

Bayesian phylogenetics



the one 'true' tree?

- the methods we've learned so far try to get a single tree that best describes the data
- however, they admit that they don't search everywhere, and that it is difficult to find the "best" tree
- are we doing a good job reporting a single tree ??



Bayesian phylogenetics

- using Bayesian principles, we will search for a set of plausible trees (weighed by their probability) instead of a single best tree
- in this method, the "space" that you search in is limited by **prior** information and the **data**
- the **posterior** distribution of trees can be translated to a probability of any branching event
 - allows estimate of uncertainty!
 - BUT incorporates prior beliefs





World championship medalists



1999	Mario De Clercq (BEL)	Erwin Vervecken (BEL)	Adri van der Poel (NED)
2000	Richard Groenendaal (NED)	Mario De Clercq (BEL)	Sven Nys (BEL)
2001	Erwin Vervecken (BEL)	Petr Dlask (CZE)	Mario De Clercq (BEL)
2002	Mario De Clercq (BEL)	Tom Vannoppen (BEL)	Sven Nys (BEL)
2003	Bart Wellens (BEL)	Mario De Clercq (BEL)	Erwin Vervecken (BEL)
2004	Bart Wellens (BEL)	Mario De Clercq (BEL)	Sven Vanthourenhout (BEL)
2005	Sven Nys (BEL)	Erwin Vervecken (BEL)	Sven Vanthourenhout (BEL)
2006	Erwin Vervecken (BEL)	Bart Wellens (BEL)	Francis Mouray (FRA)
2007	Erwin Vervecken (BEL)	Jonathan Page (USA)	Enrico Franzoi (ITA)
2008	Lars Boom (NED)	Zdeněk Štybar (CZE)	Sven Nys (BEL)

World championship medalists



1989	Danny De Bie (BEL)	Adri van der Poel (NED)	Christophe Lavainne (FRA)
1990	Henk Baars (NED)	Adri van der Poel (NED)	Bruno Lebras (FRA)
1991	Radomir Simunek (CZE)	Adri van der Poel (NED)	Bruno Lebras (FRA)
1992	Mike Kluge (GER)	Carl Camrda (CZE)	Adri van der Poel (NED)
1993	Dominique Arnould (FRA)	Mike Kluge (GER)	Wim De Vos (NED)
1994	Paul Herijgers (BEL)	Richard Groenendaal (NED)	Erwin Vervecken (BEL)
1995	Dieter Runkel (SUI)	Richard Groenendaal (NED)	Beat Wabel (SUI)
1996	Adri van der Poel (NED)	Daniele Pontoni (ITA)	Luca Bramati (ITA)
1997	Daniele Pontoni (ITA)	Thomas Frischknecht (SUI)	Luca Bramati (ITA)
1998	Mario De Clercq (BEL)	Erwin Vervecken (BEL)	Henrik Djernis (DEN)

Bayes rule:

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

- H represents a specific hypothesis
- $\Pr(H)$ is called the *prior* probability of H that was assumed before new data, D, became available.
- $\Pr(D|H)$ is called the *conditional probability* of seeing D if H is true. It is also called a *likelihood function* when it is considered as a function of H for fixed D
- $\Pr(D)$ is called the *marginal probability* of D: the a priori probability of witnessing the D under all possible hypotheses. It can be calculated as the sum of the product of all probabilities of any complete set of mutually exclusive hypotheses and corresponding conditional probabilities:

$$\Pr(D) = \sum \Pr(D|H_i)\Pr(H_i)$$

- $\Pr(H|D)$ is called the *posterior probability* of H given D.

Bayes rule

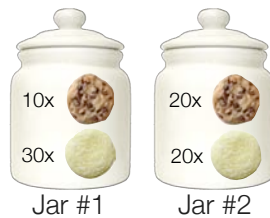
$$\Pr(H|D) = \frac{\Pr(D|H)\Pr(H)}{\Pr(D)}$$

- I choose a cookie from a jar at random, and it turns out to be a sugar cookie
- how probable is it that I picked from Jar #1?

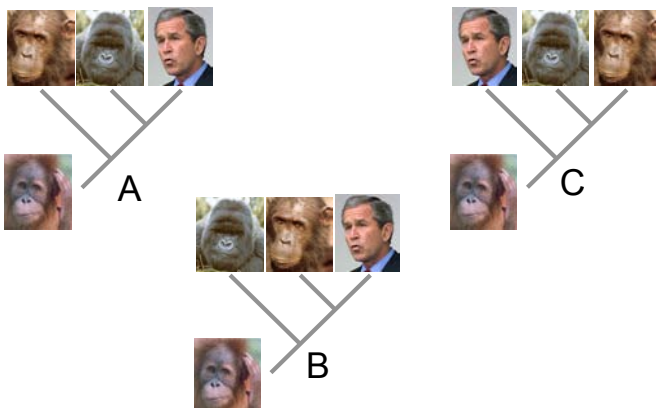
$$\Pr(\text{Jar1}|\text{sc}) = \frac{\Pr(\text{sc}|\text{Jar1})\Pr(\text{Jar1})}{[\Pr(\text{sc}|\text{Jar1})\Pr(\text{Jar1})] + [\Pr(\text{sc}|\text{Jar2})\Pr(\text{Jar2})]}$$

$$= \frac{0.75 * 0.5}{(0.75 * 0.5) + (0.5 * 0.5)}$$

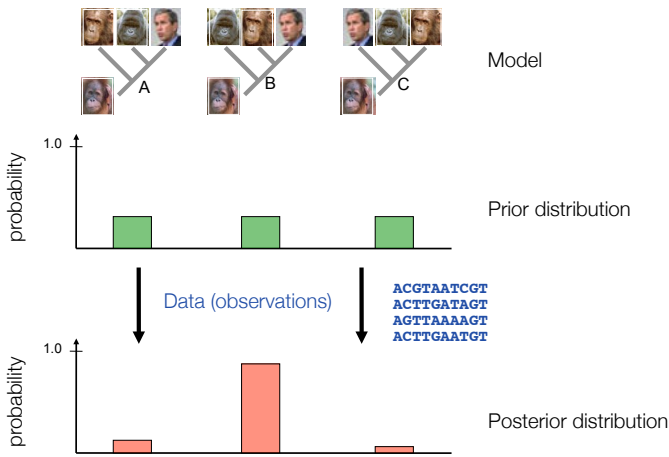
$$= 0.6$$



Bayesian phylogenetic inference



Bayesian phylogenetic inference



Bayesian inference of phylogeny

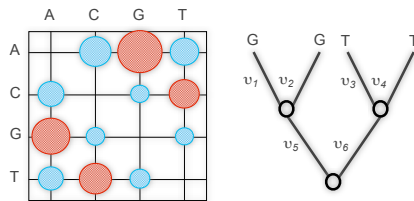
The posterior probability of a phylogenetic tree, τ :

$$f(\tau_i | X) = \frac{f(X | \tau_i) f(\tau_i)}{\sum_{j=1}^{B(s)} f(X | \tau_j) f(\tau_j)}$$

Likelihood: $L(\tau, \eta, \Theta | y_1, \dots, y_n) = \prod_{l=1}^n \Pr[y_l | \tau, \eta, \Theta]$

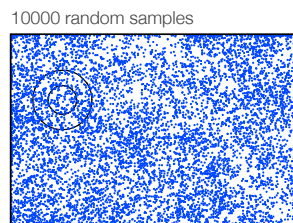
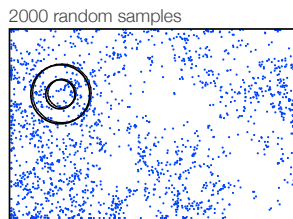
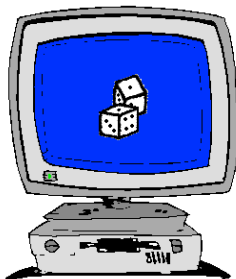
Labels: topology (red), branch lengths (blue), alignment sites (green), model parameters (green)

$B(s) = (2n-3)!!$



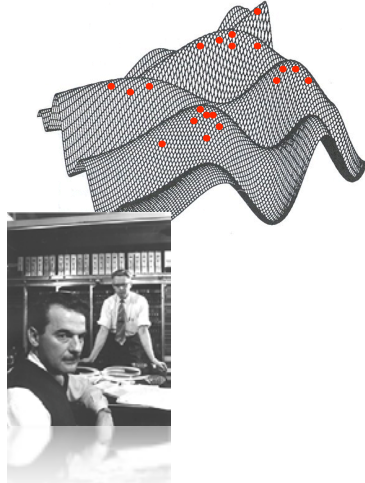
Prior: $\frac{1}{B(s)}$ ← number of topologies for s taxa

Bayesian inference: numerical integration via sampling as a solution?

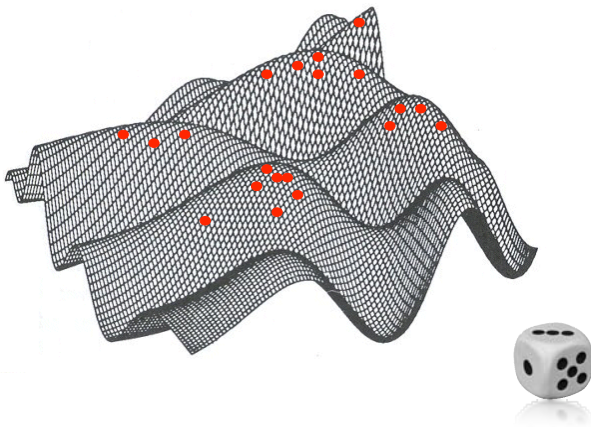


Markov chain Monte Carlo (MCMC) integration

- Metropolis et al. (1953) and Hastings (1970) proposed a stochastic algorithm (a variant of MCMC) to explore vast parameter spaces
- the frequency at which a tree is visited will be proportional to the posterior probability of this tree
 - trees with higher posterior probability will be visited more often
 - Ergodic Theorem



MCMC



Metropolis-Hastings algorithm

The algorithm starts from a random state $(\Psi = \{\tau, \mathbf{u}, \theta\})$ and 'proposes' a new state (Ψ^*)

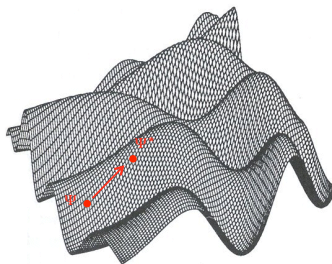
The new state is accepted with probability:

$$R = \min \left(1, \frac{f(\Psi^*|X)}{f(\Psi|X)} \times \frac{f(\Psi|\Psi^*)}{f(\Psi^*|\Psi)} \right)$$

$$= \min \left(1, \frac{f(X|\Psi^*) f(\Psi^*)}{f(X|\Psi) f(\Psi)} \times \frac{f(\Psi|\Psi^*)}{f(\Psi^*|\Psi)} \right)$$

$$= \min \left(1, \frac{f(X|\Psi^*)}{f(X|\Psi)} \times \frac{f(\Psi^*)}{f(\Psi)} \times \frac{f(\Psi|\Psi^*)}{f(\Psi^*|\Psi)} \right)$$

Likelihood ratio Prior ratio Proposal ratio



Draw random variable $q \sim (0,1)$:

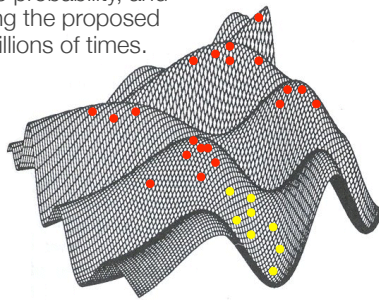
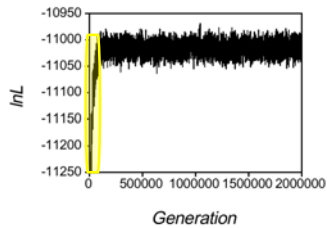
$q < R$: $\Psi = \Psi^*$
 Otherwise chain remains in original state

Metropolis-Hastings / MCMC



This process of proposing a new state, calculating the acceptance probability, and either accepting or rejecting the proposed move is repeated many millions of times.

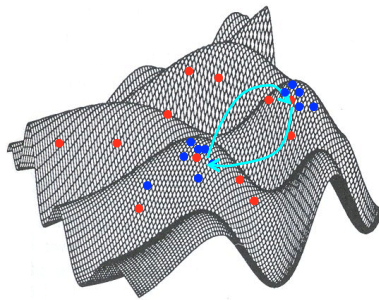
The samples from the algorithm form a Markov chain of valid, albeit correlated, samples from the posterior probability distribution



The initial sampling (=burn-in) is required to reach stationarity and should be discarded when summarizing the results

Metropolis-coupled MCMC

How to cross deep valleys?



Running multiple chains, some of which are heated:



Heating results in lower peaks and filling in valleys
After a step in the chain, a swap is attempted
If the swap is accepted, then the states are exchanged

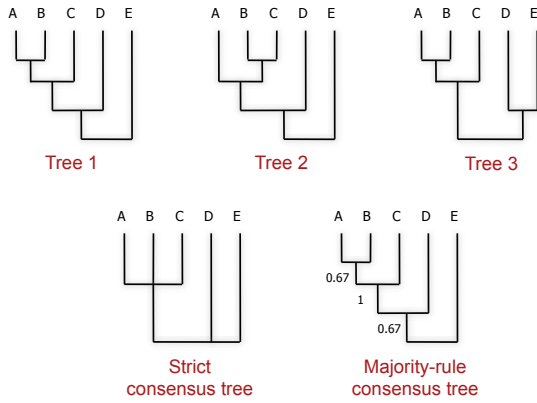
Bayesian inference in phylogeny

We have almost no prior knowledge for the parameters of interest....

So why bother doing Bayesian inference?

- ! A Bayesian analysis expresses its results as the probability of the hypothesis given the data
- ! MCMC is a stochastic algorithm and thus is able to avoid getting stuck in a local suboptimal solution.
- ! By sampling a set of plausible trees, MCMC allows estimating of the uncertainty of any branching event

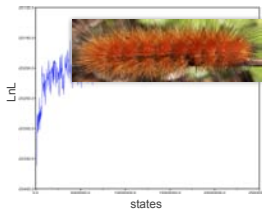
summarizing trees



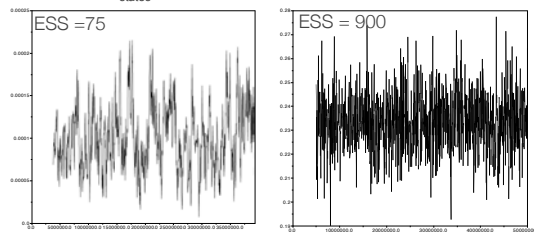
MCMC diagnostics

measure the behavior of a single chain

visually inspect output traces



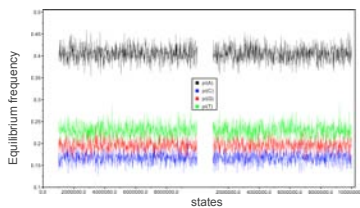
measure autocorrelation within a chain: the effective sampling size (ESS)



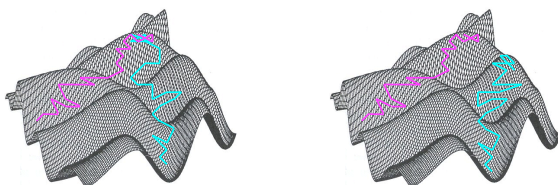
MCMC diagnostics

compare samples from different runs

Visually inspect output traces



compare different runs = comparing the variance among and within runs.



improving convergence

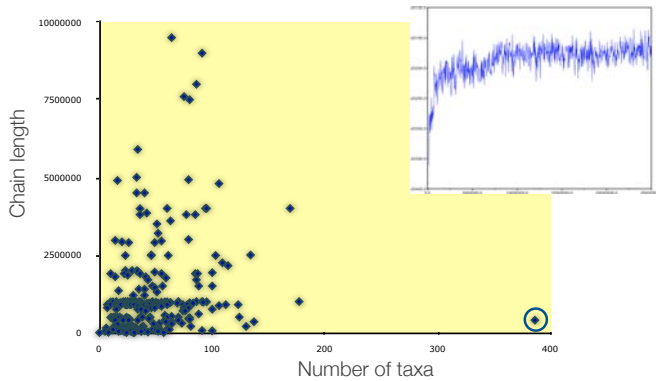
(Only if convergence diagnostics indicate problem!)

- run the chain longer
- increase the number of heated chains
- change heating temperature to bring acceptance rate of swaps between adjacent chains into the range 10 % to 70 %.
- change tuning parameters of proposals to bring acceptance rate into the range 10 % to 70 %
- propose changes to 'difficult' parameters more often
- use different proposal mechanisms
- make the model more realistic

improving MCMC mixing

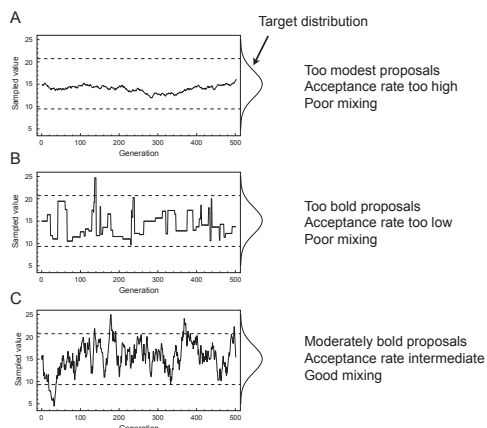
1. chain length

Bayesian analyses gives you a lot of rope...
... to hang yourself with (J.P. Huelsenbeck)



improving MCMC mixing

2. proposals



a brief recap... to be followed up

1. frequentist "probability" = long-run fraction having this characteristic.
2. Bayesian "probability" = degree of believability.
3. A frequentist is a person whose long-run ambition is to be wrong 5% of the time.
4. A Bayesian is one who, vaguely expecting a horse, and catching a glimpse of a donkey, strongly believes he has seen a mule.

- charles annis