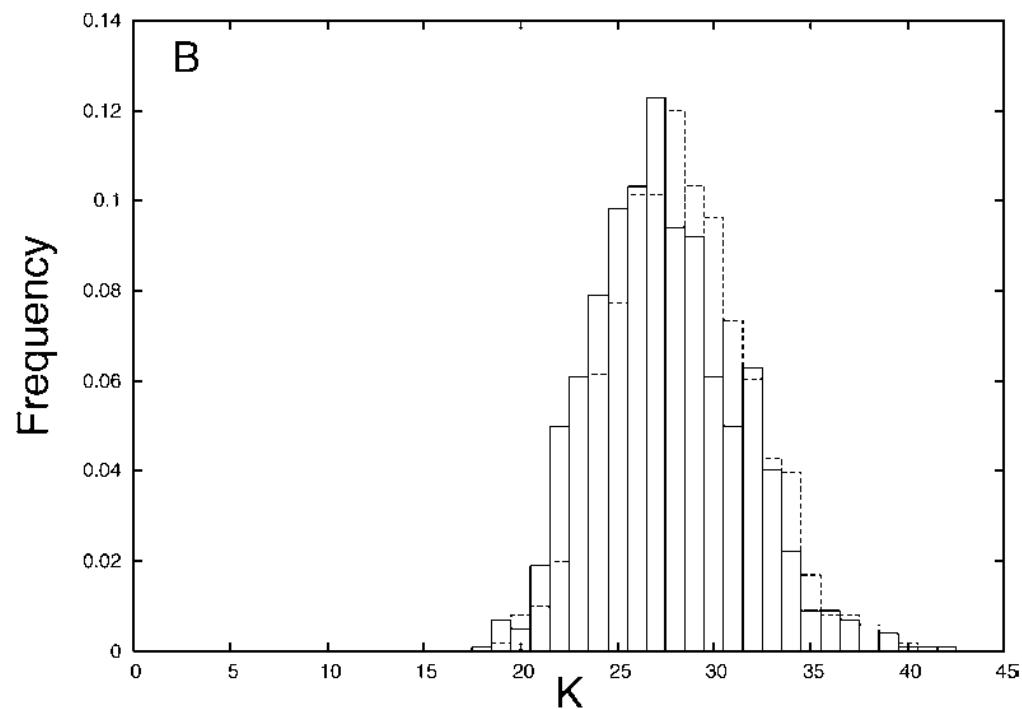
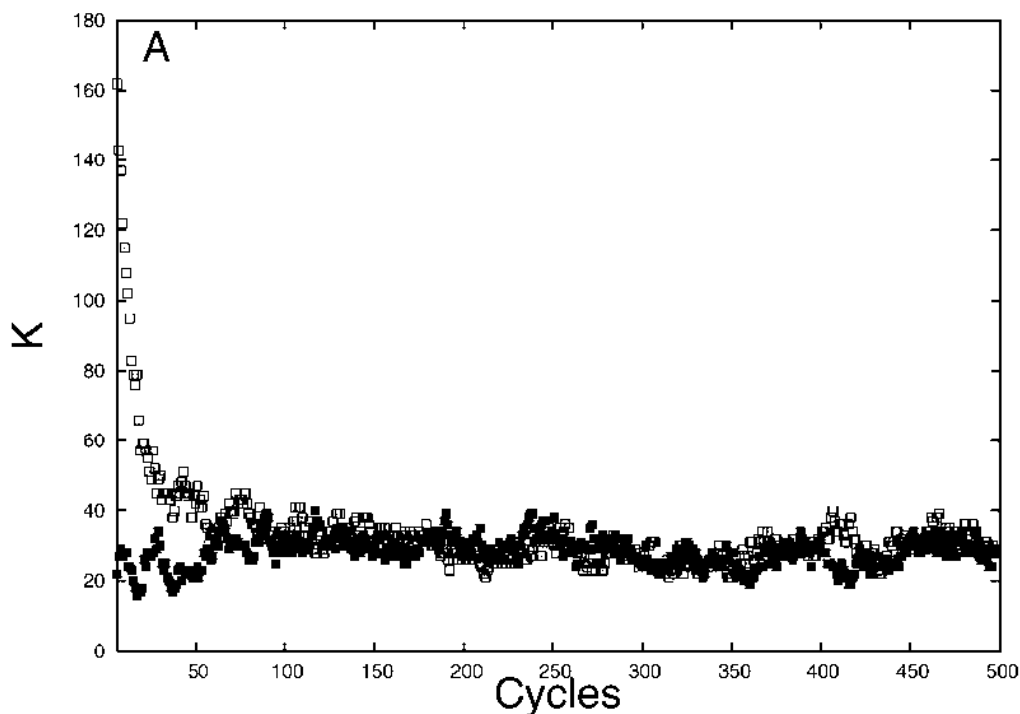


# Dirichlet process prior applet

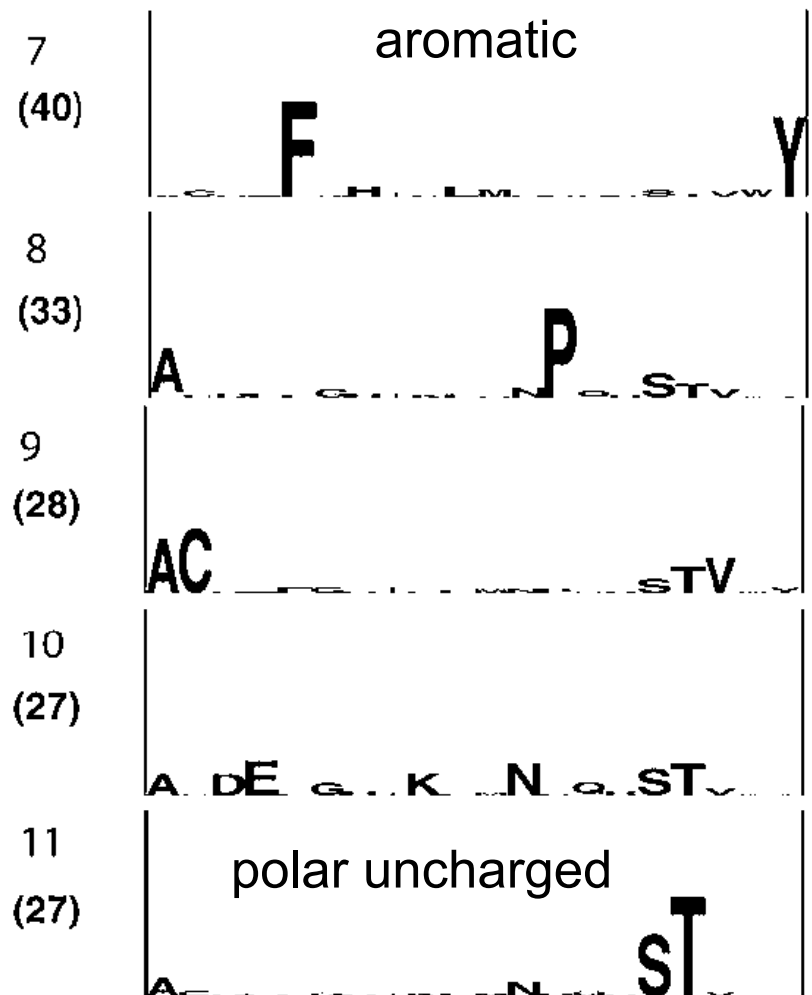
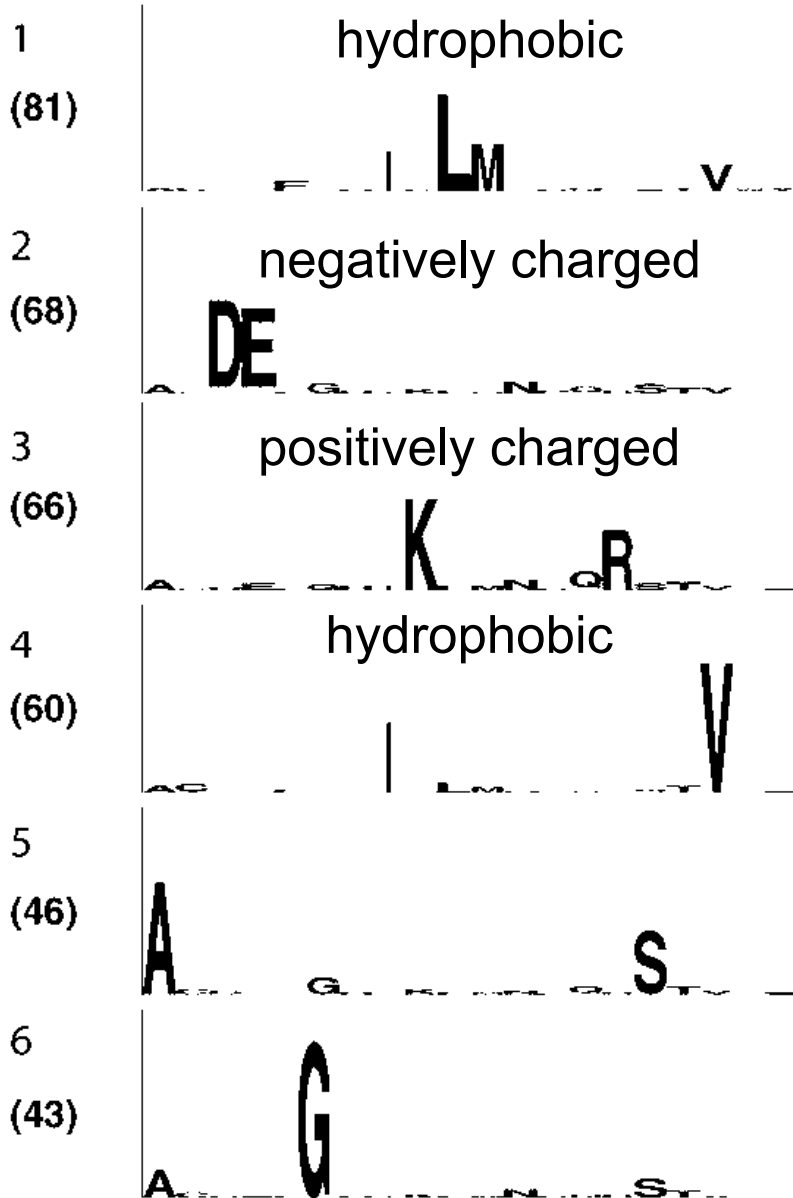
<https://plewis.github.io/applets/dpp/>

# Example 1: Elongation factor 2 (software: PhyloBayes)



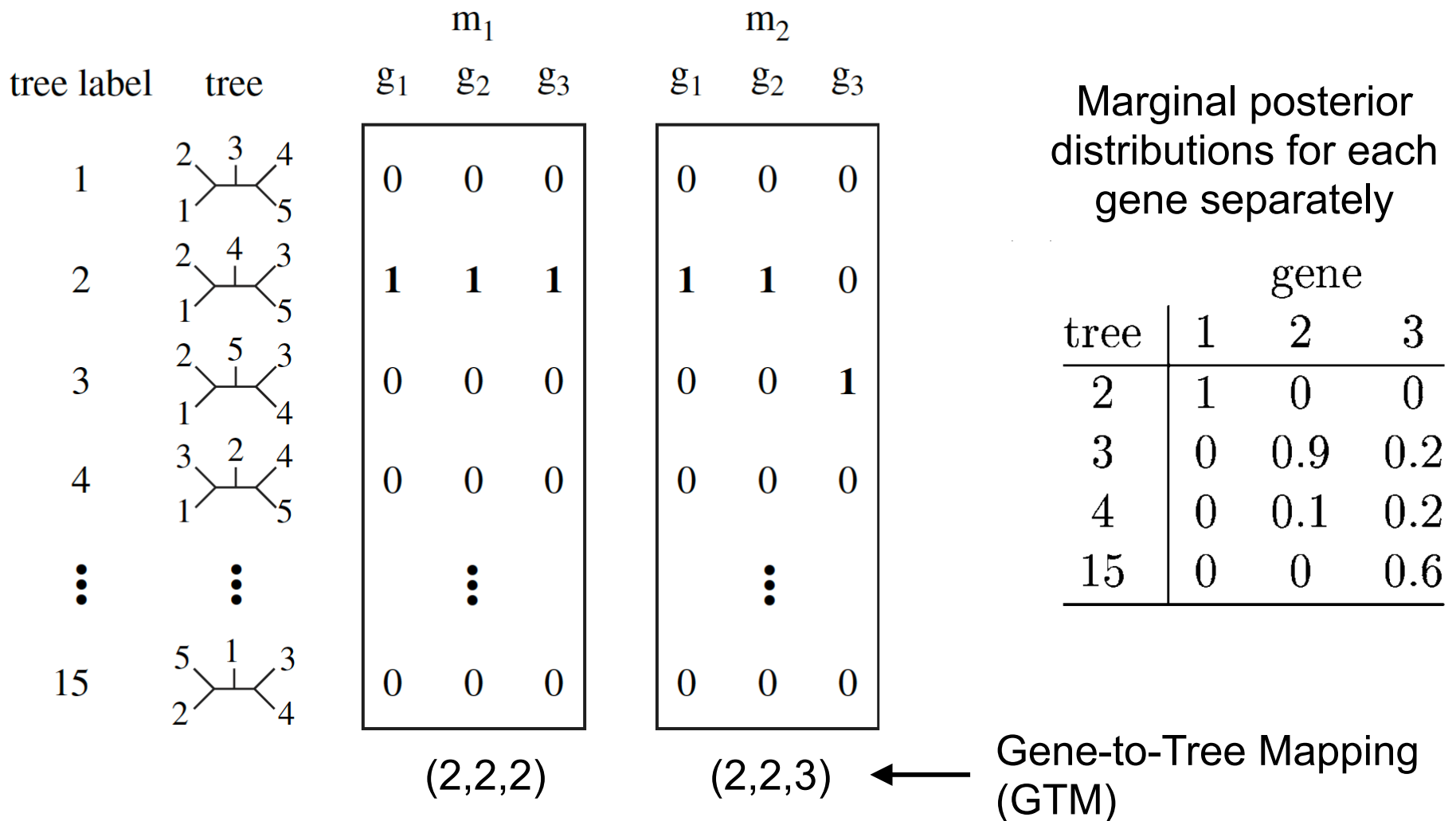
This model is implemented in PhyloBayes software: <http://www.atgc-montpellier.fr/phylobayes/>

D

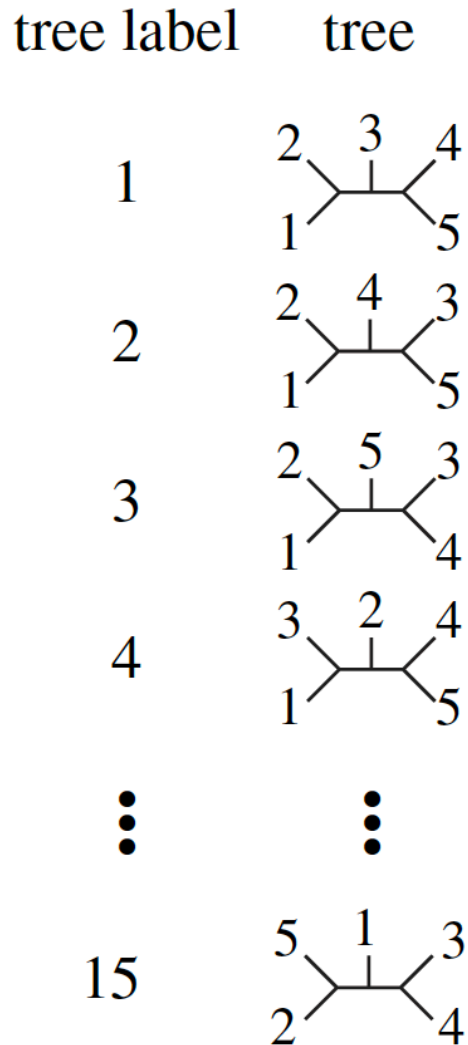


L=leucine, M=methionine  
D=aspartic acid, E=glutamic acid  
K=lysine, R=arginine, G=glycine,  
V=valine, I=isoleucine, A=alanine,  
F=phenylalanine, Y=tyrosine  
S=serine, T=threonine

# Example 2: Bayesian concordance analyses (software: BUCKy)



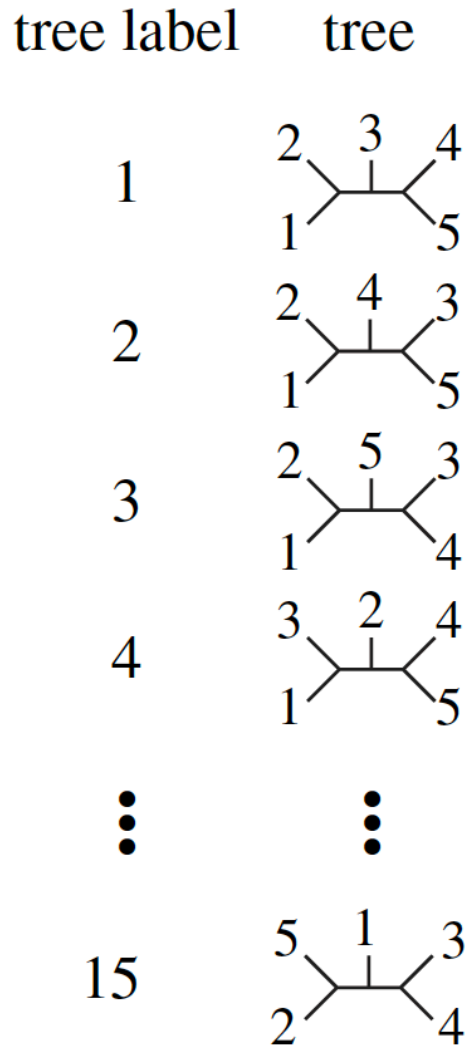
# Concordance Factors (CF)



GTM	Posterior	K	CF 12 345
(2,3,3)	0.6600		
(2,3,4)	0.0600		
(2,3,15)	0.1800		
(2,4,3)	0.0067		
(2,4,4)	0.0733		
(2,4,15)	0.0200		

1.0000

# Concordance Factors (CF)

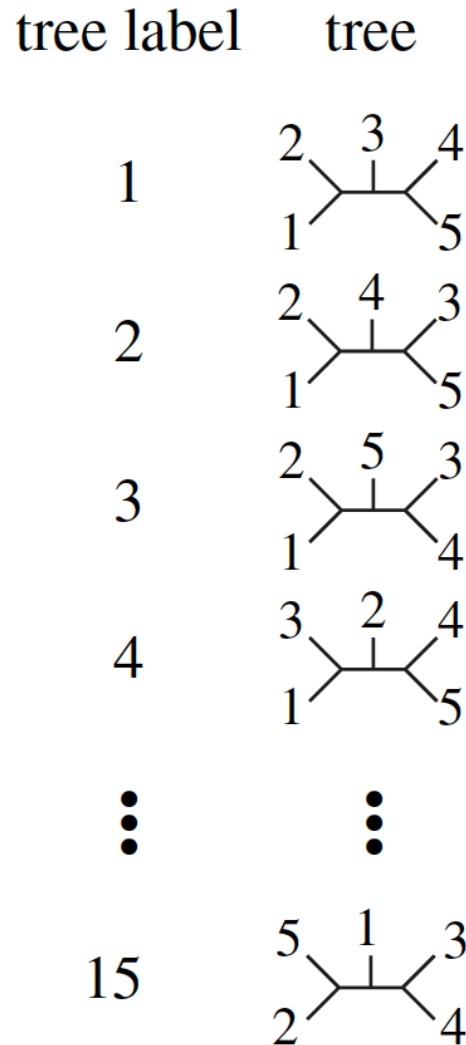


GTM	Posterior	K	CF 12 345
(2,3,3)	0.6600	2	
(2,3,4)	0.0600	3	
(2,3,15)	0.1800	3	
(2,4,3)	0.0067	3	
(2,4,4)	0.0733	2	
(2,4,15)	0.0200	3	

1.0000

2.27

# Concordance Factors (CF)



GTM	Posterior	K	CF 12 345
(2,3,3)	0.6600	2	1
(2,3,4)	0.0600	3	2/3
(2,3,15)	0.1800	3	2/3
(2,4,3)	0.0067	3	2/3
(2,4,4)	0.0733	2	1/3
(2,4,15)	0.0200	3	1/3

1.0000

2.27

0.86

# Dirichlet Process Priors

- To encourage **few, large** groups, use a **small** alpha value
- To encourage **lots of small** groups, use a **large** alpha value
- In practice, **hierarchical models** are used (i.e. alpha is a hyperparameter that can be estimated, so you need not worry about choosing the appropriate value for alpha)
- Bottom line: DP models are very nice for automatically grouping sites into clusters that have some property in common