

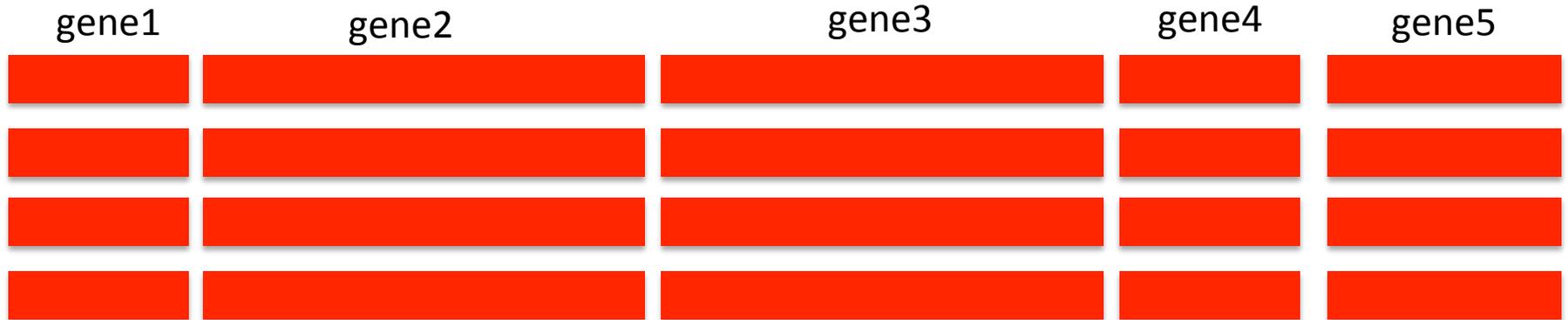
Recap: Maximum Likelihood

- We can model DNA substitutions using Markov models, expressed as instantaneous rate matrices
- Likelihood = $P(D|M)$
- We use matrix exponentiation of our rate matrix to calculate the probability of a substitution as a function of branch length $P_{ij}(t)$
- Felsenstein's pruning algorithm allows us to calculate the total likelihood of the tree

Fancy Things

- Error? Bootstrapping!!
- Rate heterogeneity
- Model testing
- Additional models (codon, amino acid, etc.)

Rate Heterogeneity: Partitions



	A	G	C	T
A	-3λ	λ	λ	λ
G	λ	-3λ	λ	λ
C	λ	λ	-3λ	λ
T	λ	λ	λ	-3λ

Rate Heterogeneity: Partitions

	A	G	C	T
A	-3λ	λ	λ	λ
G	λ	-3λ	λ	λ
C	λ	λ	-3λ	λ
T	λ	λ	λ	-3λ

	A	G	C	T
A	-3λ	λ	λ	λ
G	λ	-3λ	λ	λ
C	λ	λ	-3λ	λ
T	λ	λ	λ	-3λ

gene1

gene2

gene3

gene4

gene5



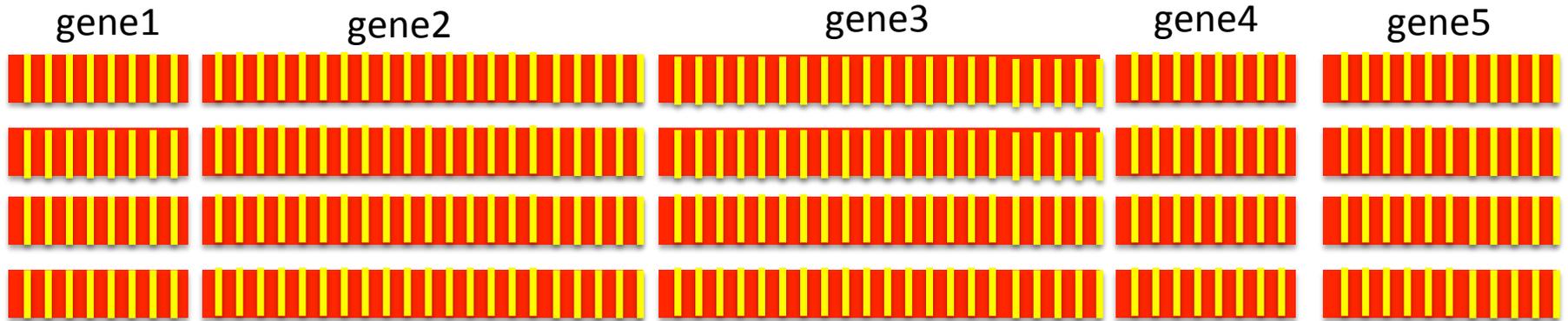
	A	G	C	T
A	-3λ	λ	λ	λ
G	λ	-3λ	λ	λ
C	λ	λ	-3λ	λ
T	λ	λ	λ	-3λ

	A	G	C	T
A	-3λ	λ	λ	λ
G	λ	-3λ	λ	λ
C	λ	λ	-3λ	λ
T	λ	λ	λ	-3λ

	A	G	C	T
A	-3λ	λ	λ	λ
G	λ	-3λ	λ	λ
C	λ	λ	-3λ	λ
T	λ	λ	λ	-3λ

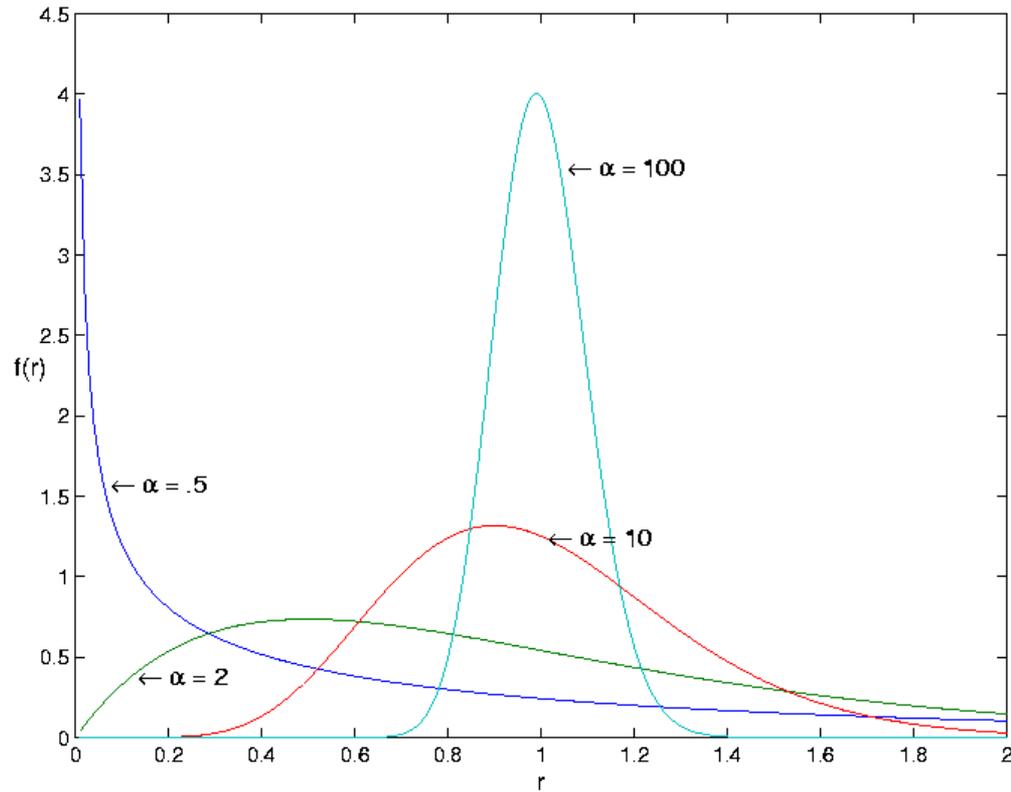
Estimate model parameters separately for different genes!

Rate Heterogeneity: Partitions



Estimate model parameters separately for different codon positions!

Rate Heterogeneity: gamma (Γ)



Model testing

Likelihood ratio test for **nested** models:

$$LR = 2 * (-\ln L_1 - -\ln L_2)$$

$$\text{HKY85 } -\ln L = 1787.08$$

$$\text{GTR } -\ln L = 1784.82$$

$$LR = 2(1787.08 - 1784.82) = 4.53, df = 4$$

critical value ($P = 0.05$) = 9.49 ($\sim \chi^2$ distribution)

Model testing

Aikake Information Criterion (**AIC**) (non nested):

$$AIC = 2k - 2 \ln(\hat{L})$$

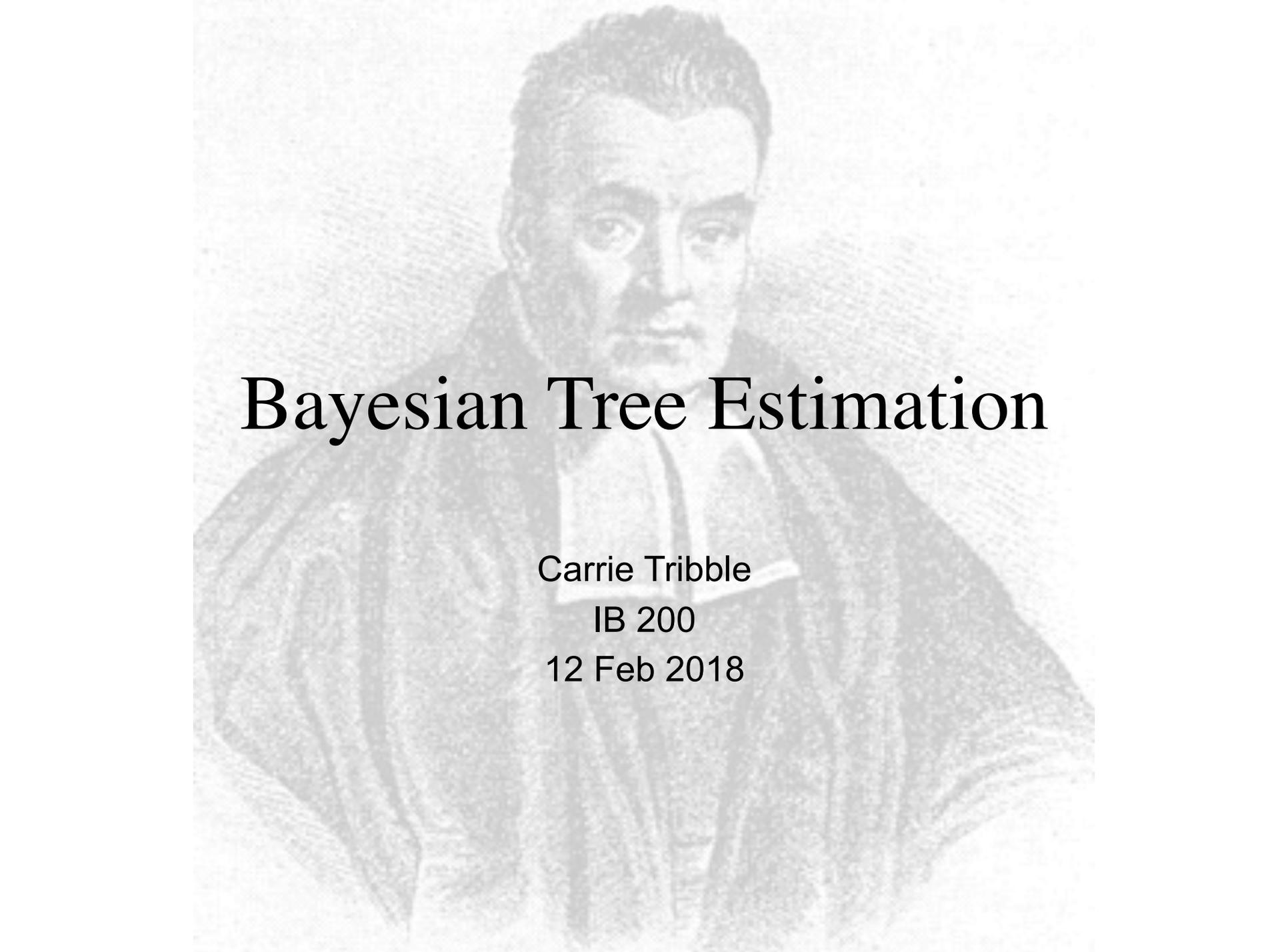
Find model with lowest AIC

- Models need not be nested
- Can correct for small sample sizes (AICc)
- BIC is a similar alternative

Additional models

- Amino Acid
- Codon
- non-markov
- What would a model of morphology look like?



A faded, grayscale portrait of a man with dark hair, wearing a suit and tie, looking directly at the camera. The image is semi-transparent and serves as a background for the text.

Bayesian Tree Estimation

Carrie Tribble

IB 200

12 Feb 2018

Basic probability theory

Probability is a quantitative measurement of the likelihood of an outcome of some random process.

- **Joint probability** = $P(A,B)$ = probability of A and B
- **Conditional probability** = $P(A|B)$ = probability of A given B
- **Marginal probability** = $P(A) = P(A,B) + P(A, \text{not } B)$

	Tenured	Assistant	total
Publish in Nature	33	36	69
Publish in Science	18	22	40
total	51	58	109

Basic probability theory

Probability is a quantitative measurement of the likelihood of an outcome of some random process.

- **Joint probability** = $P(A,B)$ = probability of A and B
- **Conditional probability** = $P(A|B)$ = probability of A given B
- **Marginal probability** = $P(A) = P(A,B) + P(A, \text{not } B)$

	Tenured	Assistant	total
Publish in Nature	33	36	69
Publish in Science	18	22	40
total	51	58	109

$$P(\text{Nature, Tenured}) = 33/109 = 0.30$$

Basic probability theory

Probability is a quantitative measurement of the likelihood of an outcome of some random process.

- **Joint probability** = $P(A,B)$ = probability of A and B
- **Conditional probability** = $P(A|B)$ = probability of A given B
- **Marginal probability** = $P(A) = P(A,B) + P(A, \text{not } B)$

	Tenured	Assistant	total
Publish in Nature	33	36	69
Publish in Science	18	22	40
total	51	58	109

$$P(\text{Tenured} | \text{Nature}) = 33/69 = 0.48$$

Basic probability theory

Probability is a quantitative measurement of the likelihood of an outcome of some random process.

- **Joint probability** = $P(A,B)$ = probability of A and B
- **Conditional probability** = $P(A|B)$ = probability of A given B
- **Marginal probability** = $P(A) = P(A,B) + P(A, \text{not } B)$

	Tenured	Assistant	total
Publish in Nature	33	36	69
Publish in Science	18	22	40
total	51	58	109

$$P(\text{Nature}) = 69/109 = 0.63$$

Basic probability theory



$$P(A | B) = \frac{P(B | A) P(A)}{P(B)},$$

Reverend Thomas Bayes, 1701 - 1761

Interpretations of probability

Frequentists

- Probability of the data
- Point estimate
- Based on frequency of event over time
- “Infinite series of trials”
- P-values, null hypotheses, etc.

Bayesians

- Probability of hypothesis
- Distribution of probabilities
- Based on degree of belief
- Updates hypotheses with data

Bayesian inference in science

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

Bayesian inference in science

The diagram illustrates Bayes' theorem with the following components and labels:

- Likelihood**: An orange arrow points from this label to the term $P(D|H)$ in the numerator of the fraction.
- Prior probability of hypothesis**: An orange arrow points from this label to the term $P(H)$ in the numerator of the fraction.
- Posterior probability**: An orange arrow points from this label to the term $P(H|D)$ on the left side of the equation.
- Marginal probability of data**: An orange arrow points from this label to the term $P(D)$ in the denominator of the fraction.

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

Bayesian phylogenetics

$$f(\Psi, v, \theta | X) = \frac{f(X | \Psi, v, \theta) f(\Psi, v, \theta)}{f(X)}$$

X = character matrix

Ψ = tree topology

v = branch lengths

θ = character evolution parameters

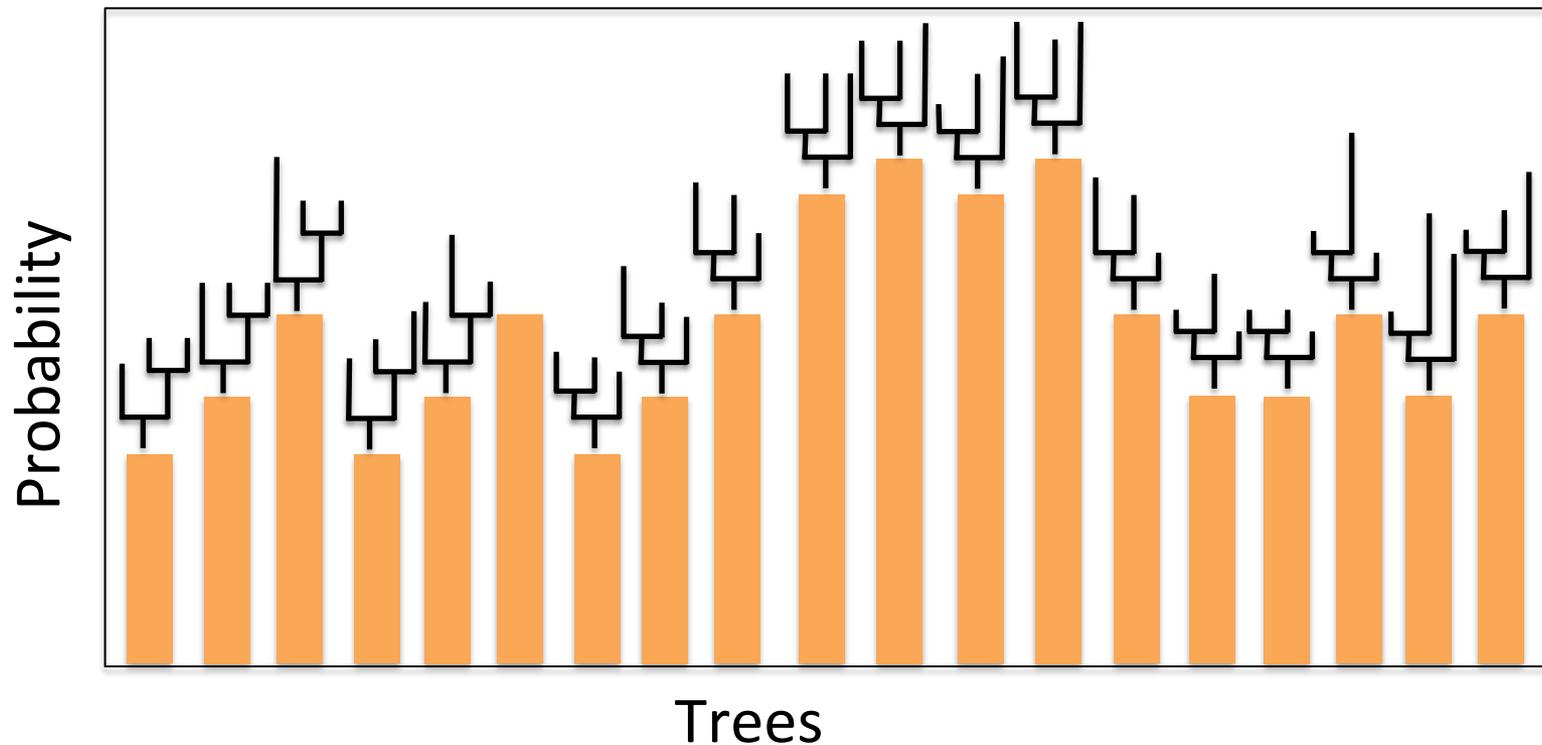
Bayesian phylogenetics

$$f(\Psi, v, \theta | X) = \frac{f(X | \Psi, v, \theta) f(\Psi, v, \theta)}{f(X)}$$

$$f(X) = \sum_{\Psi} \int_{\theta} \int_v f(X | \Psi, v, \theta) f(\Psi, v, \theta) d\theta dv$$

Bayesian phylogenetics: the posterior

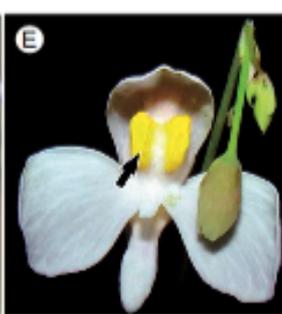
Goal is distribution of trees rather than THE tree



Bayesian phylogenetics: the posterior

Summarizing the posterior:

- *maximum a posteriori* (MAP): tree with highest posterior probability topology
- *maximum clade credibility* (MCC): tree with the highest product of clade probabilities
- *50% Majority rule consensus tree*: tree with only clades that occur in more than 50% of posterior



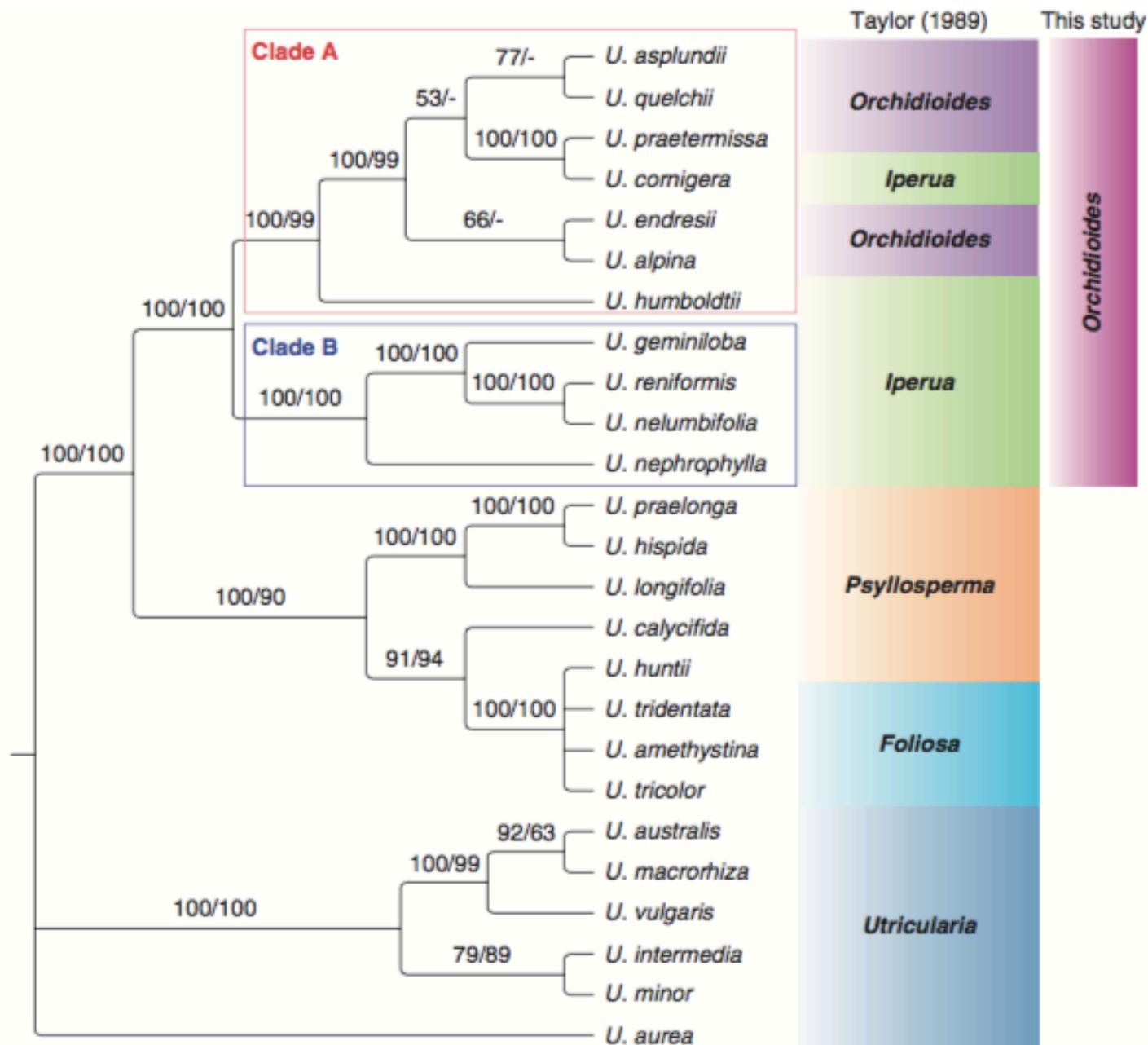


FIG. 3. Bayesian inference for the combined analysis (*rps16* + *trnL-F* + *matK* + ITS). Numbers above the branches are the posterior probabilities followed by maximum likelihood bootstraps. —, branches with support value <50.

Bayesian phylogenetics: the priors

- Are (often) distributions
- Can be informative or uninformative

DID THE SUN JUST EXPLODE?
(IT'S NIGHT, SO WE'RE NOT SURE.)



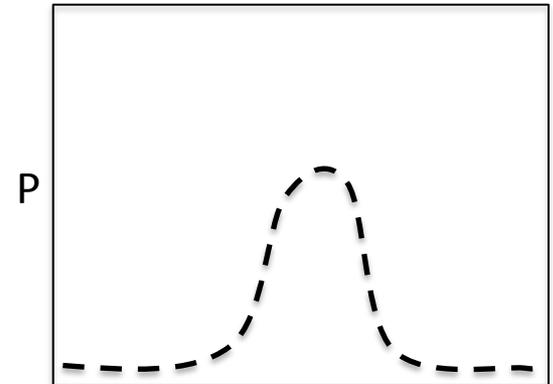
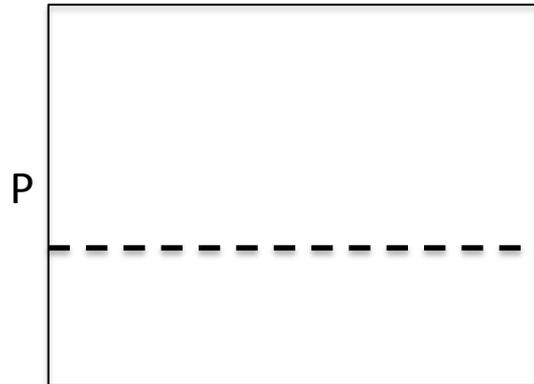
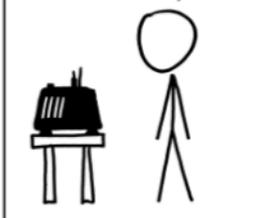
FREQUENTIST STATISTICIAN:

THE PROBABILITY OF THIS RESULT HAPPENING BY CHANCE IS $\frac{1}{36} = 0.027$. SINCE $p < 0.05$, I CONCLUDE THAT THE SUN HAS EXPLODED.



BAYESIAN STATISTICIAN:

BET YOU \$50 IT HASN'T.



Bayesian phylogenetics: the priors

- Make our assumptions explicit
- Incorporate additional information
- Can be tested

DID THE SUN JUST EXPLODE?
(IT'S NIGHT, SO WE'RE NOT SURE.)

THIS NEUTRINO DETECTOR MEASURES
WHETHER THE SUN HAS GONE NOVA.

THEN, IT ROLLS TWO DICE. IF THEY
BOTH COME UP SIX, IT LIES TO US.
OTHERWISE, IT TELLS THE TRUTH.

LET'S TRY:
DETECTOR! HAS THE
SUN GONE NOVA?



FREQUENTIST STATISTICIAN:

THE PROBABILITY OF THIS RESULT
HAPPENING BY CHANCE IS $\frac{1}{36} = 0.027$.
SINCE $p < 0.05$, I CONCLUDE
THAT THE SUN HAS EXPLODED.



BAYESIAN STATISTICIAN:

BET YOU \$50
IT HASN'T.



Markov chain Monte Carlo

MCMC

$$f(X) = \sum_{\Psi} \int_{\theta} \int_{v} f(X|\Psi, v, \theta) f(\Psi, v, \theta) d\theta dv$$

We can't calculate this, so we use MCMC instead.

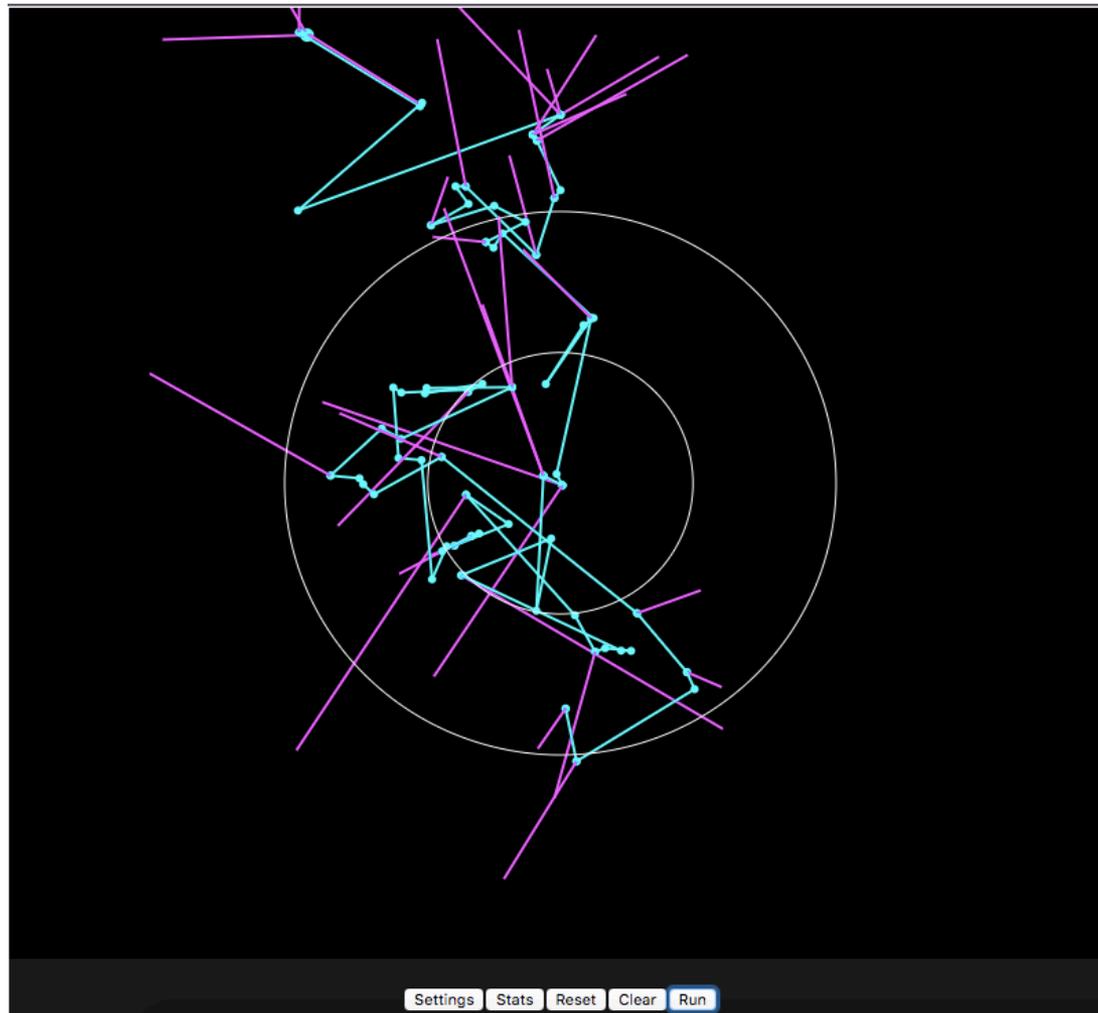
Markov chain = memoryless process

Monte Carlo = simulation based on random sampling

Markov chain Monte Carlo: Metropolis algorithm

1. Draw values for parameters θ by drawing from prior
2. Propose new values for θ' using a proposal distribution
3. Calculate acceptance ratio: $R = \min \left[1, \frac{f(X|\theta')}{f(X|\theta)} \times \frac{f(\theta')}{f(\theta)} \right]$
4. If $R \geq 1$, we set $\theta = \theta'$ and go back to step 2 (uphill)
5. If $R < 1$, draw random number $0 \leq \alpha \leq 1$. If $\alpha \leq R$, we set $\theta = \theta'$ and go back to step 2
6. Frequency we end up in region of parameter space is proportional to region's posterior probability

Markov chain Monte Carlo: Metropolis algorithm



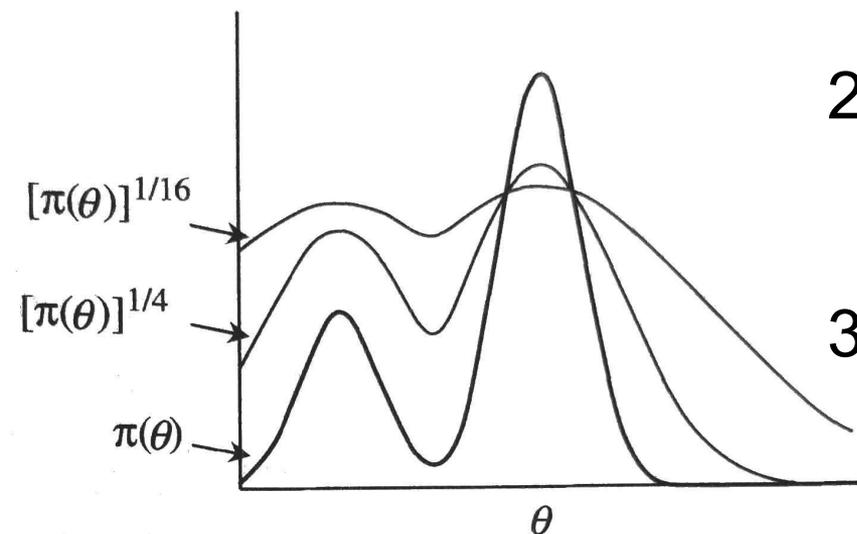
Markov chain Monte Carlo: Metropolis-coupled (MC³)

1. Hot and cold chains

2. Cold chain is 'normal' a la
Metropolis MCMC

3. Hot chain is 'flattened' by raising
posterior probability to $0 < \beta < 1$
to explore space faster

4. Chains occasionally switch
states



Markov chain Monte Carlo: reversible jump (rjMCMC)

1. Samples from multiple models with different dimensions
2. Draws in proportion to posterior probability of models
3. Allows for model selection and/or model averaging

Bayesian phylogenetics: model testing

1. rjMCMC

2. Bayes Factors: $K = \frac{P(M_1|D)}{P(M_2|D)} / \frac{P(M_1)}{P(M_2)}$

3. How is this similar/ different to likelihood ratio test or AIC?