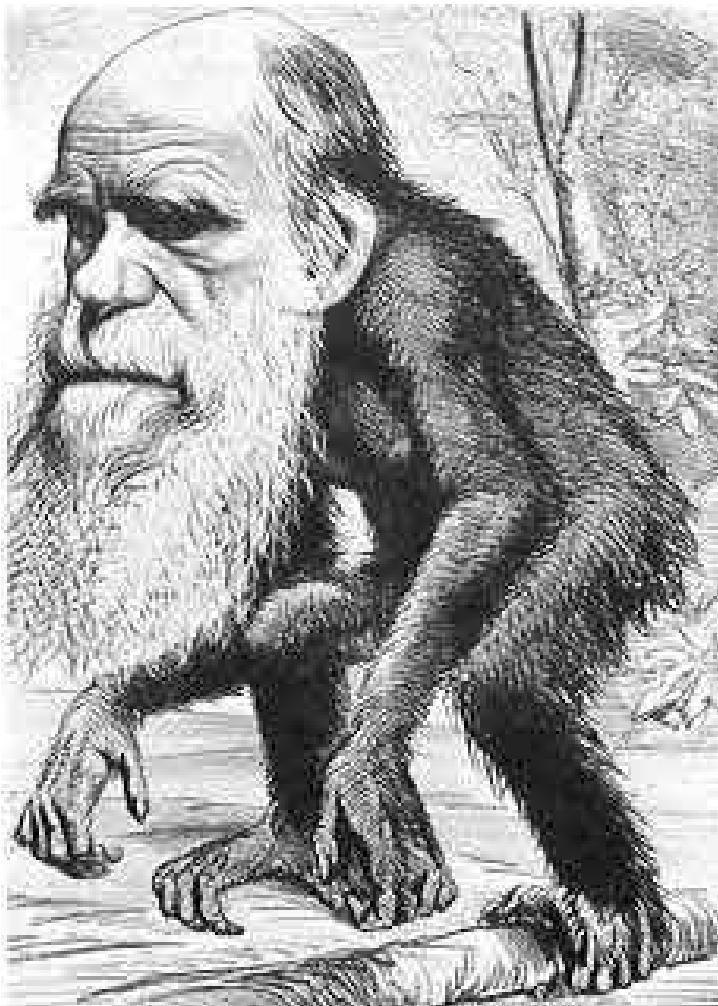
“utter bollocks”

“biggest intellectual fraud
since Chomsky”

“this isn’t history, it’s history put in nested boxes!”



What is evolution?



Variation

Heritability

Differential survival

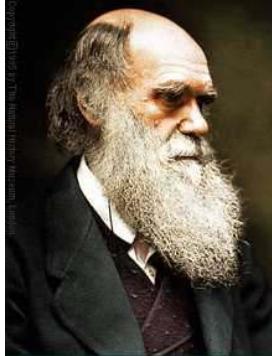
When and where did 🐨👩💬 originate?

What **differences** are there between 🐨👩💬's?

How are 🐨👩💬 **related** to other 🐨👩💬's?

What **processes** shaped 🐨👩💬?

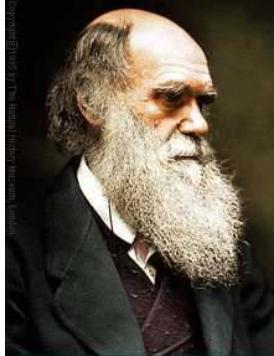
Can we infer what 🐨👩💬 were in the **past**?



Darwin (1871)

"Languages, like organic beings, can be classed in groups under groups...

"The formation of different languages and of distinct species, and the proofs that both have been developed through a gradual process, are **curiously parallel**"



Darwin (1871)

"Languages, like organic beings, can be classed in groups under groups..."

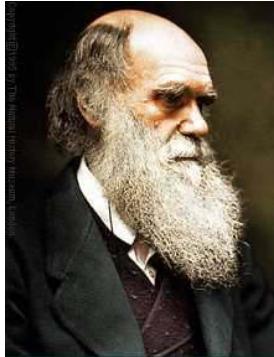
"The formation of different languages and of distinct species, and the proofs that both have been developed through a gradual process, are **curiously parallel**"



Schleicher (1863)

Darwinism Tested by the Science of Language

"same process has long been generally assumed for linguistic organisms"



Darwin (1871)

"Languages, like organic beings, can be classed in groups under groups..."

"The formation of different languages and of distinct species, and the proofs that both have been developed through a gradual process, are **curiously parallel**"



Schleicher (1863)

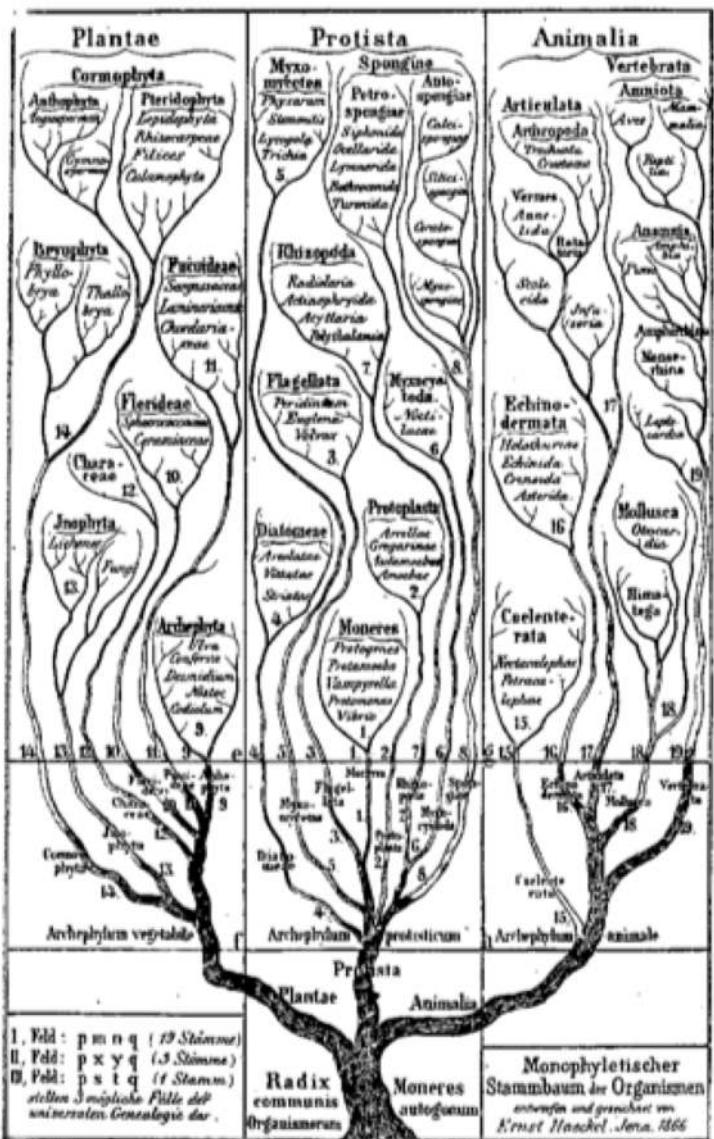
Darwinism Tested by the Science of Language

"same process has long been generally assumed for linguistic organisms"

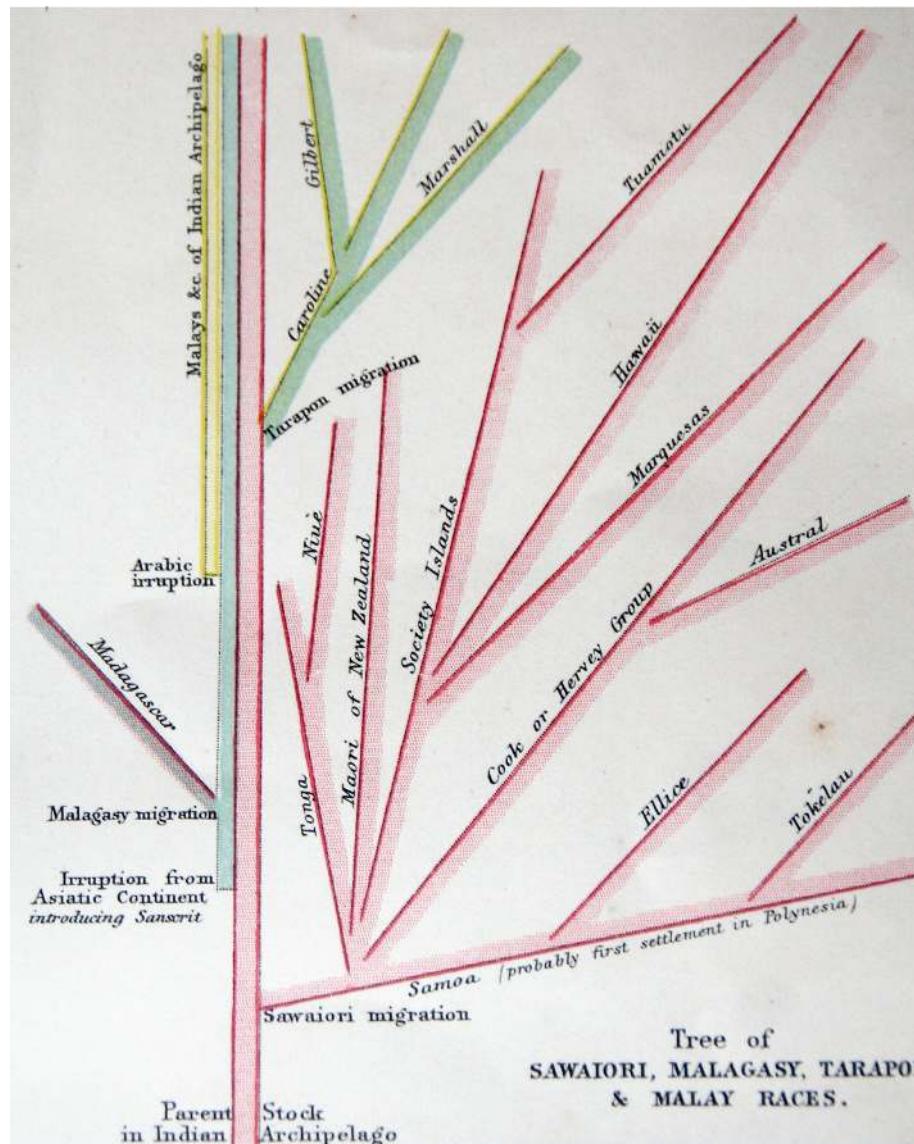


Brugmann (1884)

Importance of using "shared innovations" to define groups and not "shared retentions"



Haeckel 1866



Encyclopedia Britannica 1888





Lexicostatistics

1. Wordlists of “Basic” vocabulary.
2. Count shared cognates between language pairs (=“retention rate”)
3. Cluster languages with highest similarity

Swadesh (1950, 1952, 1955)

	Taboo	Blood	To Suck
Fijian	tabu	drā	sucu-ma
Tahitian	tapu	toto	ngote
Maori	tapu	toto	ngote
Hawaiian	kapu	koko	omo
Marquesan	tapu	toto	omo

Identified by Systematic Sound Correspondences
- e.g. Maori “t” = “k” in Hawaiian.

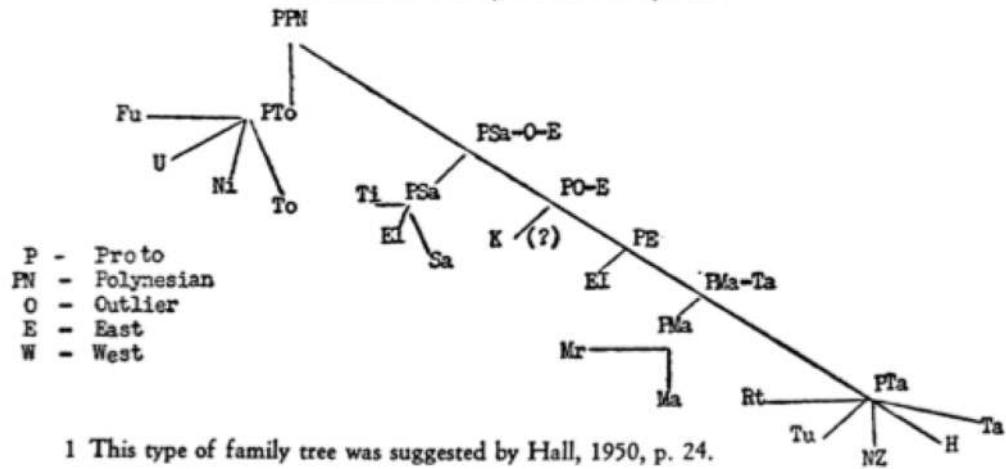
Elbert 1953

TABLE 2
Polynesian cognate percentages

U	N ₁	T ₀	T ₁	E	S _a	S ₁ ¹	F ₁ ¹	O _J ¹	N _u ¹	K	EI	Mr	Ma	Rt	Tu	NZ	H	Ta
63	62	74	83	79	74	57	62	62	54	52	54	57	56	66	62	61	58	58
	72	86	78	74	70	53	59	58	52	51	53	53	51	62	62	61	55	59
	64	68	61	63	50	59	49	54	49	49	48	55	47	56	55	54	49	51
		70	64	66	46	55	55	49	45	48	49	45	58	53	54	49	52	
			81	76	66	66	62	59	59	62	67	63	71	68	71	67	66	
				78	66	61	62	59	58	61	62	63	66	66	67	68	66	
					64	60	61	55	53	53	55	52	67	62	57	59	60	
					55	60	52	55	52	55	56	60	58	60	60	60	59	
					57	52	54	54	54	59	53	55	61	60	55	54		
						51	54	53	53	51	53	56	55	56	55	56	53	
							53	49	45	44	53	50	52	52	49	48		
								47	49	45	54	51	51	51	51	49	50	
								64	63	64	62	63	64	62				
									73	75	72	70	69	68				
									73	69	67	70	67	67				
										83	83	79	85					
										79	77	83						
											71	73						
												76						

1 Percentages based on incomplete data.

TABLE 4
A tentative family tree for Polynesia¹



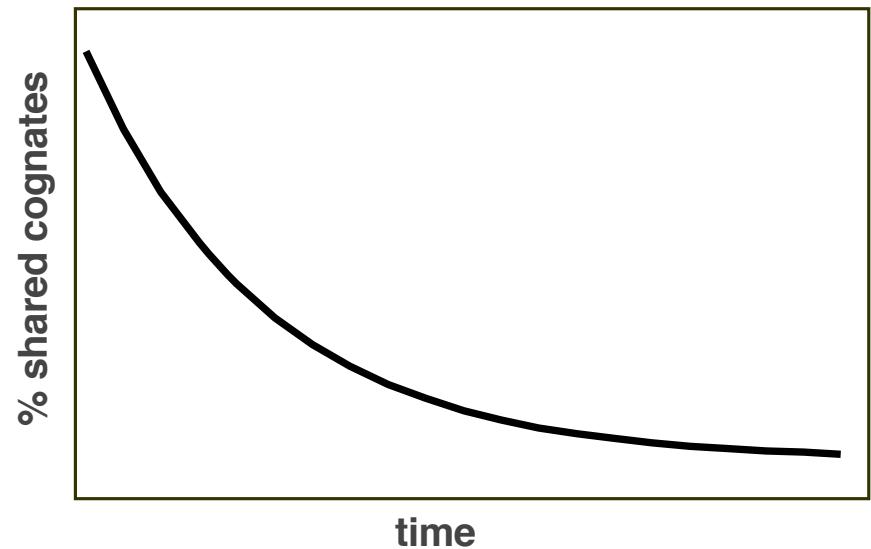


Glottochronology

Loss of cognates happens at a constant rate
 (=radioactive decay)

19% loss per 1000 years (Lees 1953)

$$time = \frac{\log(\% \text{ shared cognates})}{2 \log(\text{retention rate})}$$



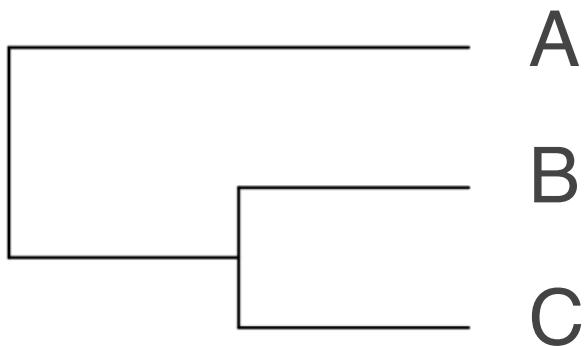
Retention Rate

- Rate at which cognates are lost.
- Swadesh 1952
 - Compared Old & Modern English, 1000 years apart
 - Shared 85% similarity: $r=0.85/1000$
- Lees 1953
 - Surveyed ‘control cases’
 - Average rate of 0.81% \pm 0.176



Language	Similarity	Age	Retention Rate
Old English vs. Modern English	76.6%	1,000	0.766
Plautine Latin vs. Early Spanish	65.5%	1,800	0.790
Plautine Latin vs. Molière's French	62.5%	1,850	0.776
Old High German vs. Modern German	84.2%	1,100	0.854
Middle Egyptian vs. Coptic	53.0%	2,200	0.760
Koine Greek vs. Modern Athenian Greek	69.0%	2,070	0.836
Koine Greek vs. Modern Cypriot	67.8%	2,070	0.829
Ancient Classical Chinese vs. Mod. Mandarin	79.6%	1,000	0.795
Old Norse vs. Modern Swedish	85.0%	1,020	0.854
Classical Latin vs. Modern Tuscan	68.6%	2,150	0.839
Classical Latin vs. Modern Portuguese	62.9%	2,150	0.806
Classical Latin vs. Modern Rumanian	56.0%	2,150	0.764
Classical Latin vs. Modern Catalan	60.6%	2,150	0.793

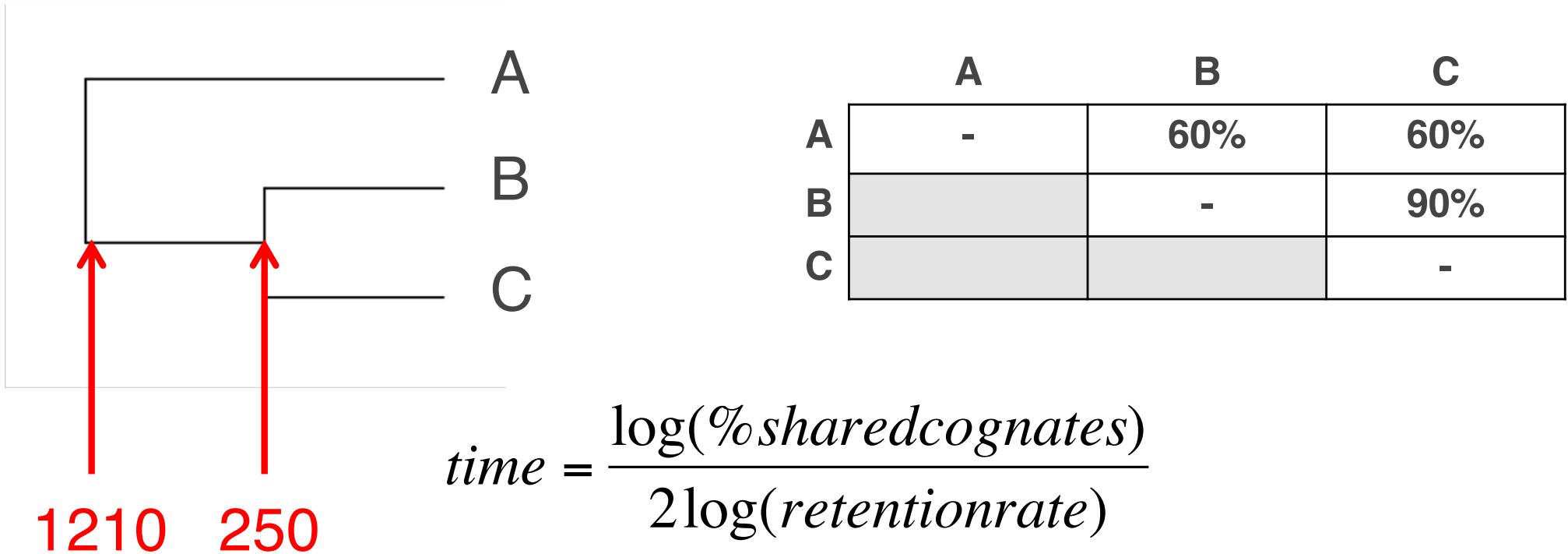
Example



	A	B	C
A	-	60%	60%
B		-	90%
C			-

$$time = \frac{\log(\% shared cognates)}{2 \log(retention rate)}$$

Example



B-C: 90% similarity

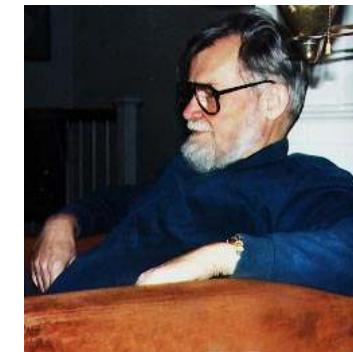
$$\begin{aligned} &= \text{Log}(0.90) / 2 \text{ Log}(0.81) \\ &= 0.250 \\ &= 250 \text{ years (* 1000)} \end{aligned}$$

A-B: 60% similarity

$$\begin{aligned} &= \text{Log}(0.60) / 2 \text{ Log}(0.81) \\ &= 1.21 \\ &= 1,210 \text{ years (* 1000)} \end{aligned}$$

The Rise of Lexicostatistics...

IN THE LAST DECADE glottochronology has excited international interest and acquired a literature of its own. To the anthropologist it promises a measure of time depth for language families without documented history, and yet another linguistic example of regularity in cultural phenomena.



Hymes (1960): “Lexicostatistics so far”

“... a significant work—one which may conceivably be as revolutionary for Oceanic linguistics and culture history as was the work of Greenberg (1949–54) for the interpretation of African languages and cultures”

Murdock (1964) p.117



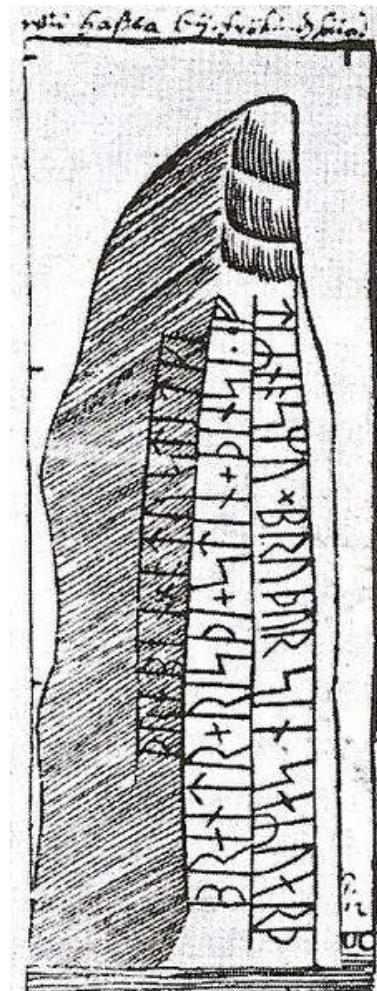
...and the fall of Lexicostatistics

Major Criticism: Universality of Rates

Old Norse & Icelandic?

- Glottochronology: 200 years.
- Reality: 1000 years

Bergsland & Vogt 1962: “Our findings clearly disprove the basic assumption of glottochronology ‘that fundamental vocabulary changes at a constant rate’ ”



Jungner, Hugo; Elisabeth Svärdström (1940-1971). Sveriges runinskrifter: V. Västergötlands runinskrifter. Stockholm: Kungl. Vitterhets Historie och Antikvitets Akademien. ISSN 0562-8016. p. 260

Fallout.

"a tradition of hostility towards probabilistic modelling in historical linguistics" (Sankoff '73)

"In summary, glottochronology is not accurate; all its basic assumptions have been severely criticized. It should not be accepted, it should be rejected"
(Campbell '04)

"Linguists don't do dates" (McMahon & McMahon '03)



...and yet...

- 1997: "(nine minimum centuries divergence in glottochronological counts, for whatever that may be worth)"
- 2002: "For what it is worth, the glottochronological method yields an estimate of about 1300 years as the period of common development for Polynesian".



U.P.G.M.A

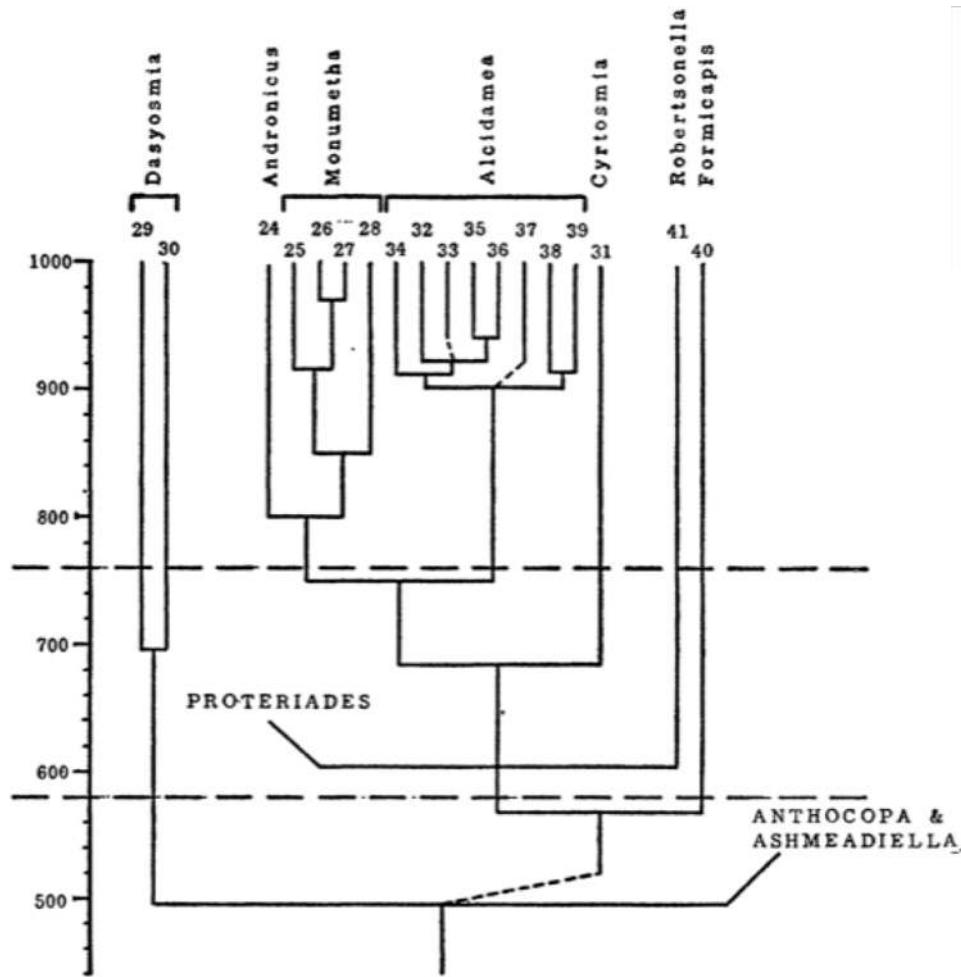


FIG. 6. Diagram of relationships for the genus *Hoplitis* obtained by the weighted variable group method.

A QUANTITATIVE APPROACH TO A PROBLEM
IN CLASSIFICATION¹

CHARLES D. MICHENER AND ROBERT R. SOKAL²

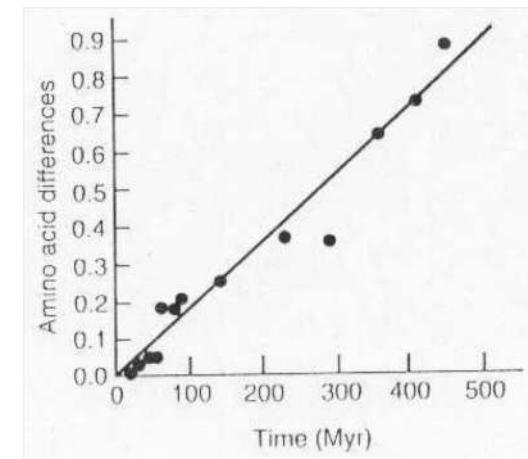
Department of Entomology, University of Kansas, Lawrence



Molecular Clock

Zuckerkandl & Pauling 1962

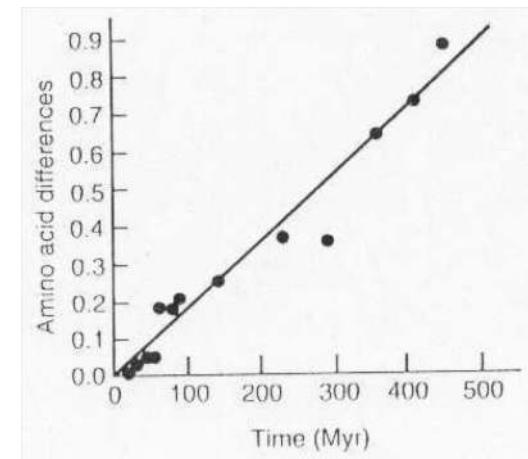
Number of AA differences were proportional to species divergence times.



Molecular Clock

Zuckerkandl & Pauling 1962

Number of AA differences were proportional to species divergence times.



Kimura 1968

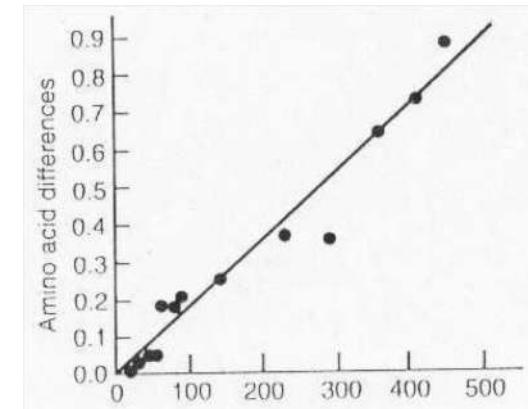
The average time taken for one base pair replacement within a genome is therefore

$$28 \times 10^6 \text{ yr} \div \left(\frac{4 \times 10^9}{300} \right) \div 1.2 \doteq 1.8 \text{ yr}$$

Molecular Clock

Zuckerkandl & Pauling 1962

Number of AA differences were proportional to species divergence times.



Kimura 1968

The average time taken for one base pair replacement within a genome is therefore

$$28 \times 10^6 \text{ yr} \div \left(\frac{4 \times 10^9}{300} \right) \div 1.2 \div 1.8 \text{ yr}$$

Sarich & Wilson 1967

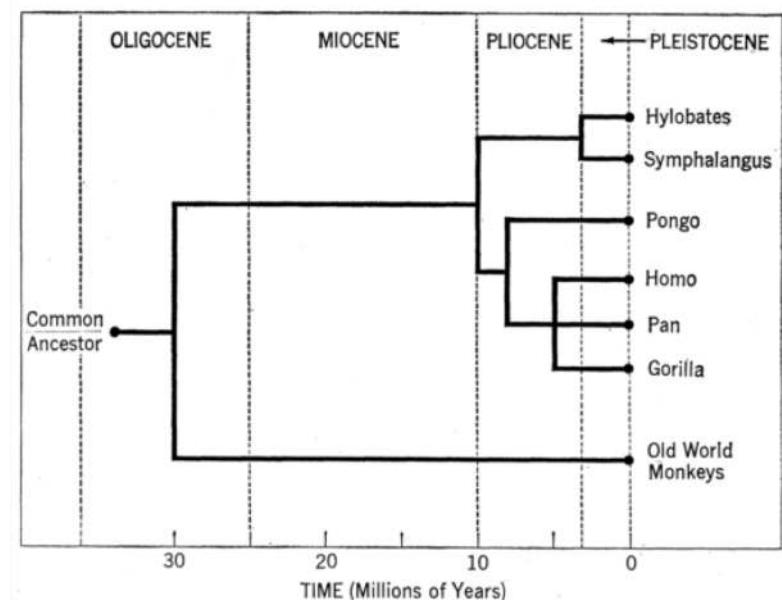


Fig. 1. Times of divergence between the various hominoids, as estimated from immunological data. The time of divergence of hominoids and Old World monkeys is assumed to be 30 million years.

Problems?

Kirsch 1969

SEROLOGICAL DATA AND PHYLOGENETIC INFERENCE:
THE PROBLEM OF RATES OF CHANGE

JOHN A. W. KIRSCH

Felsenstein 1978

CASES IN WHICH PARSIMONY OR COMPATIBILITY
METHODS WILL BE POSITIVELY MISLEADING¹

JOSEPH FELSENSTEIN

Britten 1986

Rates of DNA Sequence Evolution Differ
Between Taxonomic Groups

ROY J. BRITTEN

The Cladistics Wars



Solutions.

Cavalli-Sforza &
Edwards 1967

Yang 1993

Sanderson 1997

Drummond et al. 2006

Phylogenetic Analysis Models and Estimation Procedures

L. L. CAVALLI-SFORZA AND A. W. F. EDWARDS*

Maximum-Likelihood Estimation of Phylogeny from DNA Sequences When Substitution Rates Differ over Sites¹

Ziheng Yang

A Nonparametric Approach to Estimating Divergence Times in the Absence of Rate Constancy

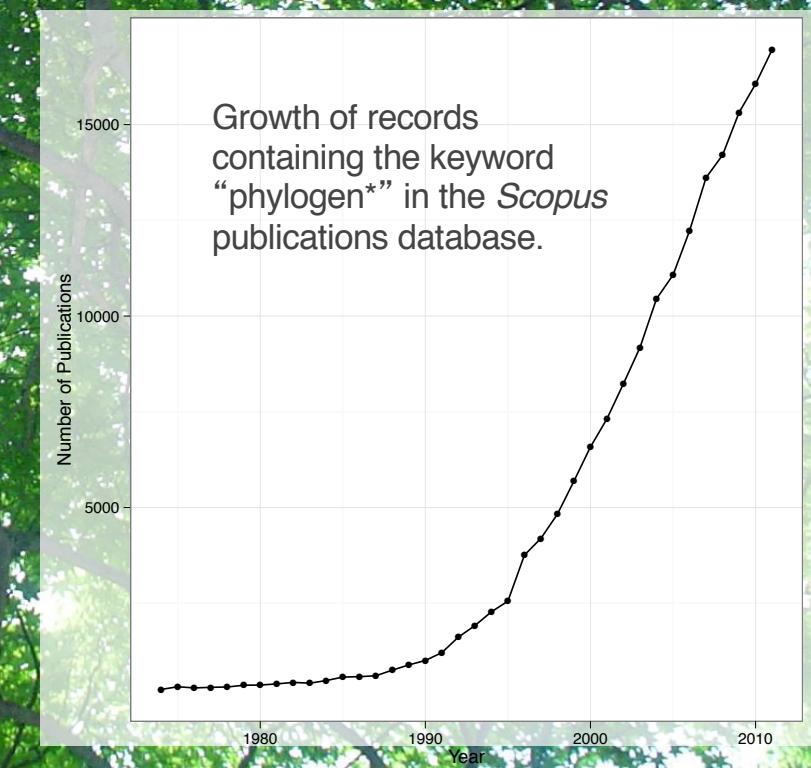
Michael J. Sanderson

Relaxed Phylogenetics and Dating with Confidence

Alexei J. Drummond[✉], Simon Y. W. Ho, Matthew J. Phillips, Andrew Rambaut^{✉}*

Phylogenetics

Range of methods in a robust, statistical/inferential framework for **testing** evolutionary hypotheses.

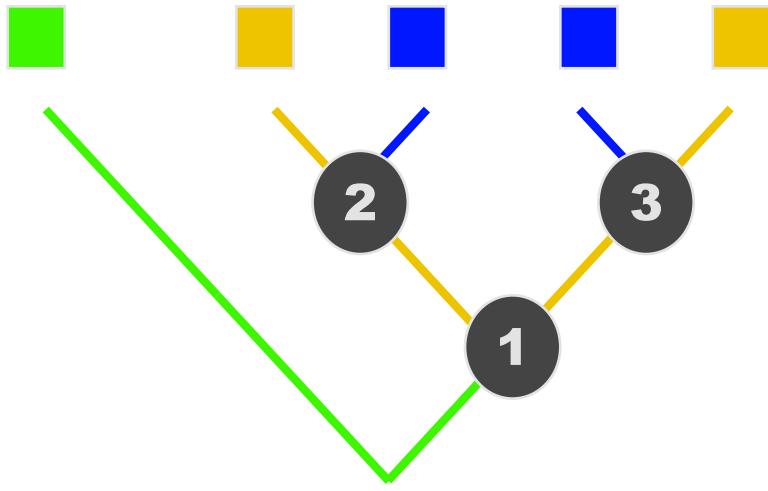


How do we build trees?

1. Distance Methods
2. Maximum Parsimony.
3. Maximum Likelihood.
4. **Bayesian Phylogenetic Analyses.**



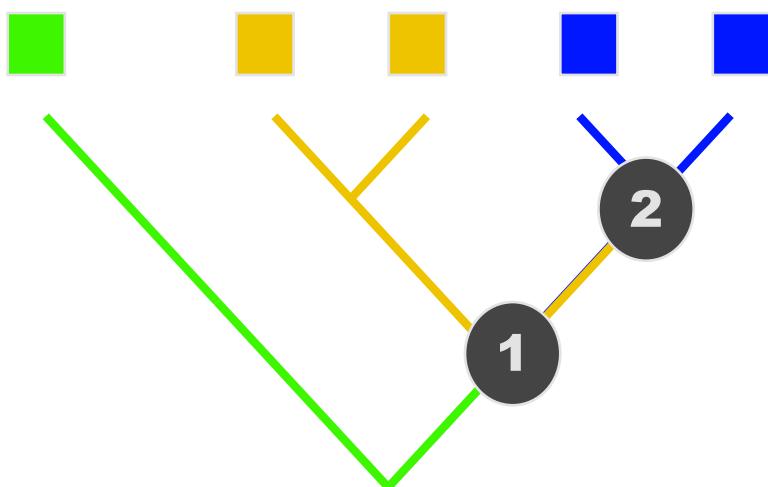
Maximum Parsimony



Unlikely that complex traits
should arise more than once

= Best tree is maximally
parsimonious

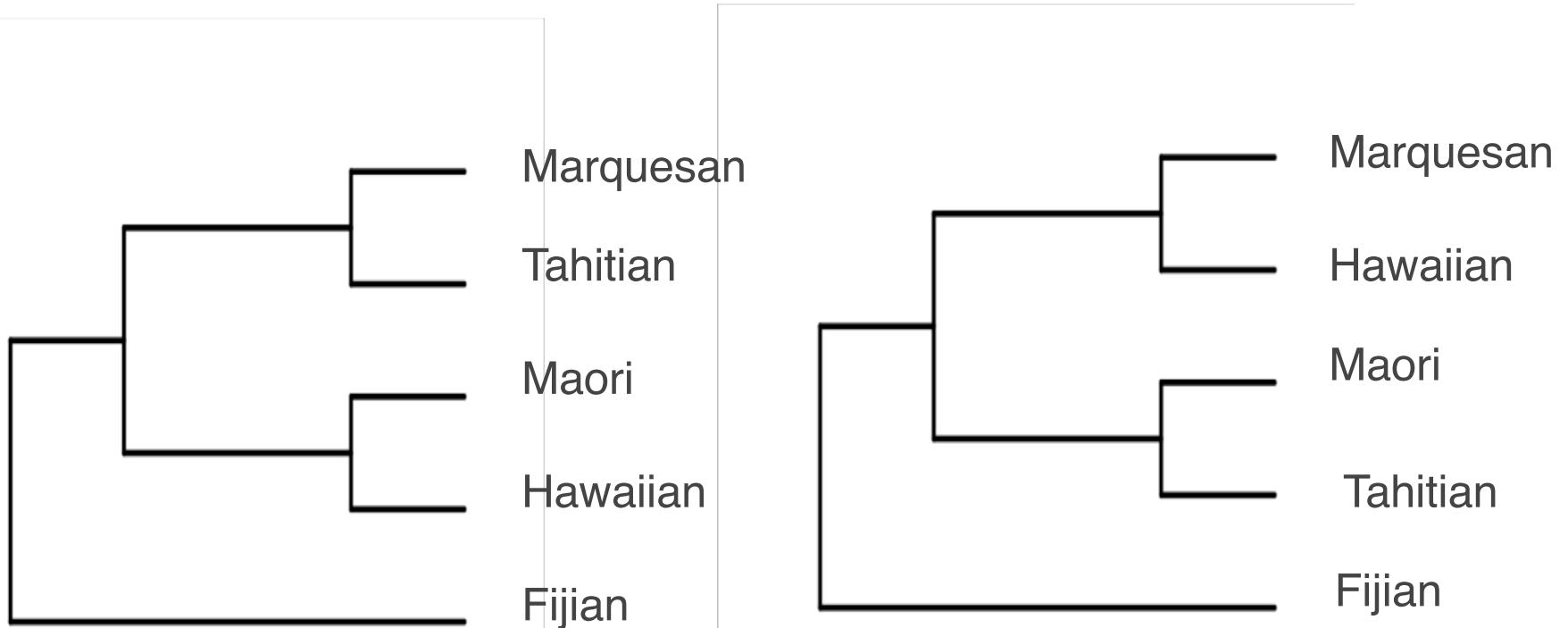
= Smallest amount of
evolution (Fewest number of
character state changes)



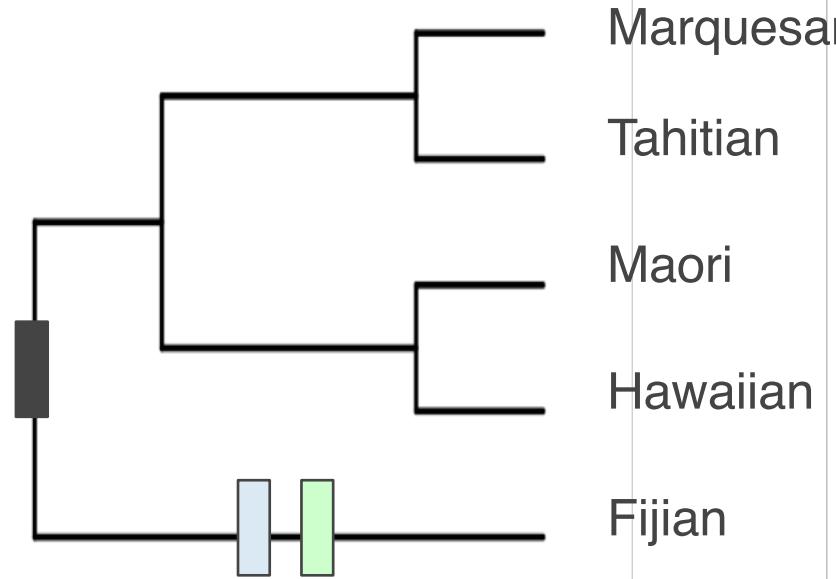
	Taboo	Blood	To Suck
Fijian	tabu	drā	sucu-ma
Tahitian	tapu	toto	ngote
Maori	tapu	toto	ngote
Hawaiian	kapu	koko	omo
Marquesan	tapu	toto	omo

	Taboo	Blood	To Suck
Fijian	tabu	drā	sucu-ma
Tahitian	tapu	toto	ngote
Maori	tapu	toto	ngote
Hawaiian	kapu	koko	omo
Marquesan	tapu	toto	omo

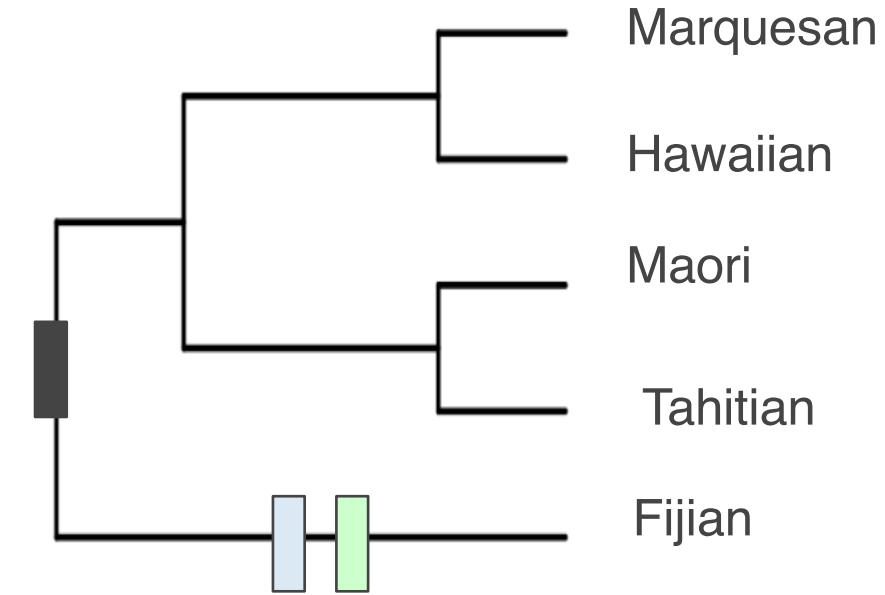
Fijian	1	1	0	1	0	0
Tahitian	1	0	1	0	1	0
Maori	1	0	1	0	1	0
Hawaiian	1	0	1	0	0	1
Marquesan	1	0	1	0	0	1



	Fijian	Tahitian	Maori	Hawaiian	Marquesan
Fijian	1	1	0	1	0
Tahitian	1	0	1	0	1
Maori	1	0	1	0	1
Hawaiian	1	0	1	0	0
Marquesan	1	0	1	0	1

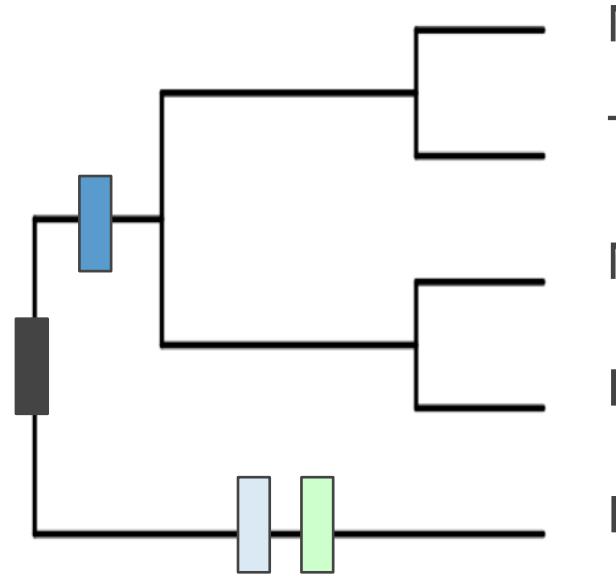


Length=3

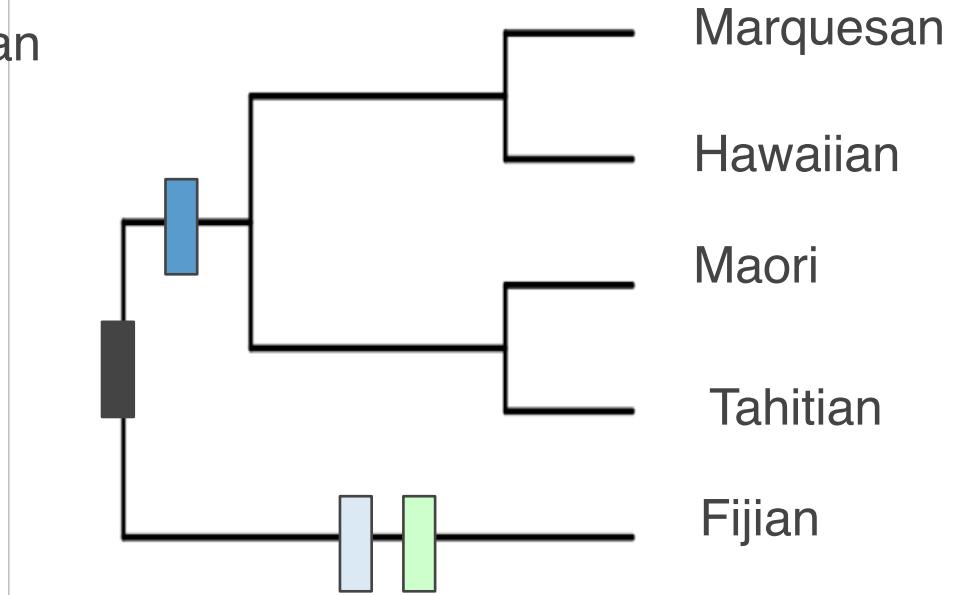


Length=3

Fijian	1	1	0	1	0	0
Tahitian	1	0	1	0	1	0
Maori	1	0	1	0	1	0
Hawaiian	1	0	1	0	0	1
Marquesan	1	0	1	0	0	1

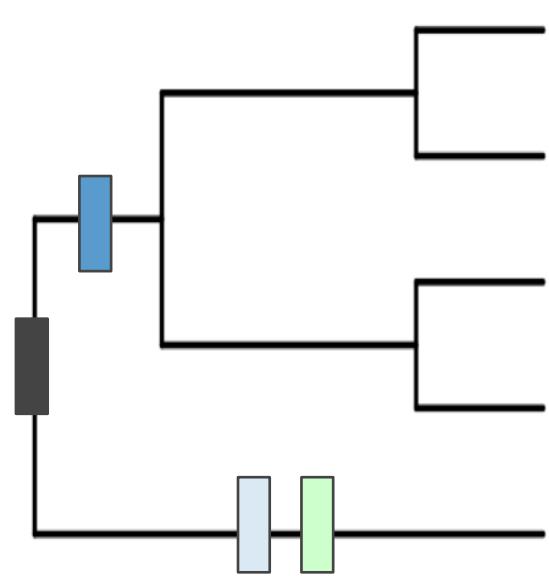


Length=4



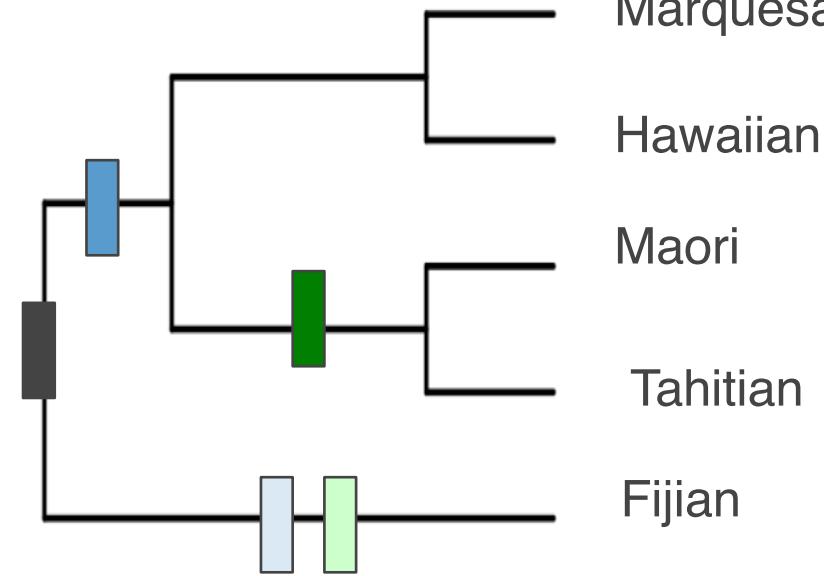
Length=4

Fijian	1	1	0	1	0	0
Tahitian	1	0	1	0	1	0
Maori	1	0	1	0	1	0
Hawaiian	1	0	1	0	0	1
Marquesan	1	0	1	0	0	1



Length=4

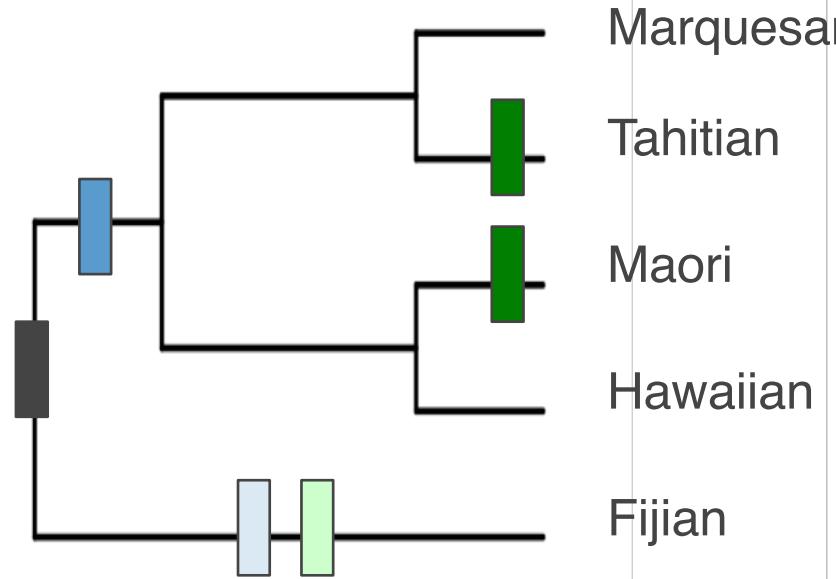
Marquesan
Tahitian
Maori
Hawaiian
Fijian



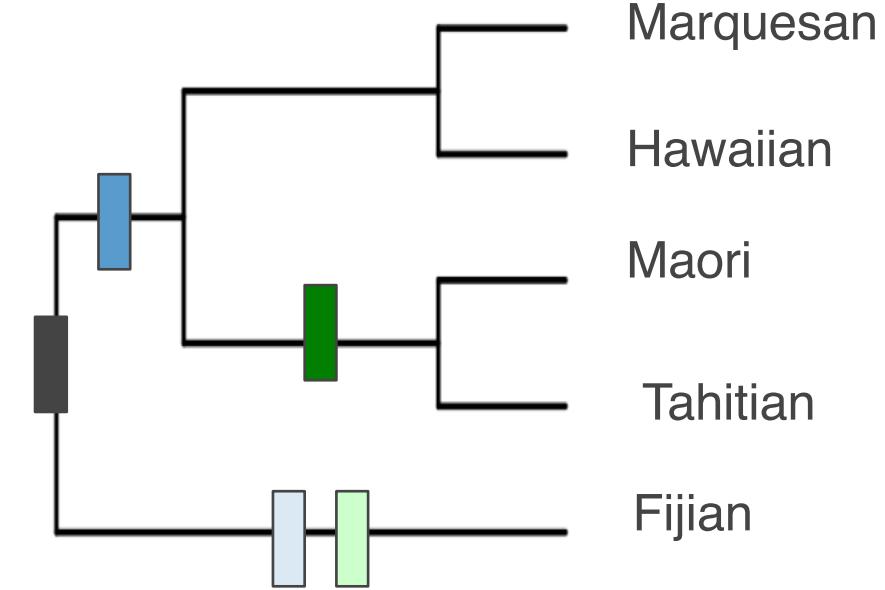
Length=5

Marquesan
Hawaiian
Maori
Tahitian
Fijian

	1	1	0	1	0	0
Fijian	1	1	0	1	0	0
Tahitian	1	0	1	0	1	0
Maori	1	0	1	0	1	0
Hawaiian	1	0	1	0	0	1
Marquesan	1	0	1	0	0	1

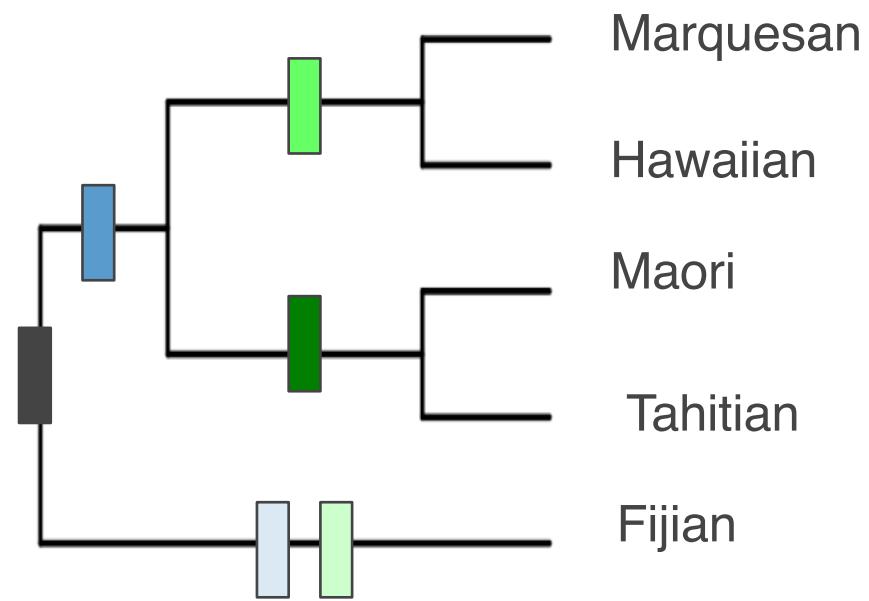
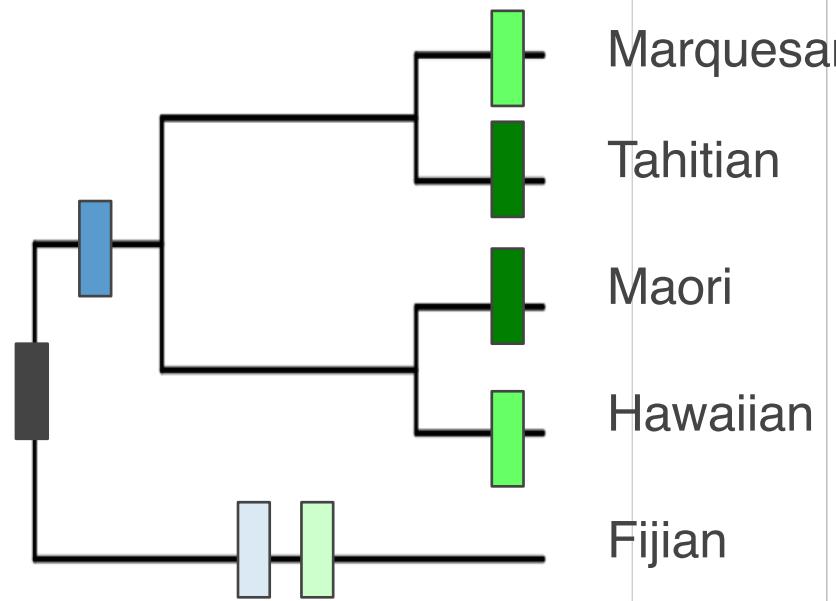


Length=6

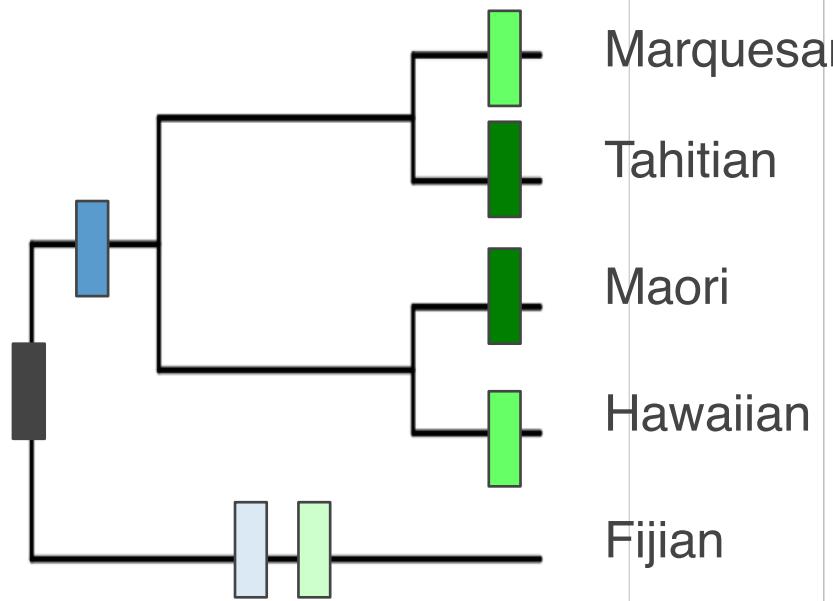


Length=5

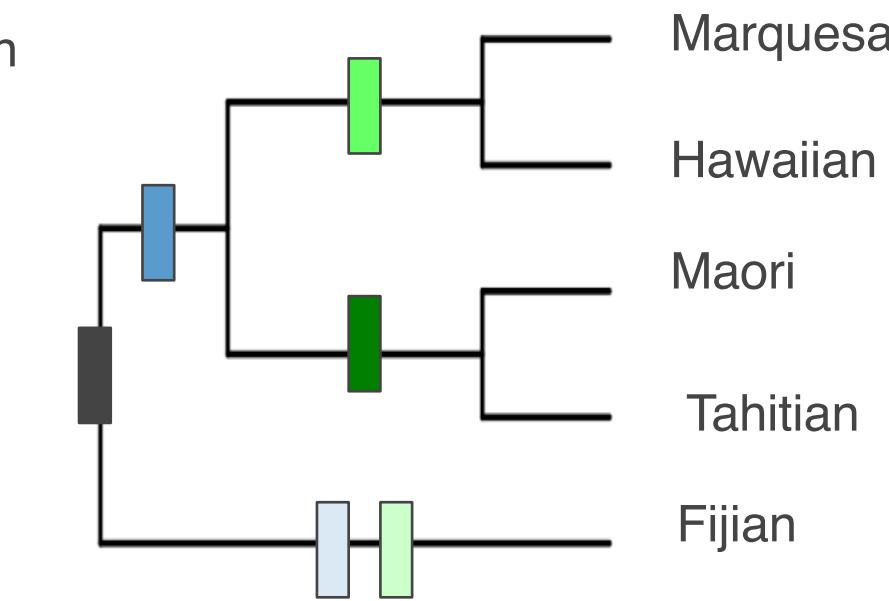
	1	1	0	1	0	0
Fijian	1	1	0	1	0	0
Tahitian	1	0	1	0	1	0
Maori	1	0	1	0	1	0
Hawaiian	1	0	1	0	0	1
Marquesan	1	0	1	0	0	1



	1	1	0	1	0	0
Fijian	1	1	0	1	0	0
Tahitian	1	0	1	0	1	0
Maori	1	0	1	0	1	0
Hawaiian	1	0	1	0	0	1
Marquesan	1	0	1	0	0	1



Length=8



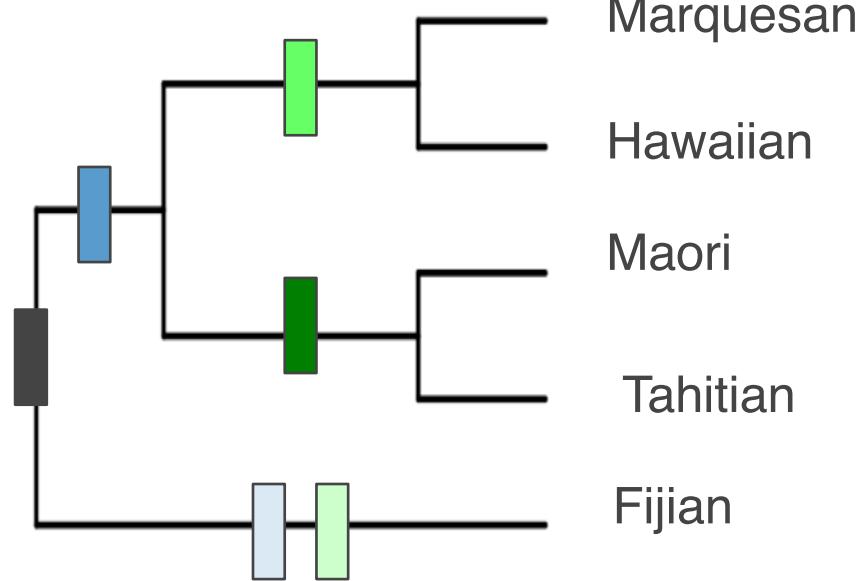
Length=6



	1	1	0	1	0	0
Fijian	1	1	0	1	0	0
Tahitian	1	0	1	0	1	0
Maori	1	0	1	0	1	0
Hawaiian	1	0	1	0	0	1
Marquesan	1	0	1	0	0	1



Innovations vs. Retentions



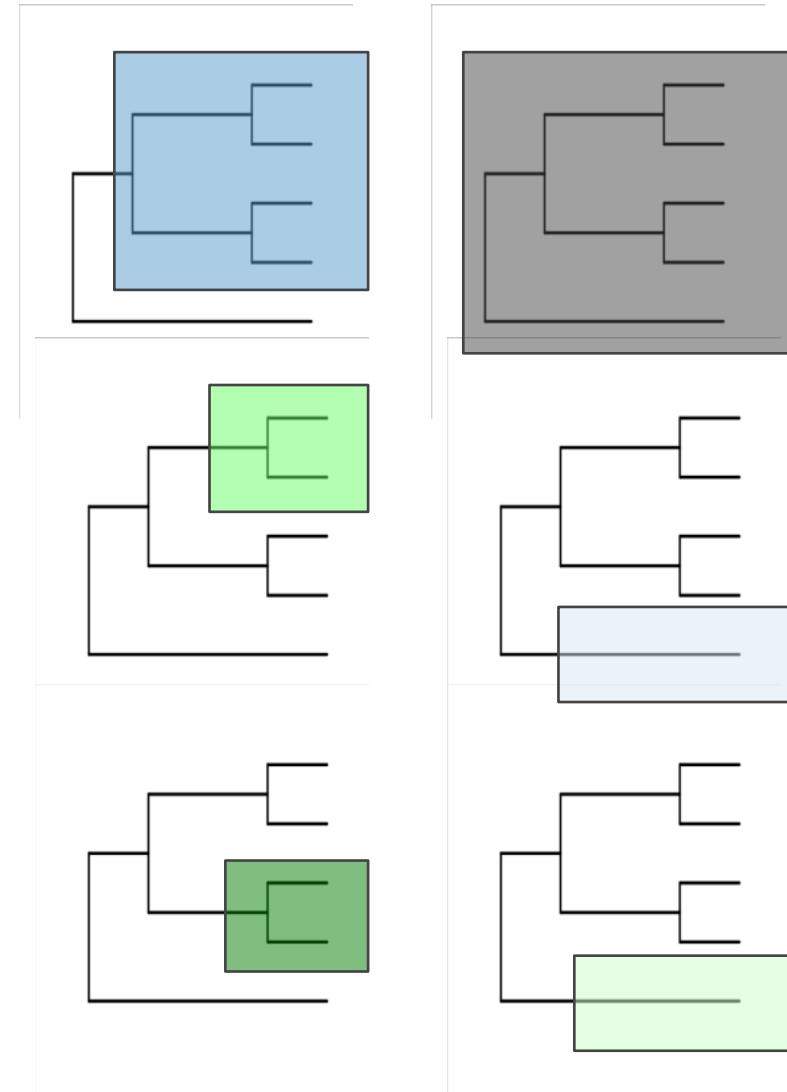
Marquesan

Hawaiian

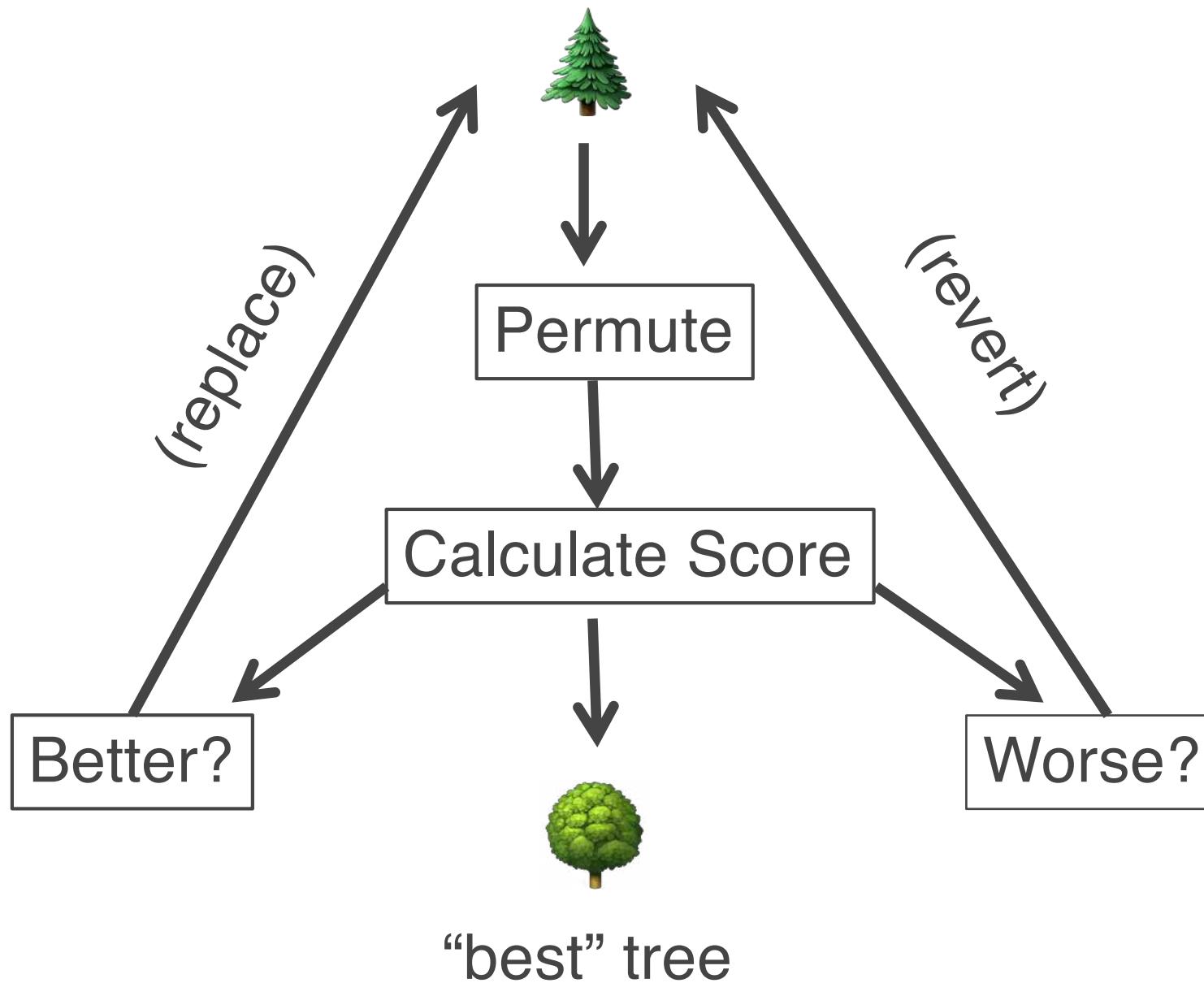
Maori

Tahitian

Fijian



Algorithm

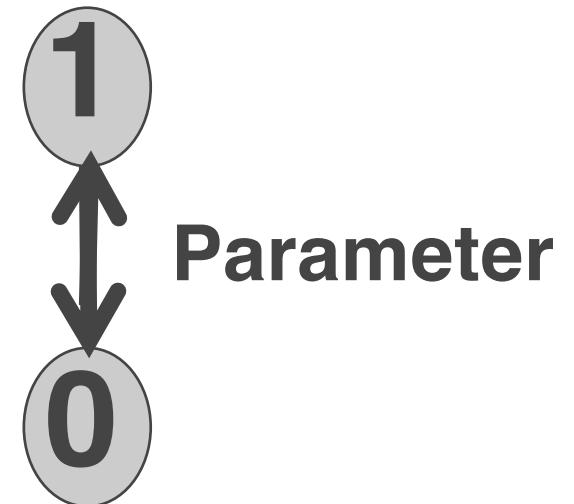


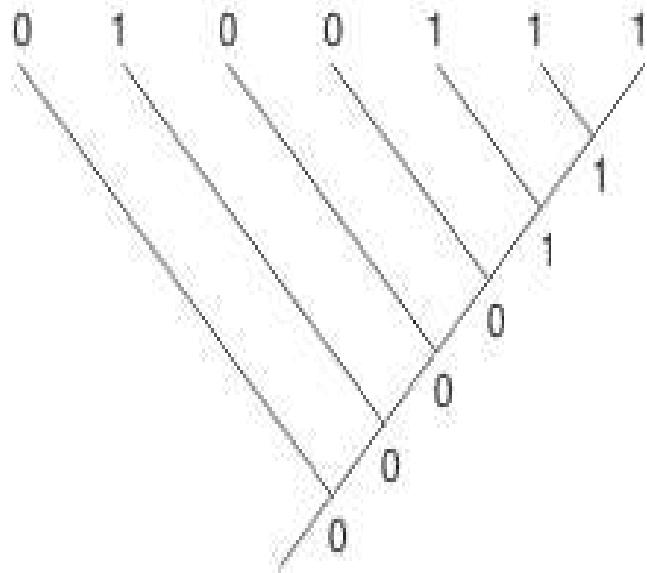
Maximum Likelihood

- Problem: Parsimony is a strong implicit assumption to make.
- Better: explicitly build in models of change

Maximum Likelihood

- Problem: Parsimony is a strong implicit assumption to make.
- Better: explicitly build in models of change
- => Maximum likelihood.
- Stochastic **model** of change
(=assumptions about how evolution works)
- **Likelihood** = fit of data to tree under a model.
 - Very small number = $\log(L_h)$
 - Closer to zero = better fit.





$$L(a) = P(0 \rightarrow 0|b_1) \times P(0 \rightarrow 0|b_2) \times P(1 \rightarrow 1|b_3) \times P(1 \rightarrow 0|b_4) \times \\ P(0 \rightarrow 0|b_5) \times P(0 \rightarrow 0|b_6) \times P(0 \rightarrow 0|b_7) \times P(0 \rightarrow 1|b_8) \times P(1 \rightarrow 1|b_9) \\ \times P(1 \rightarrow 1|b_{10}) \times P(1 \rightarrow 1|b_{11}) \times P(1 \rightarrow 1|b_{12})$$

$L_h =$
P of being in state 0, and
staying state 0 on branch 1.

x

P of being in state 0, and
staying state 0 on branch 2.

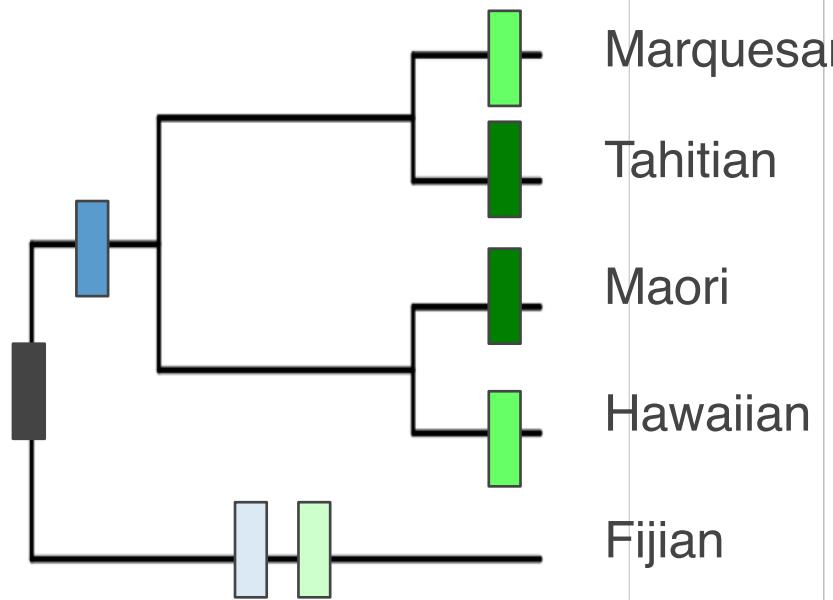
x

P of being in state 0, and
staying state 0 on branch 2.

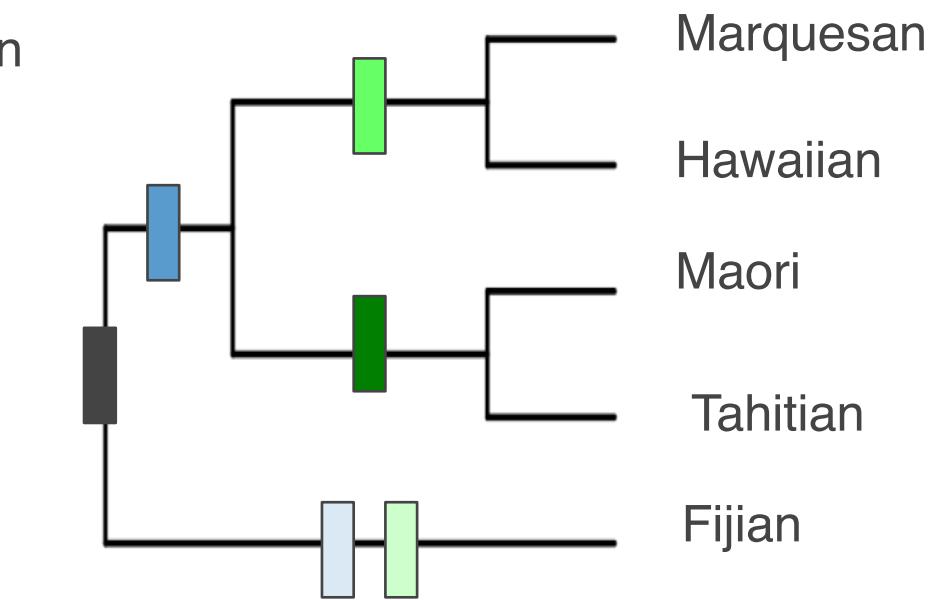
.... etc

Site Likelihood(a) = $\left(\begin{array}{c} 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{array} \right) \dots \times \dots \left(\begin{array}{c} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{array} \right) \dots \times \dots \left(\begin{array}{c} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{array} \right)$

Site Likelihood(a) = P(reconstruction 1) ... x ... P(reconstruction 5) ... x ... P(reconstruction n)



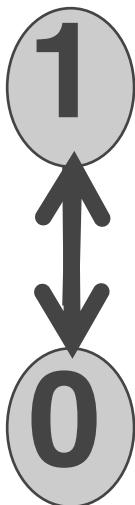
$\text{Ln}(L) = -14.804$



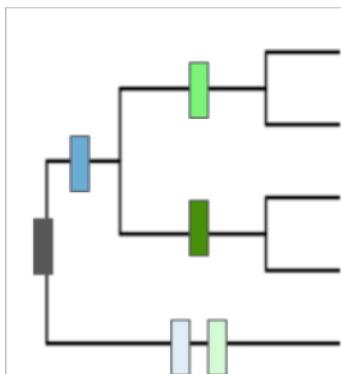
$\text{Ln}(L) = -12.007$ ←

Fijian	1	1	0	1	0	0
Tahitian	1	0	1	0	1	0
Maori	1	0	1	0	1	0
Hawaiian	1	0	1	0	0	1
Marquesan	1	0	1	0	0	1

Model Comparison with Lh



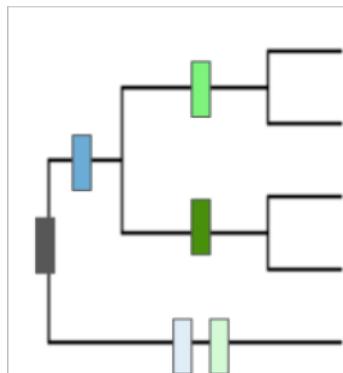
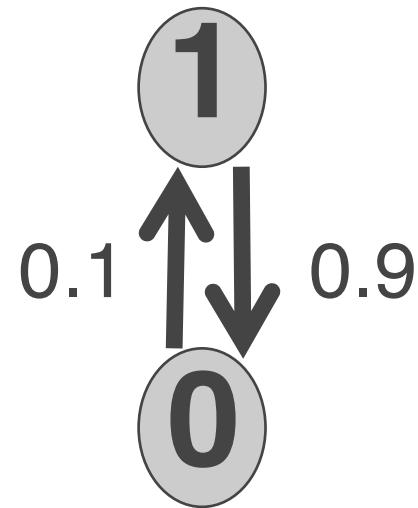
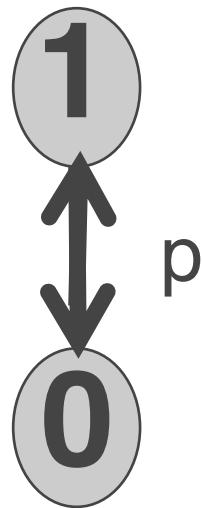
Parameter



Can we modify our
assumptions and get a **better**
explanation of the data?

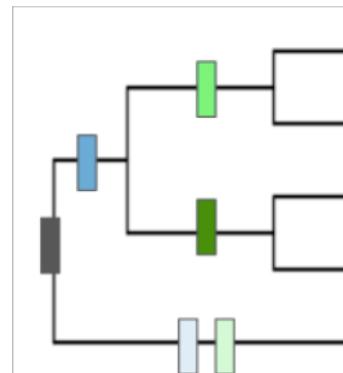
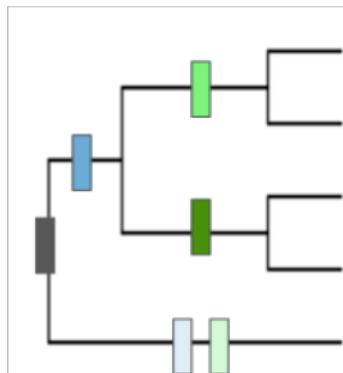
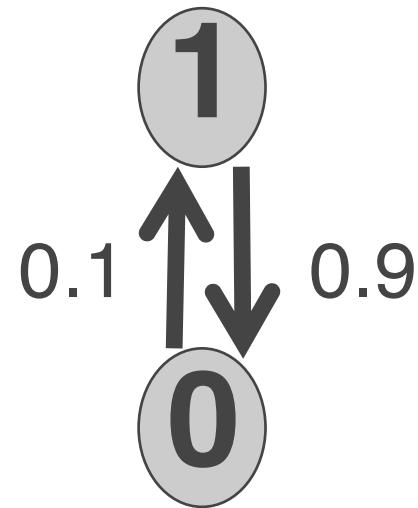
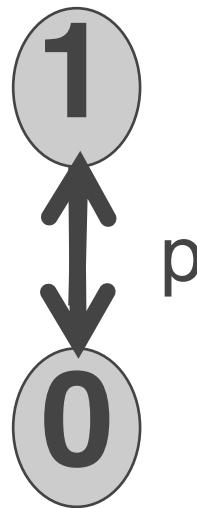
$$\ln(L) = -12.007$$

Models



$$\text{Ln}(L) = -12.007$$

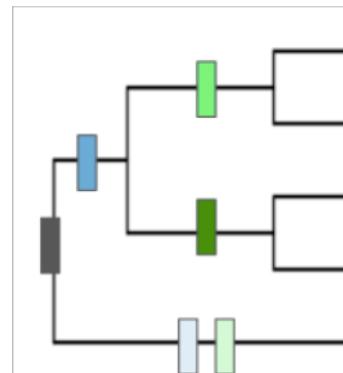
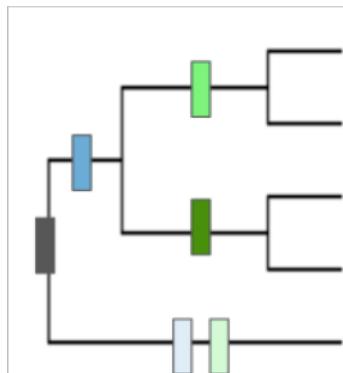
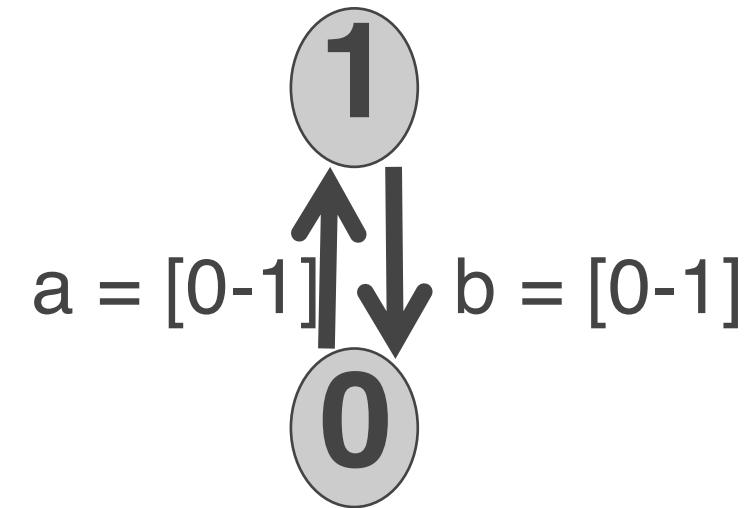
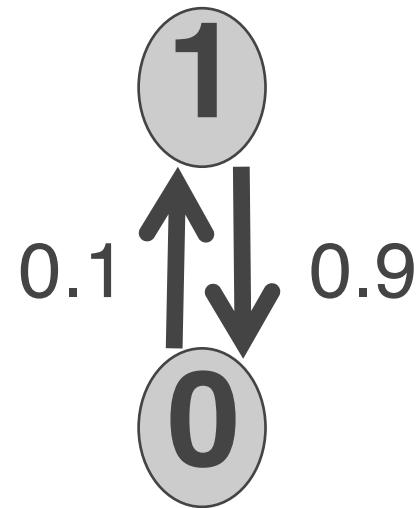
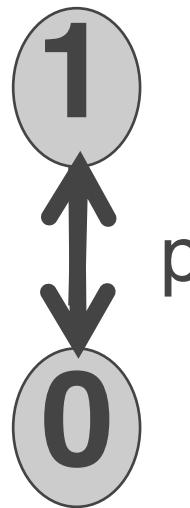
Models



$$\ln(L) = -12.007$$

$$\ln(L) = -11.310$$

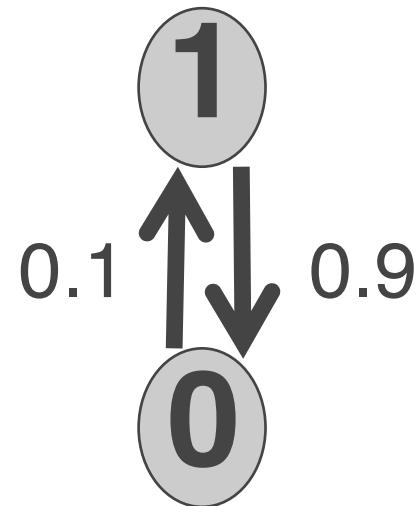
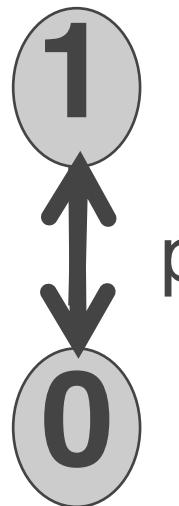
Models



$$\ln(L) = -12.007$$

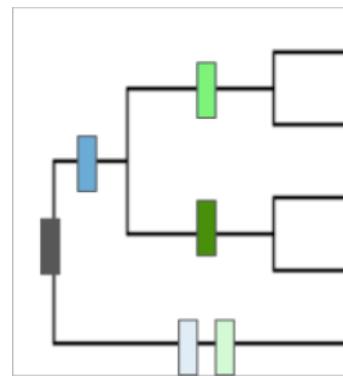
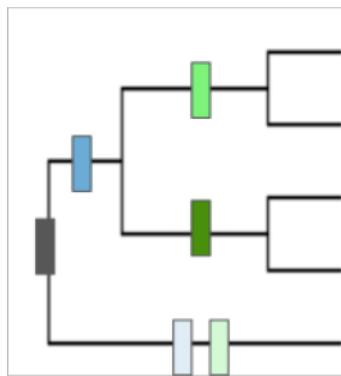
$$\ln(L) = -11.310$$

Models



$a = [0-1]$ $b = [0-1]$

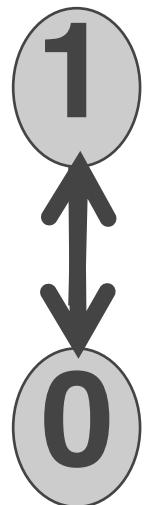
$a = 0.92, b = 0.08$



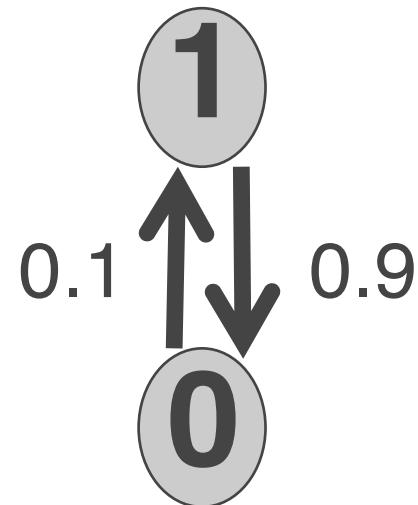
$$\ln(L) = -12.007$$

$$\ln(L) = -11.310$$

Models

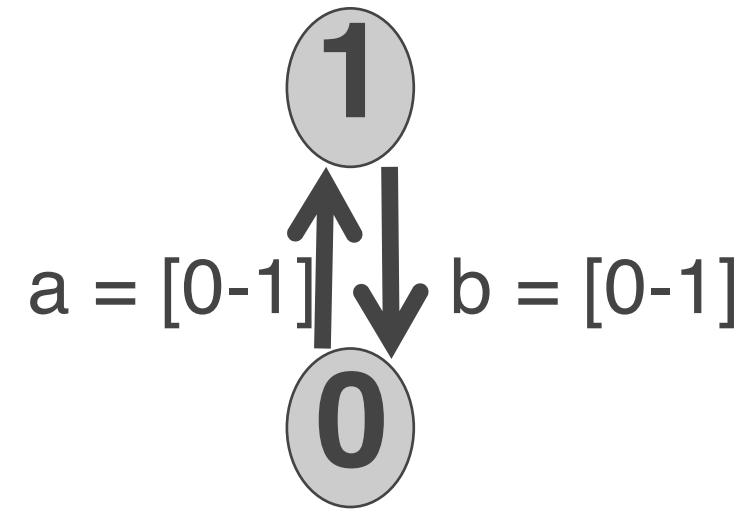


p



0.1

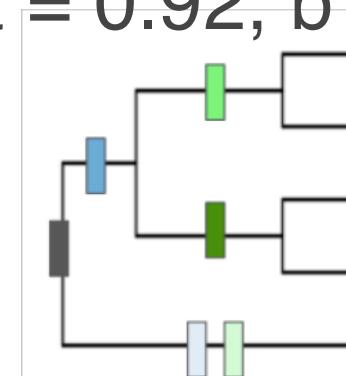
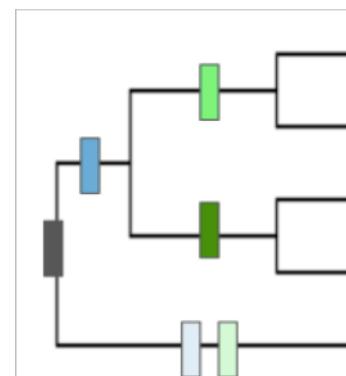
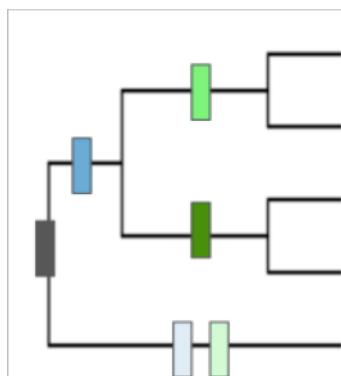
0.9



a = [0-1]

b = [0-1]

a = 0.92, b = 0.08

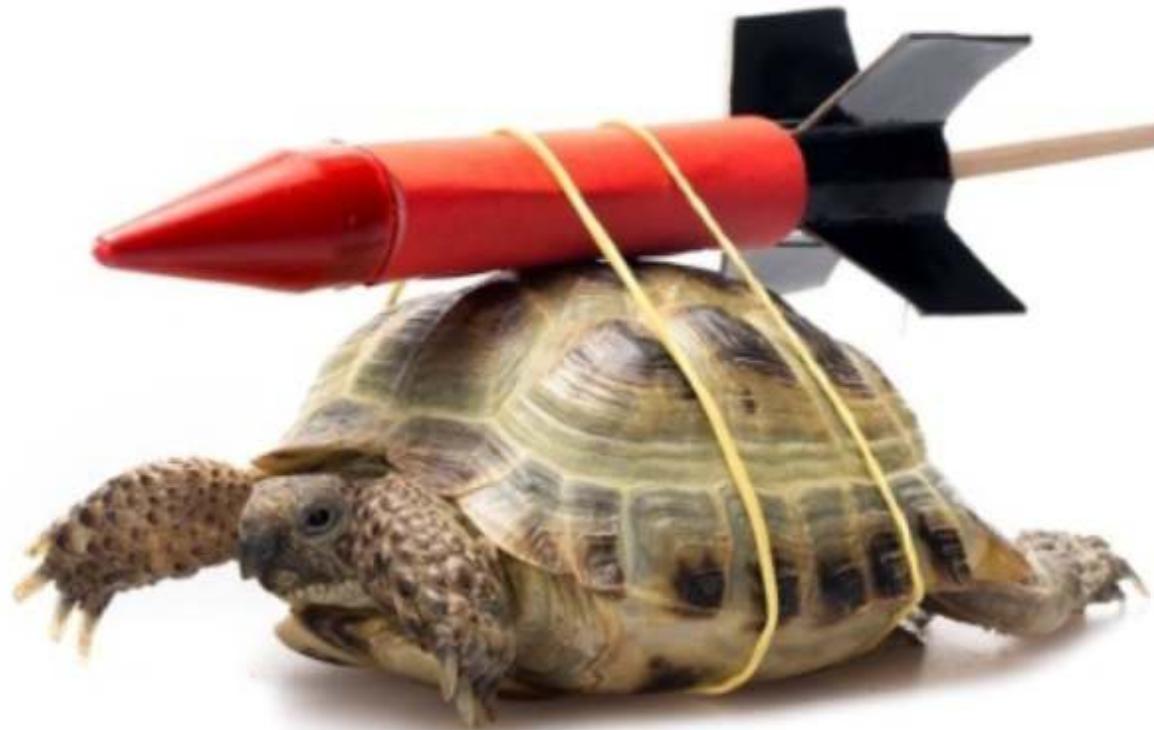


$$\ln(L) = -12.007$$

$$\ln(L) = -11.310$$

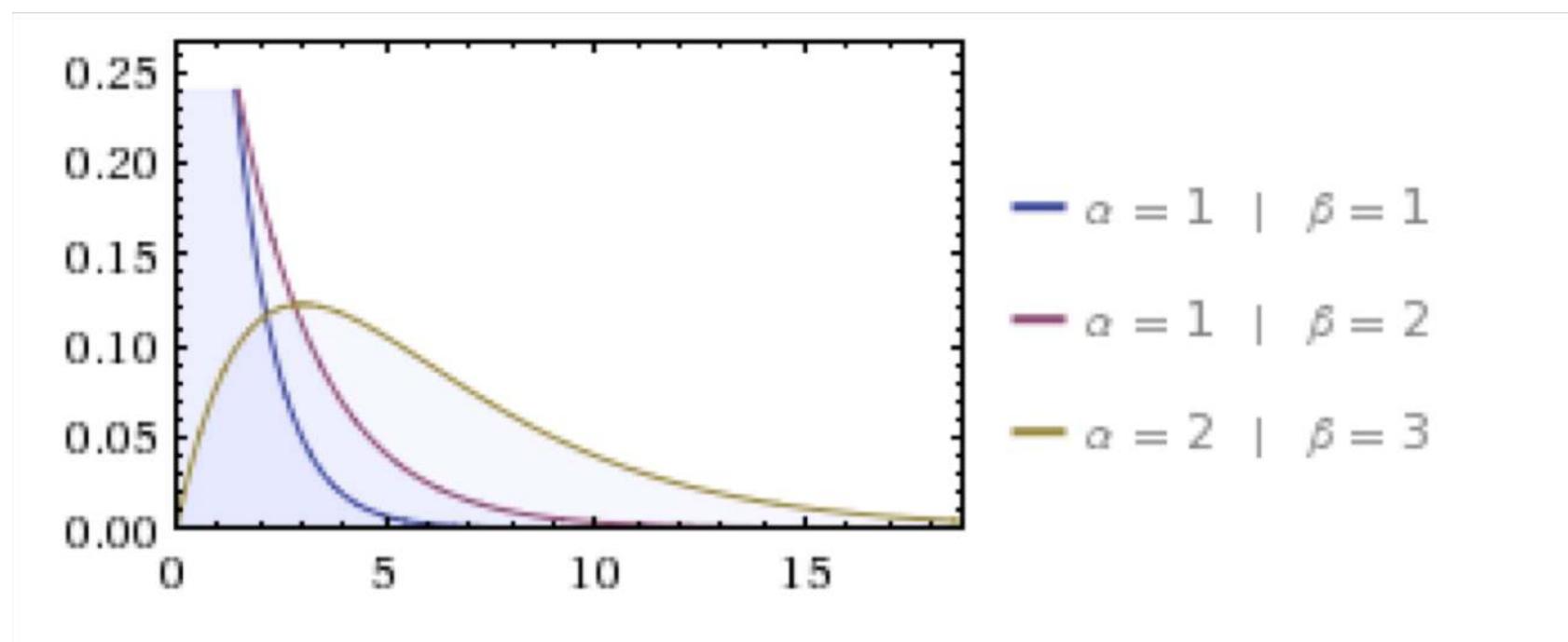
$$\ln(L) = -9.072$$

But Rates Vary!



Rate Variation - Gamma

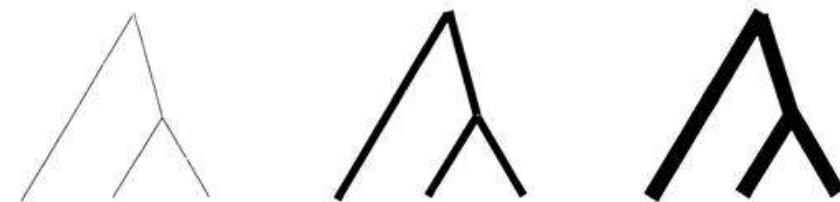
- “Site-Specific Rates”: Rates vary across **sites**
- Gamma Distribution: One parameter, α , controls the shape
- Discretized into n categories and Characters assigned a category
- Estimate the best value for α using Lh.



Rate Variation - Covarion

- Rates change across **phylogeny**
- Each cognate set is either “on” or “off”
 - “on” = evolves according to normal rates
 - “off” = evolves according to slow rate (often 0)
- Analysis estimate when sites are “on” and “off” and switching rates.
- = Each site can evolve at different rates at different times on the tree.

Among-Site Rate Variation



Site-Specific Rate Variation



Gautier 2001

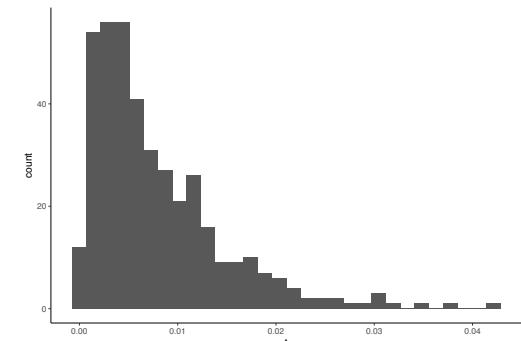
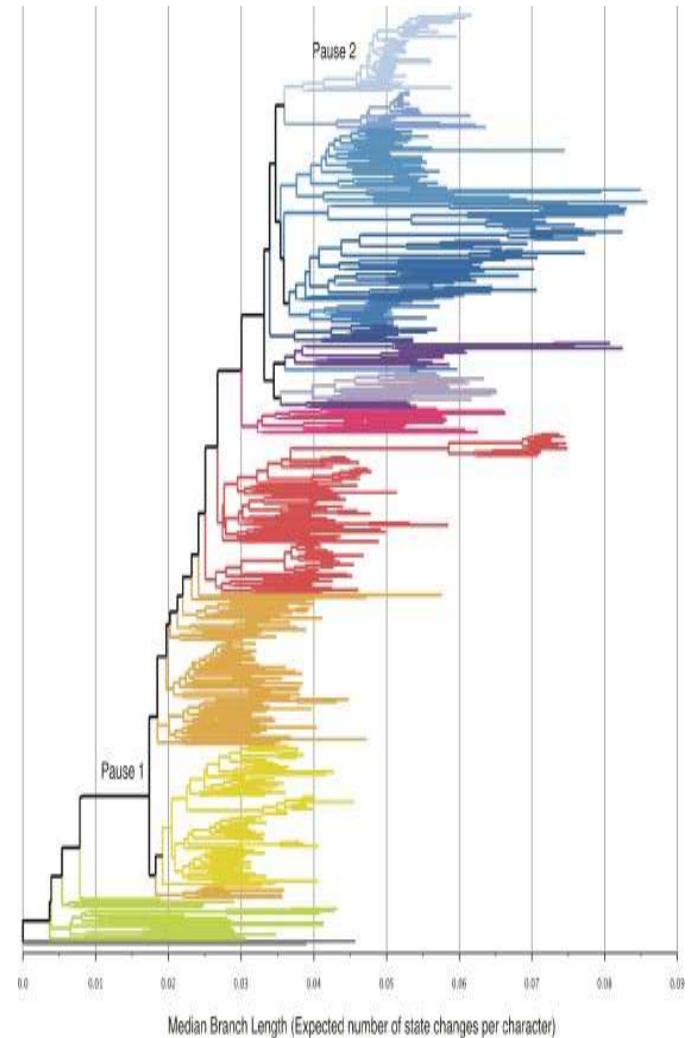


“Linguists don’t do dates.”

Phylogenetic Dating

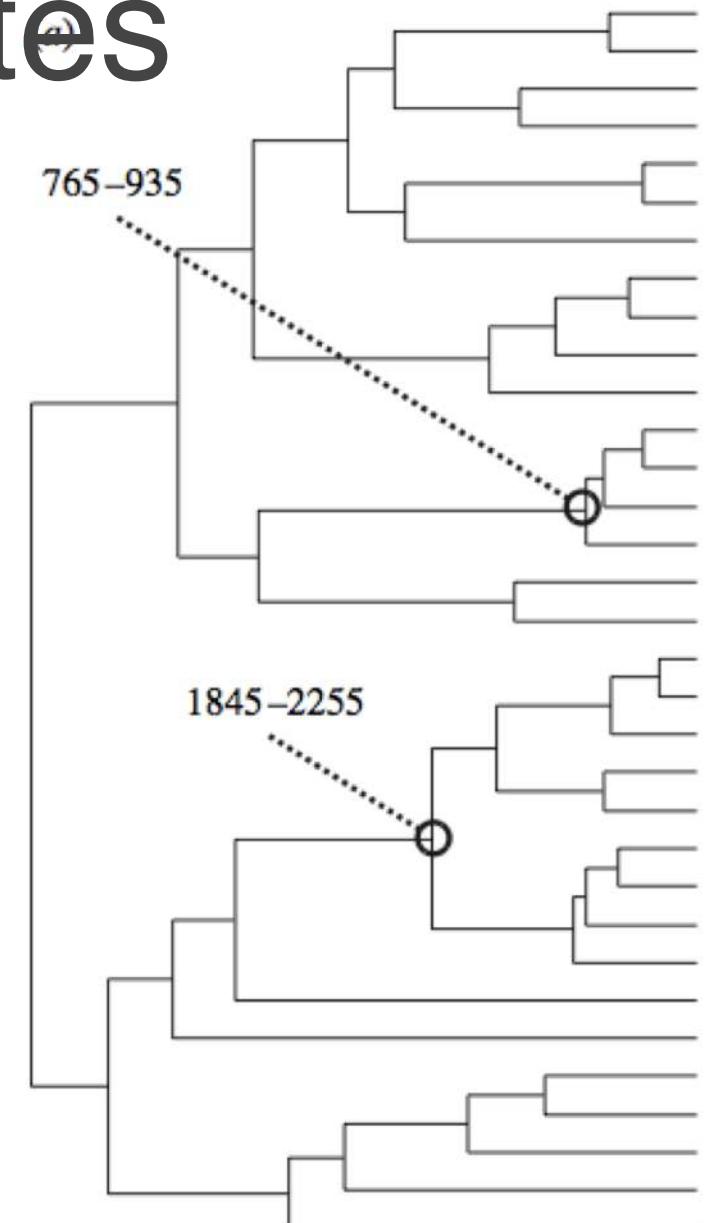
ML estimates amount of change along lineage
= (number of changes per character)

Not a global retention rate but a **per-lineage** estimate of the amount of change.



Convert Rates to Dates

- (pre)historical information to **calibrate nodes**
 - e.g. Archaeology suggests initial settlement was..
 - e.g. Historical evidence says that X and Y were separate by...
- Smooth rates over these calibrations



Strict Clock



- One rate for all languages.
- No variation
- \sim = glottochronology

Strict Clock



- One rate for all languages.
- No variation
- \sim = glottochronology

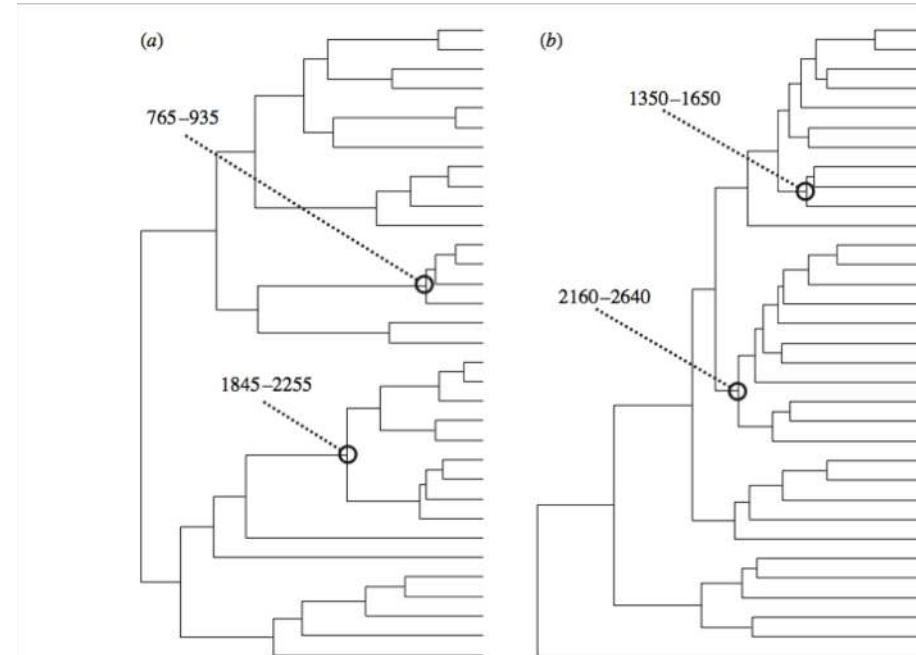
Relaxed Clock



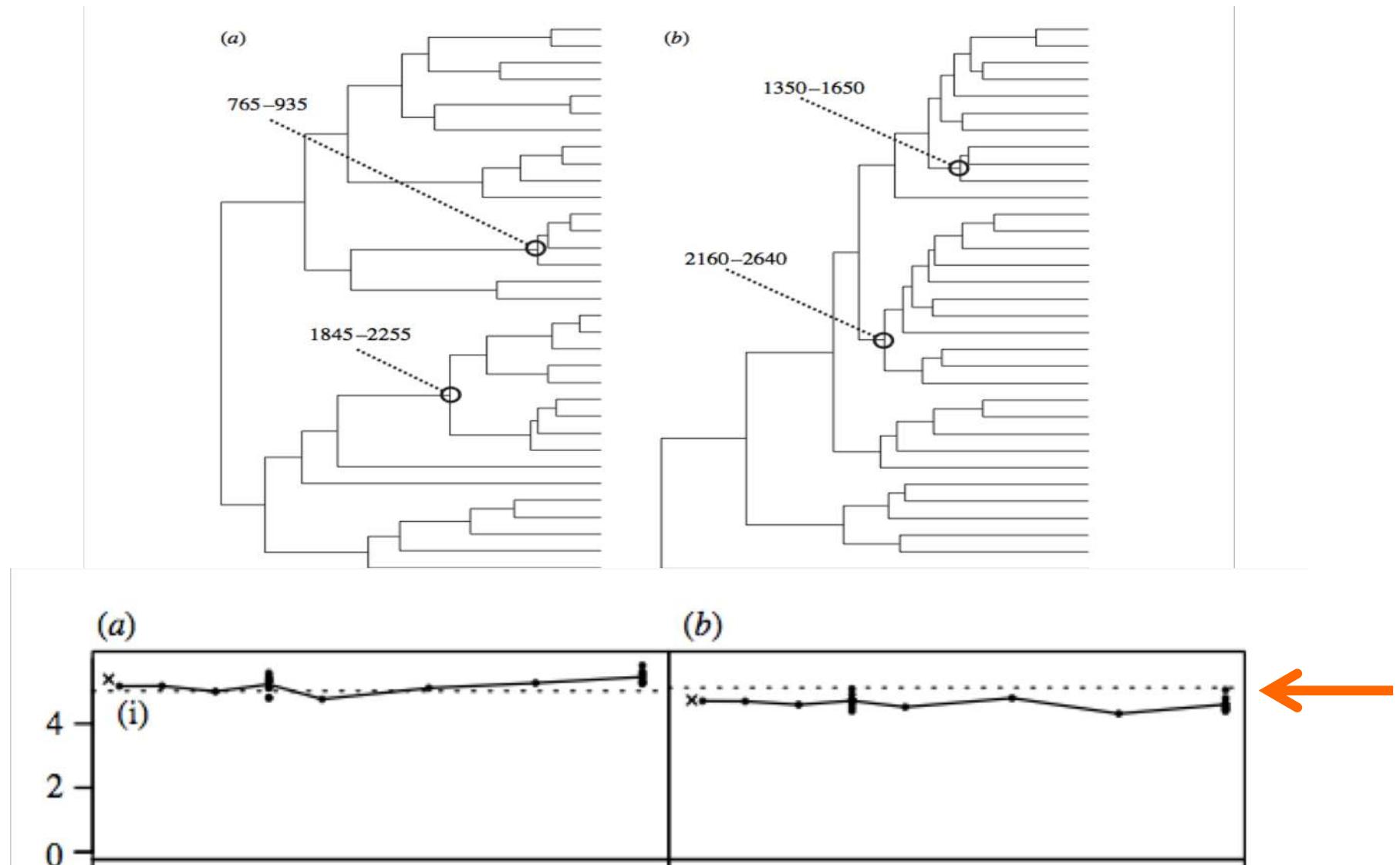
- Estimate distribution of *branch* rates from data
- Rates drawn from a parametric distribution estimated from the data
 - **LogNormal**, Exponential
- Allows rate to vary across branches (each branch could have own rate)

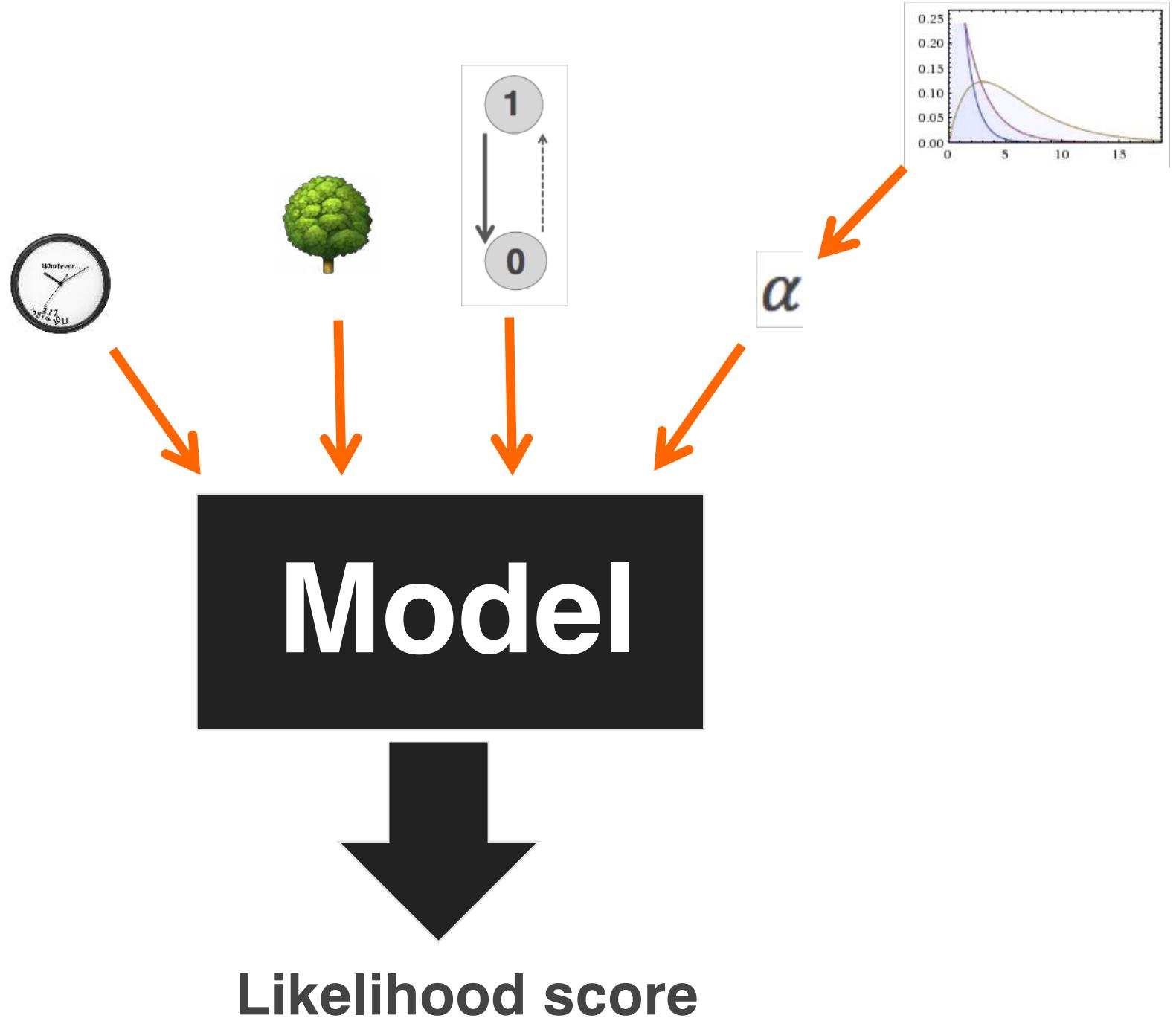
Does it work?

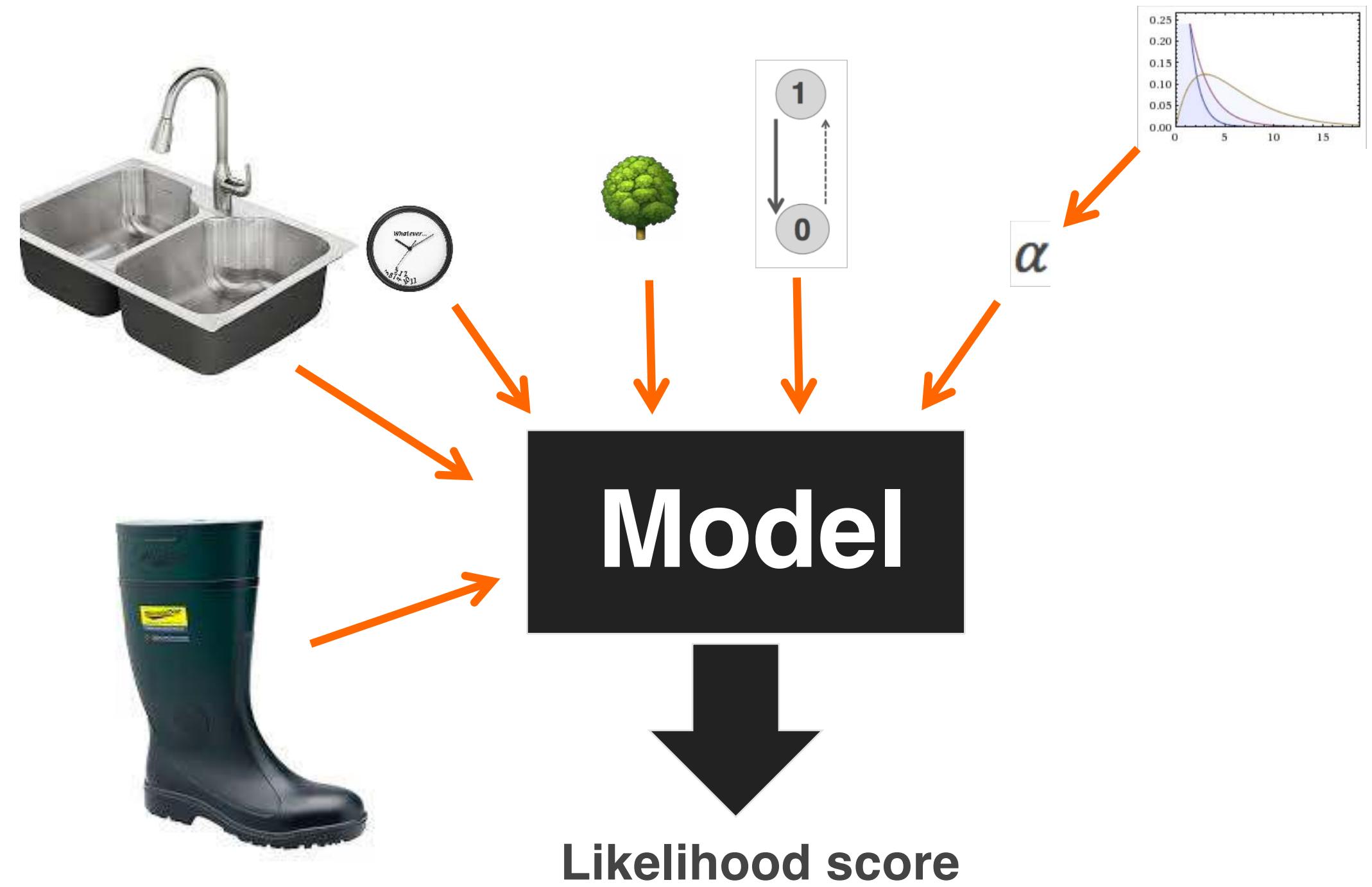
- Take fixed trees, with known dates.
- Simulate a new dataset.
- Analyse data with two calibrations
- Do we get the right answer?
- Repeat x 1000

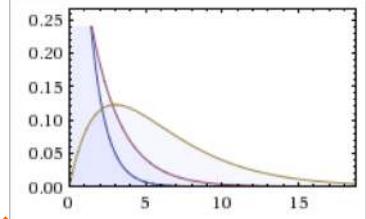
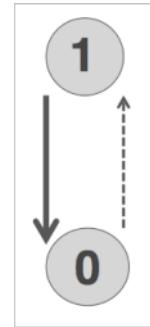


Does it work?

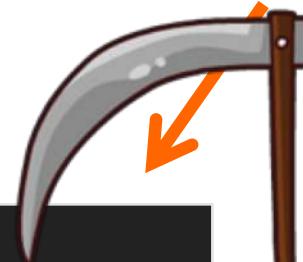








α

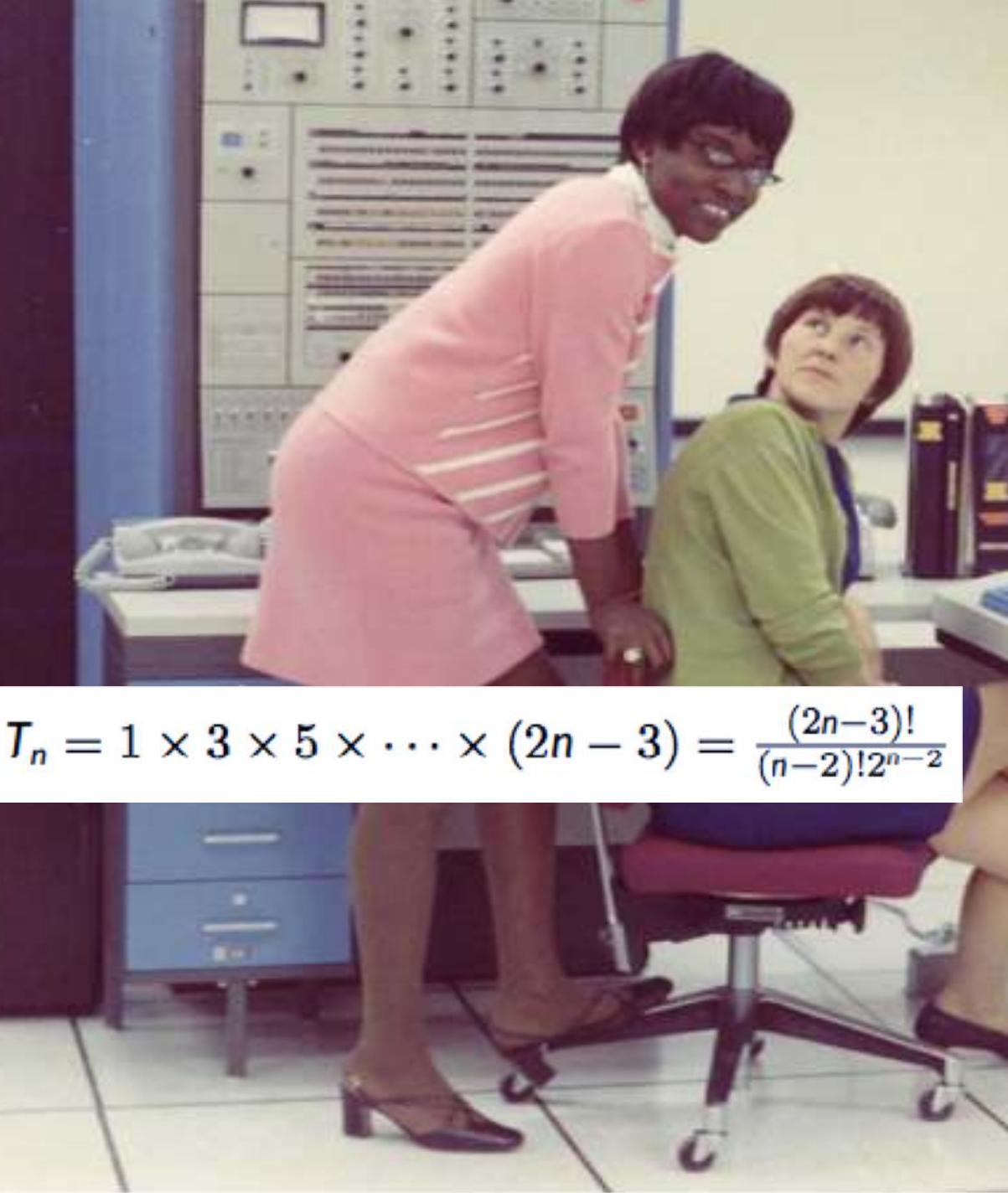


Model

OVER-PARAMETERISED

????????????





$$T_n = 1 \times 3 \times 5 \times \cdots \times (2n - 3) = \frac{(2n-3)!}{(n-2)!2^{n-2}}$$

n	Bifurcating trees
1	1
2	1
3	3
4	15
5	105
6	945
7	10,395
8	135,135
9	2,027,025
10	34,459,425
11	654,729,075
12	13,749,310,575
13	316,234,143,225
14	7,905,853,580,625
15	213,458,046,676,875
16	6,190,283,353,629,375
17	191,898,783,962,510,625
18	6,332,659,870,762,850,625
19	221,643,095,476,699,771,875
20	8,200,794,532,637,891,559,375
21	319,830,986,772,877,770,815,625
22	13,113,070,457,687,988,603,440,625

Need to handle complexity

- ML and MP give a “**point**” estimate.
- But a single tree is not enough.
 - Reality is complicated.
 - Need to estimate uncertainty around that estimate.
 - “confidence intervals” = how confident can I about my estimate.





Bayesian methods

- Extend ML methods to explicitly account for uncertainty

$$\Pr(H|D) = \frac{\Pr(H) \Pr(D|H)}{\Pr(D)}$$

General intro: McElreath “Statistical Rethinking” (<http://xcelab.net/rm/statistical-rethinking/>)

Phylogenetics: Brown “The State of Bayesian Phylogenetics: Bayes for the Uninitiated” (<https://tinyurl.com/ly2cvfd>)



Bayesian methods

- Extend ML methods to explicitly account for uncertainty

$$\Pr(H|D) = \frac{\Pr(H) \Pr(D|H)}{\Pr(D)}$$

Lh

A blue bracket is drawn under the term $\Pr(D|H)$. Above the bracket, the letters "Lh" are written in blue, likely indicating a label or a specific context for that term.

General intro: McElreath “Statistical Rethinking” (<http://xcelab.net/rm/statistical-rethinking/>)

Phylogenetics: Brown “The State of Bayesian Phylogenetics: Bayes for the Uninitiated” (<https://tinyurl.com/ly2cvfd>)



Bayesian methods

- Extend ML methods to explicitly account for uncertainty

$$\Pr(H|D) = \frac{\text{prior } \Pr(H) \text{ Pr}(D|H)}{\Pr(D)}$$

The equation shows the Bayesian formula for the posterior probability $\Pr(H|D)$. The numerator is the product of the prior probability $\Pr(H)$ and the likelihood $\Pr(D|H)$. The denominator is the total probability $\Pr(D)$. The words "prior" and "Lh" are written above their respective terms in orange and blue, with curly braces indicating they are parts of the numerator.

General intro: McElreath “Statistical Rethinking” (<http://xcelab.net/rm/statistical-rethinking/>)

Phylogenetics: Brown “The State of Bayesian Phylogenetics: Bayes for the Uninitiated” (<https://tinyurl.com/ly2cvfd>)



Bayesian methods

- Extend ML methods to explicitly account for uncertainty

$$\Pr(H|D) = \frac{\Pr(H) \Pr(D|H)}{\Pr(D)}$$

prior **Lh**

Prior Pr that data is true

The diagram illustrates Bayes' theorem. At the top, the words "prior" and "Lh" (likelihood) are written in orange and blue respectively. Below them, the formula $\Pr(H|D) = \frac{\Pr(H) \Pr(D|H)}{\Pr(D)}$ is shown. The term $\Pr(H)$ is bracketed with an orange bracket labeled "prior". The term $\Pr(D|H)$ is bracketed with a blue bracket labeled "Lh". The denominator $\Pr(D)$ is bracketed with a green bracket labeled "Prior Pr that data is true".

General intro: McElreath “Statistical Rethinking” (<http://xcelab.net/rm/statistical-rethinking/>)

Phylogenetics: Brown “The State of Bayesian Phylogenetics: Bayes for the Uninitiated” (<https://tinyurl.com/ly2cvfd>)



Bayesian methods

- Extend ML methods to explicitly account for uncertainty

$$\Pr(H|D) = \frac{\text{prior} \quad \Pr(H) \quad \text{Lh} \quad \Pr(D|H)}{\Pr(D)}$$

(Pr hypothesis given data)

Prior Pr that data is true

The diagram illustrates Bayes' theorem with the following components:

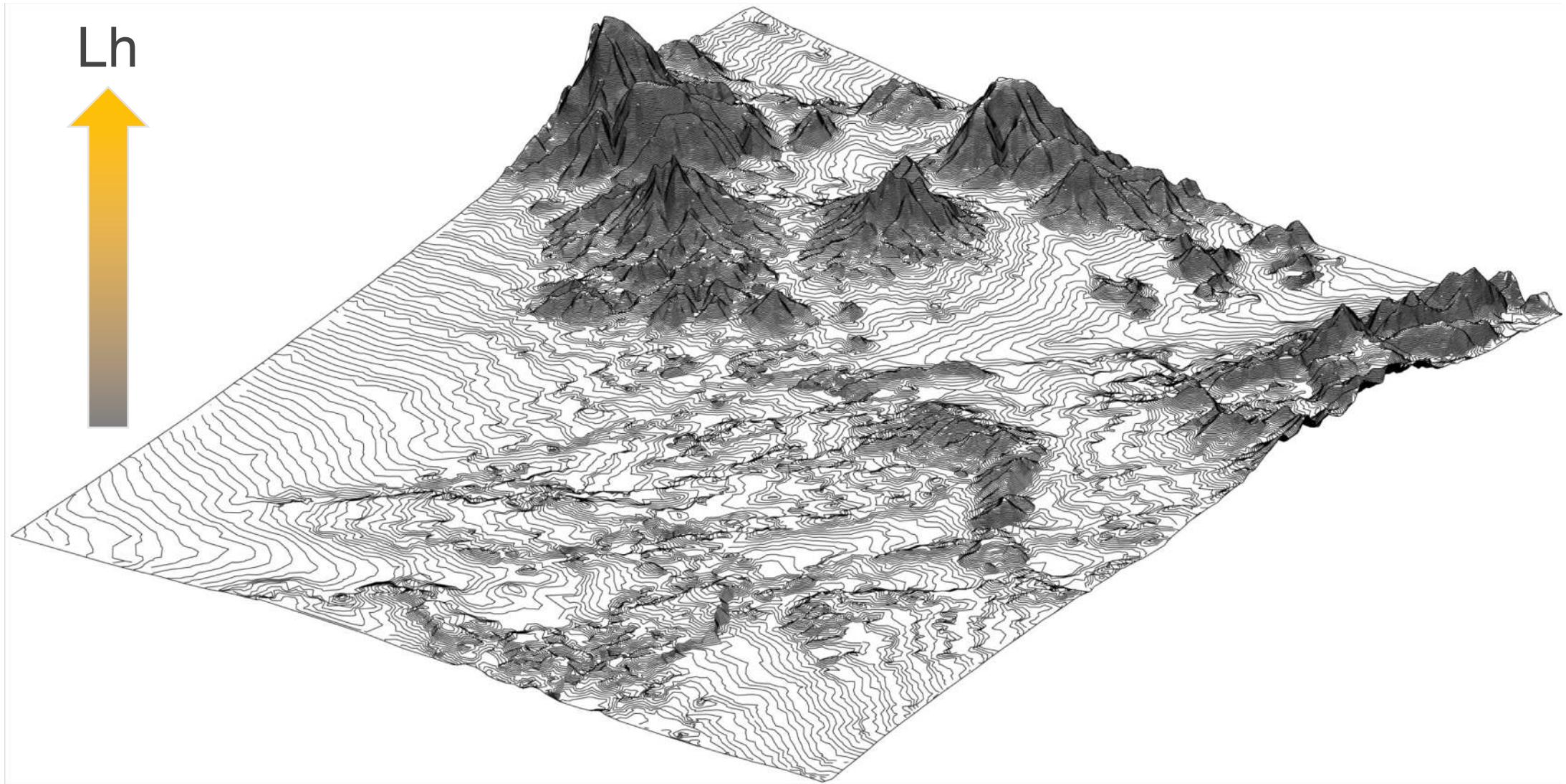
- Posterior**: $\Pr(H|D)$, labeled in purple.
- prior**: $\Pr(H)$, labeled in orange.
- Lh**: $\Pr(D|H)$, labeled in blue.
- Evidence**: $\Pr(D)$, labeled in green at the bottom.

Brackets group the terms: the numerator is $\text{prior} \quad \Pr(H) \quad \text{Lh} \quad \Pr(D|H)$ and the denominator is $\Pr(D)$.

General intro: McElreath “Statistical Rethinking” (<http://xcelab.net/rm/statistical-rethinking/>)

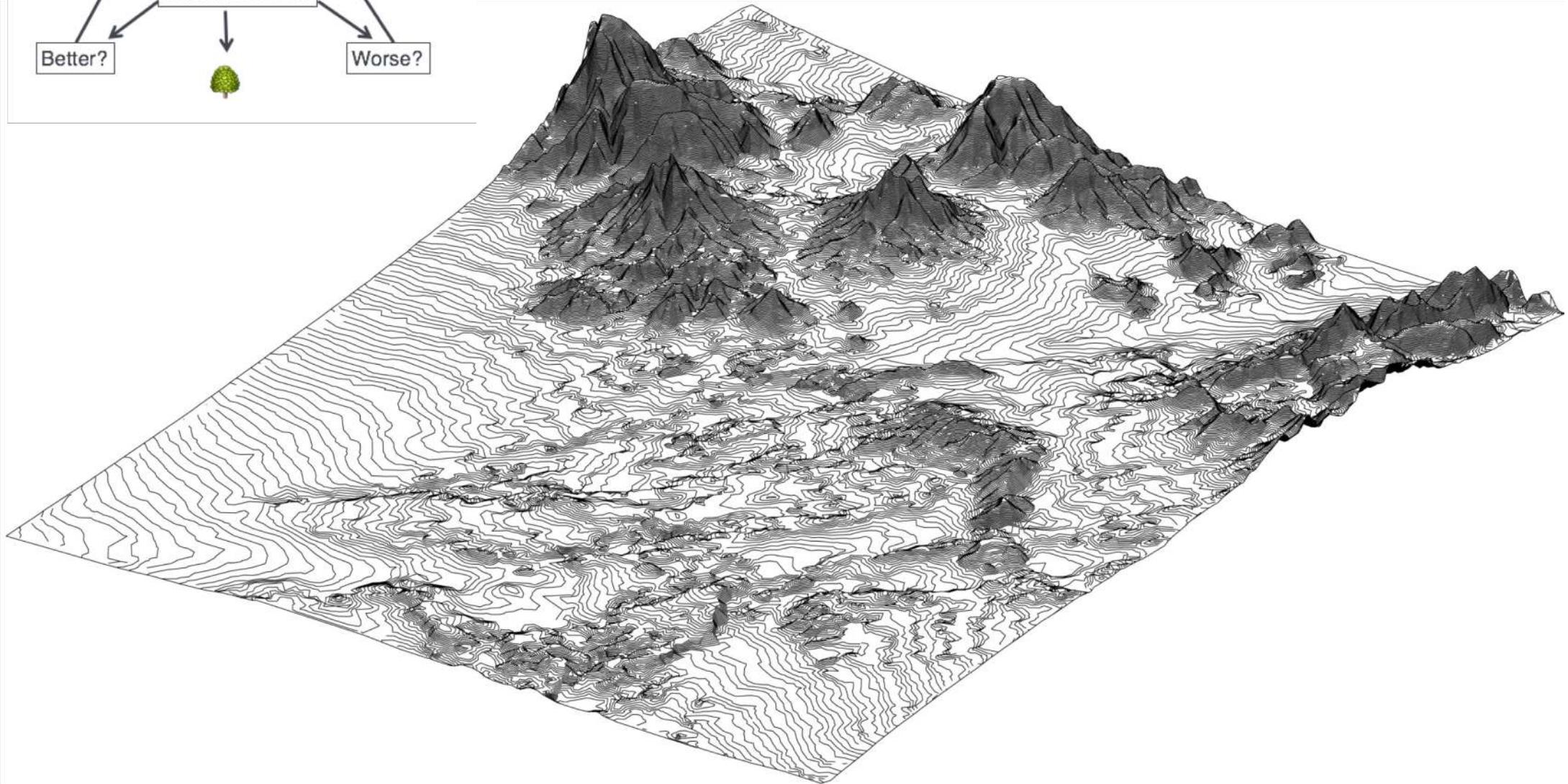
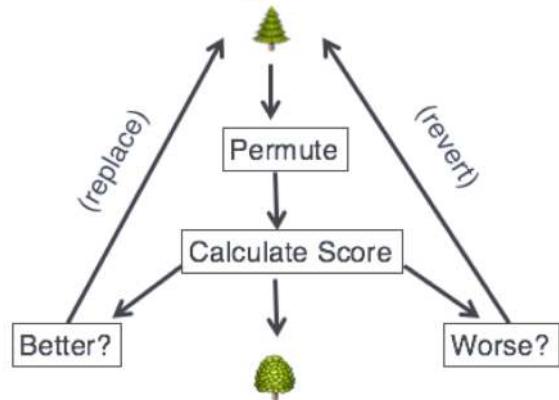
Phylogenetics: Brown “The State of Bayesian Phylogenetics: Bayes for the Uninitiated” (<https://tinyurl.com/ly2cvfd>)

Treespace



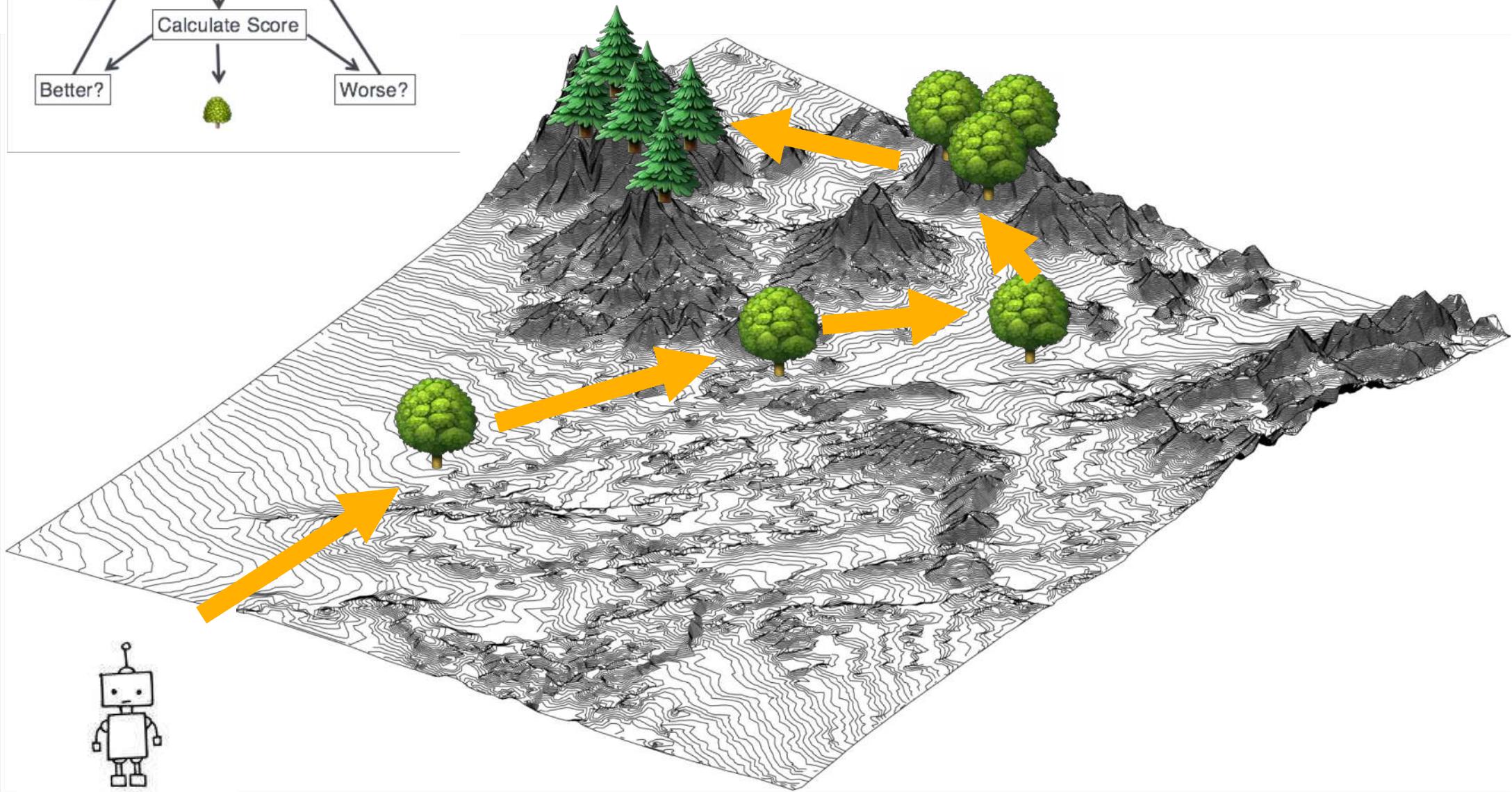
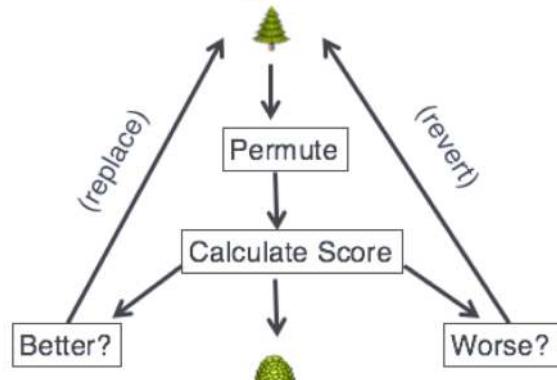
MCMC

Algorithm



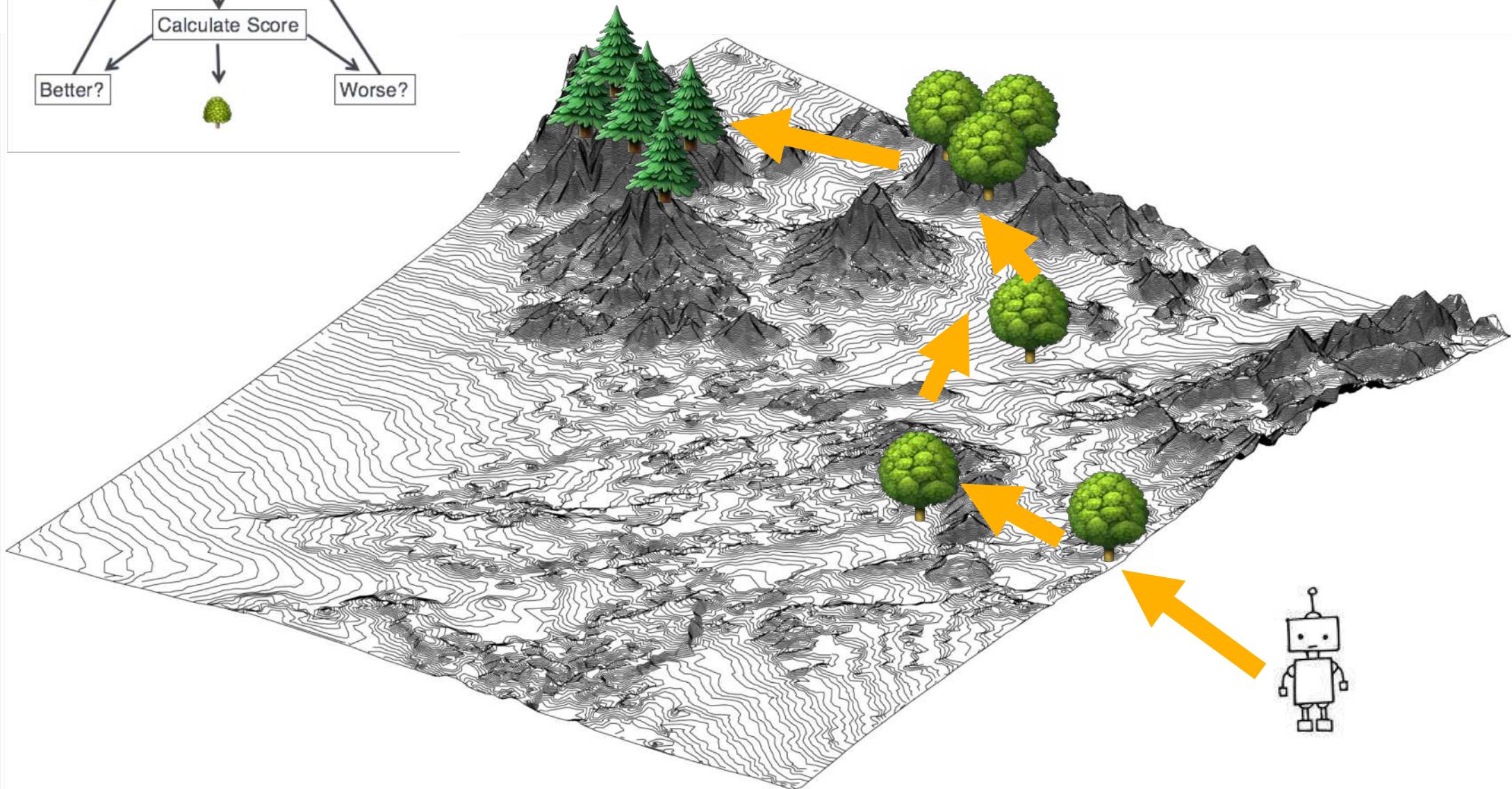
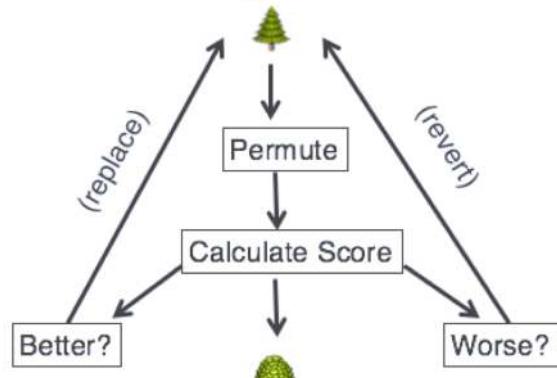
MCMC

Algorithm

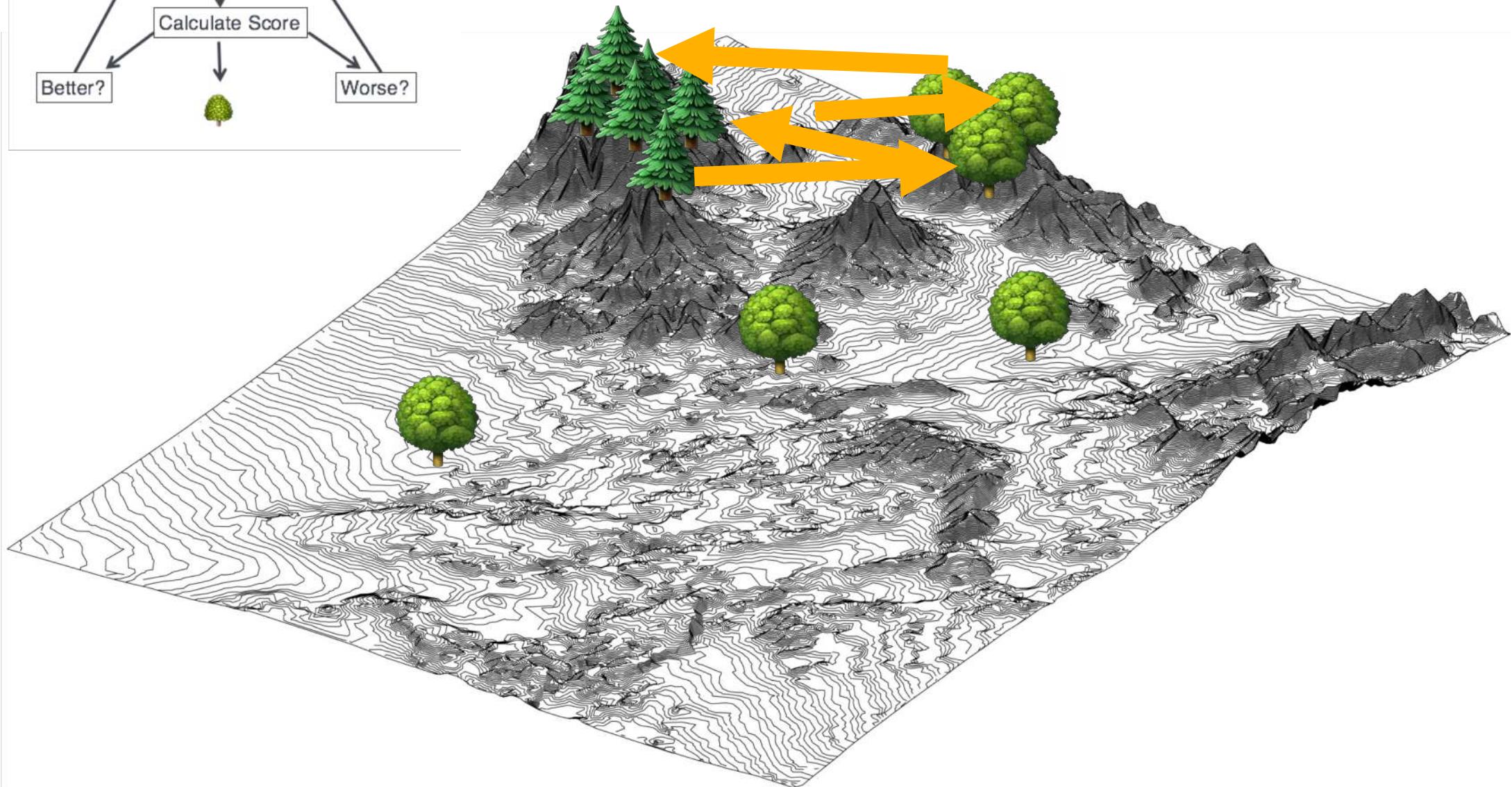
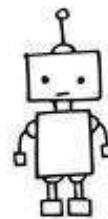
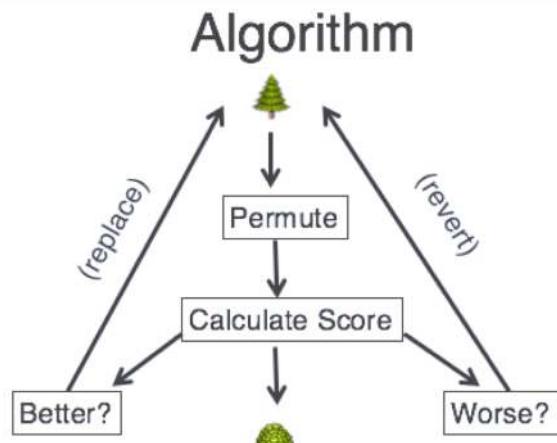


MCMC

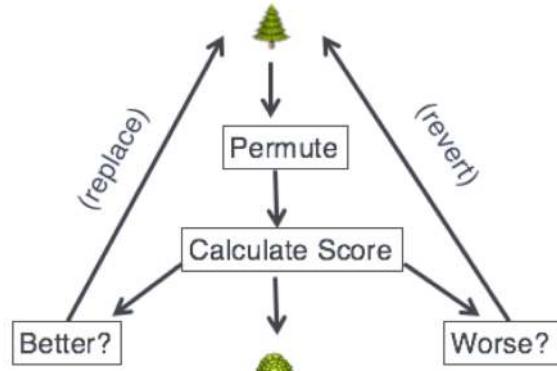
Algorithm



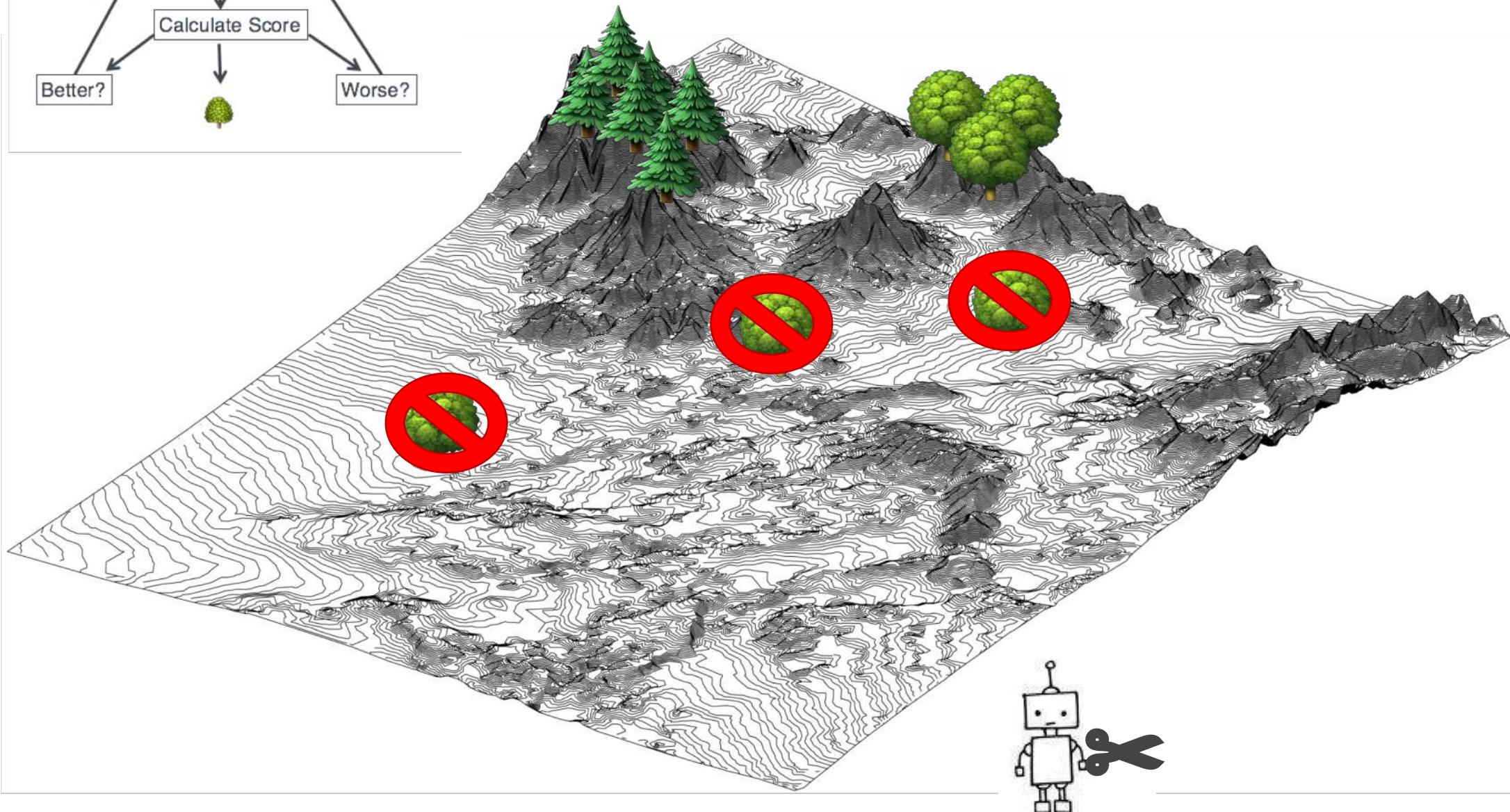
“stationarity”



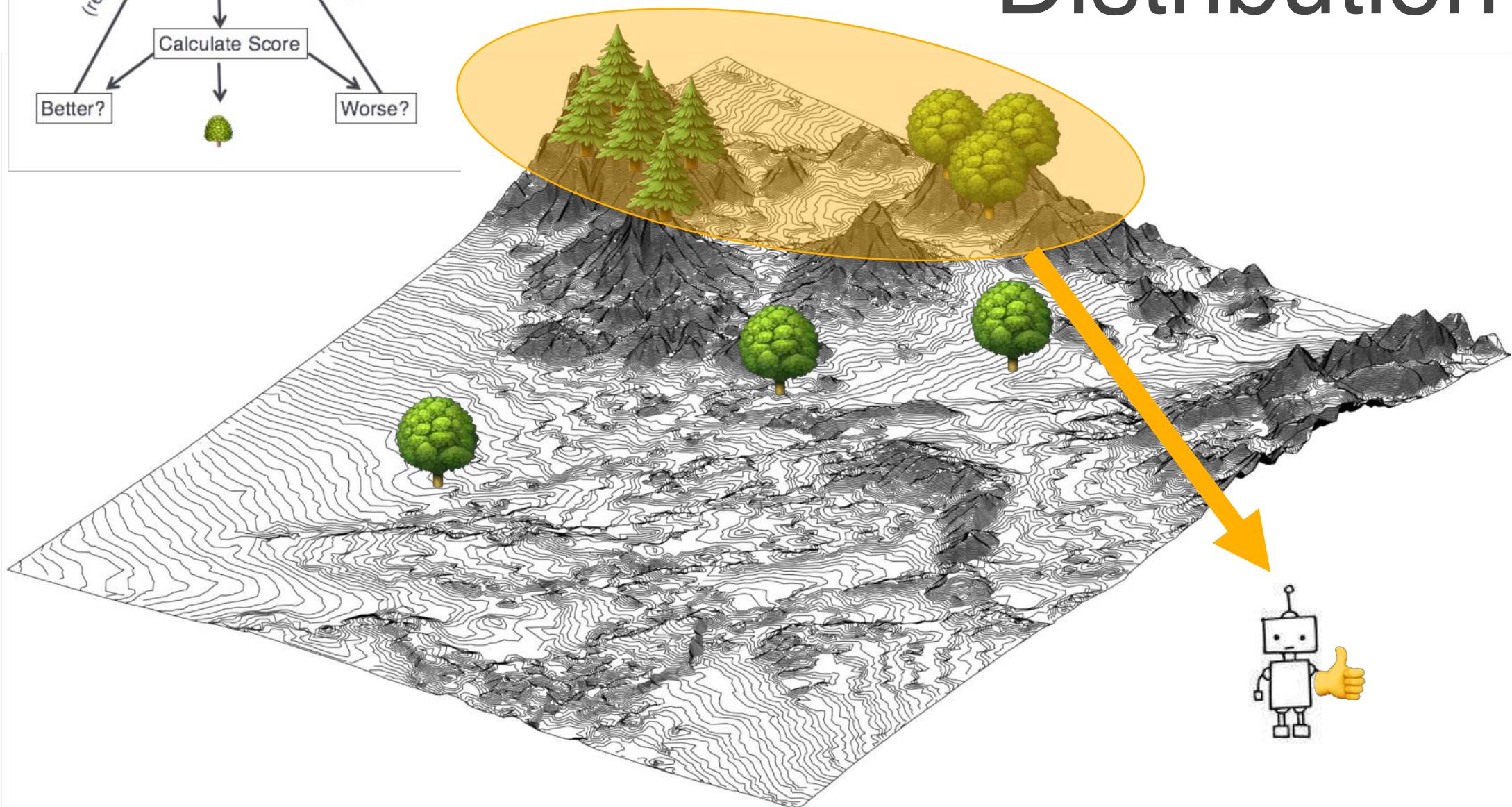
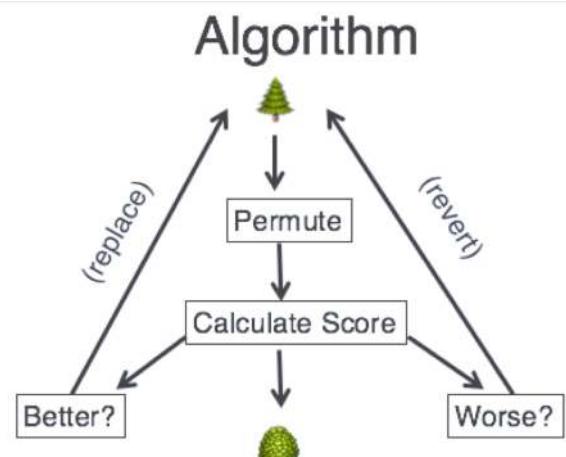
Algorithm



“burn-in”



Posterior Probability Distribution



Posterior Probability Distribution: Uto-Aztecán Languages

Covarion & Relaxed Clock + etc
Sampled 10,000 trees.



Posterior Probability Distribution: Uto-Aztec Languages

Covarion & Relaxed Clock + etc

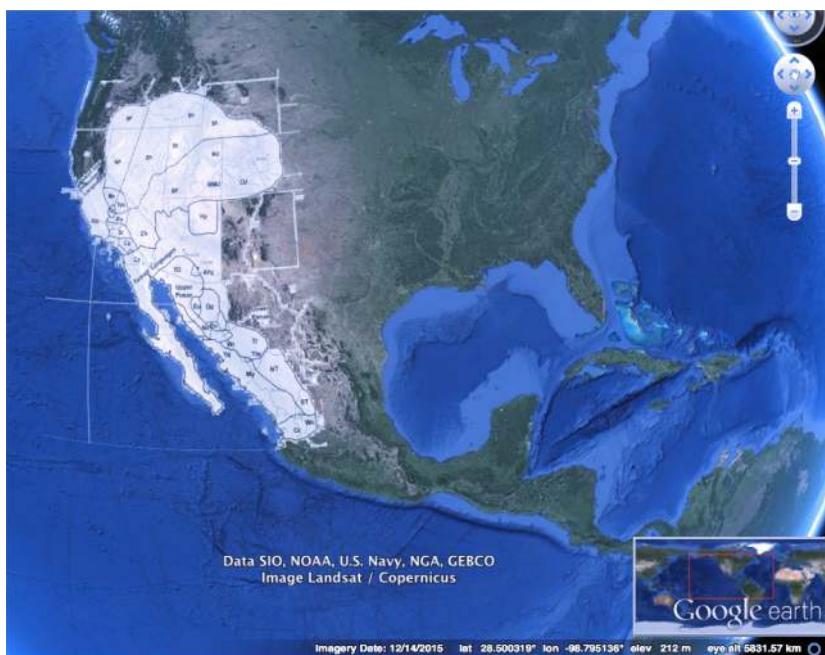
Sampled 10,000 trees.

Draw each one (Densitree, Bouckaert '10)

Some well supported regions

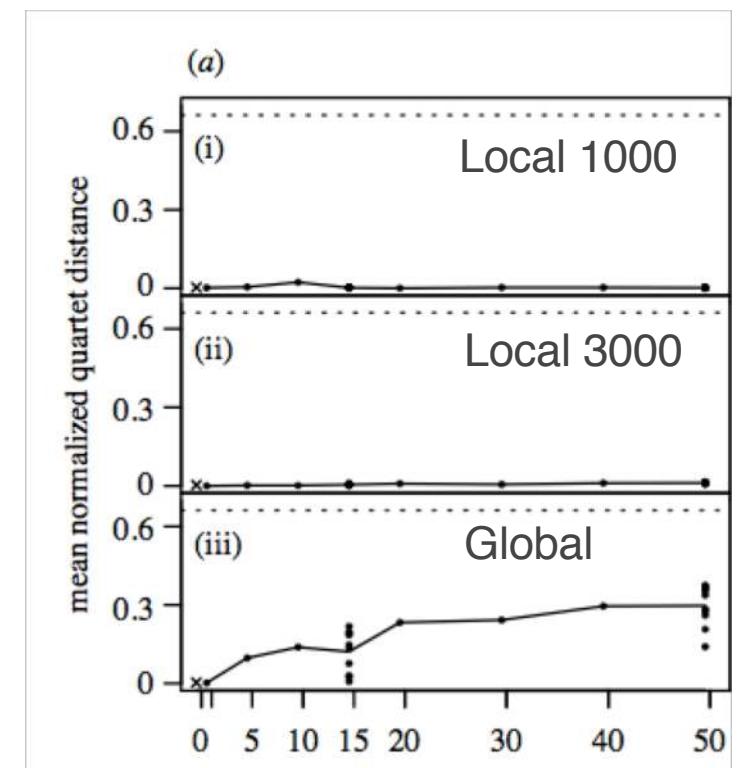
Some less...

Some conflict



Borrowing & B.P.M.

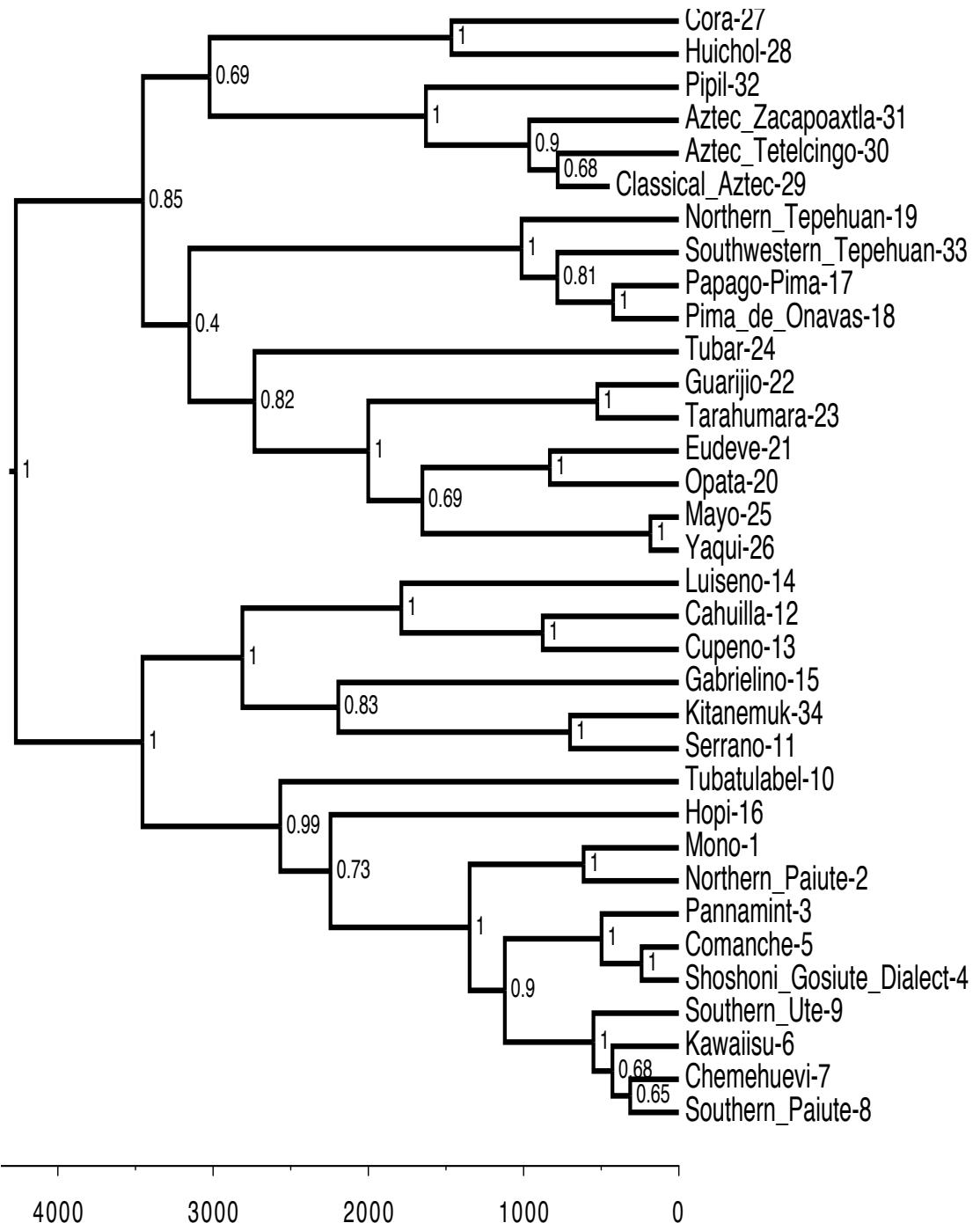
- Key benefit – robust to the effects of borrowing.
- Simulation study:
 - Take a “known” tree, simulate data on it.
 - ‘local’ vs. ‘global’ borrowing
 - 0-50% borrowing at each point.
 - Only in worst case scenarios do we see noticeable differences from true tree (global borrowing, >15% per year)
 - BUT never reaches a completely random tree.
- capture ‘majority signal’



(Greenhill et al. '10)

Posterior Probability Distribution: Uto-Aztec Languages

Maximum Clade Credibility Tree
Reduce posterior to single summary tree
Values are “Posterior probabilities”
- 1.00 = present in ALL trees in P.
- 0.50 = only half of the trees have this.



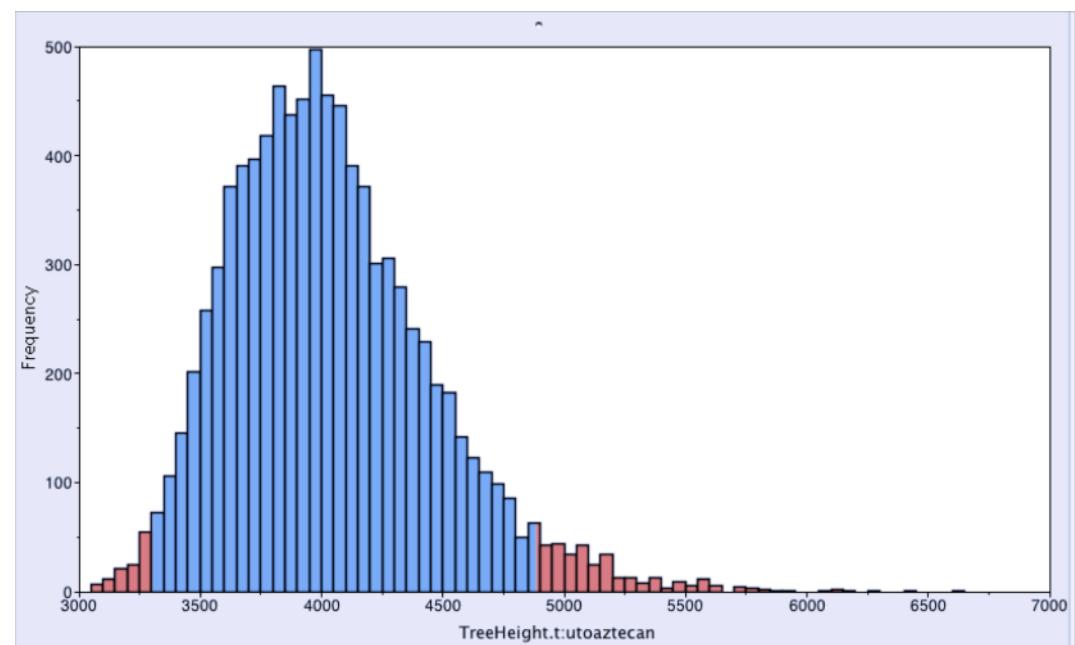
Posterior Probability Distribution: Uto-Aztecán Languages

...and any statistic we calculate
has a PP distribution.

- mean
- 95% highest posterior density interval
(95% probability that μ falls in range)



Mean = 4334 years



95% HPD: 3302 – 4884 years

Whirlwind tour

- Disciplinary parallels & similarity of questions
- Lexicostatistics and Glottochronology
- Parallel evolution of methods in Biology
- Maximum Parsimony
- Maximum Likelihood
- Bayesian Phylogenetics
- Theory
- Practical: My BEAST tutorial.
 - <https://taming-the-beast.github.io/>

