

# Comparative Biology with focus on 4 examples

*Comparative Biology*

*The Domain of Comparative Biology*

*Co-modeling in Comparative Biology*

*The purpose of Comparative Biology*

*Examples of Stochastic Comparative Modeling*

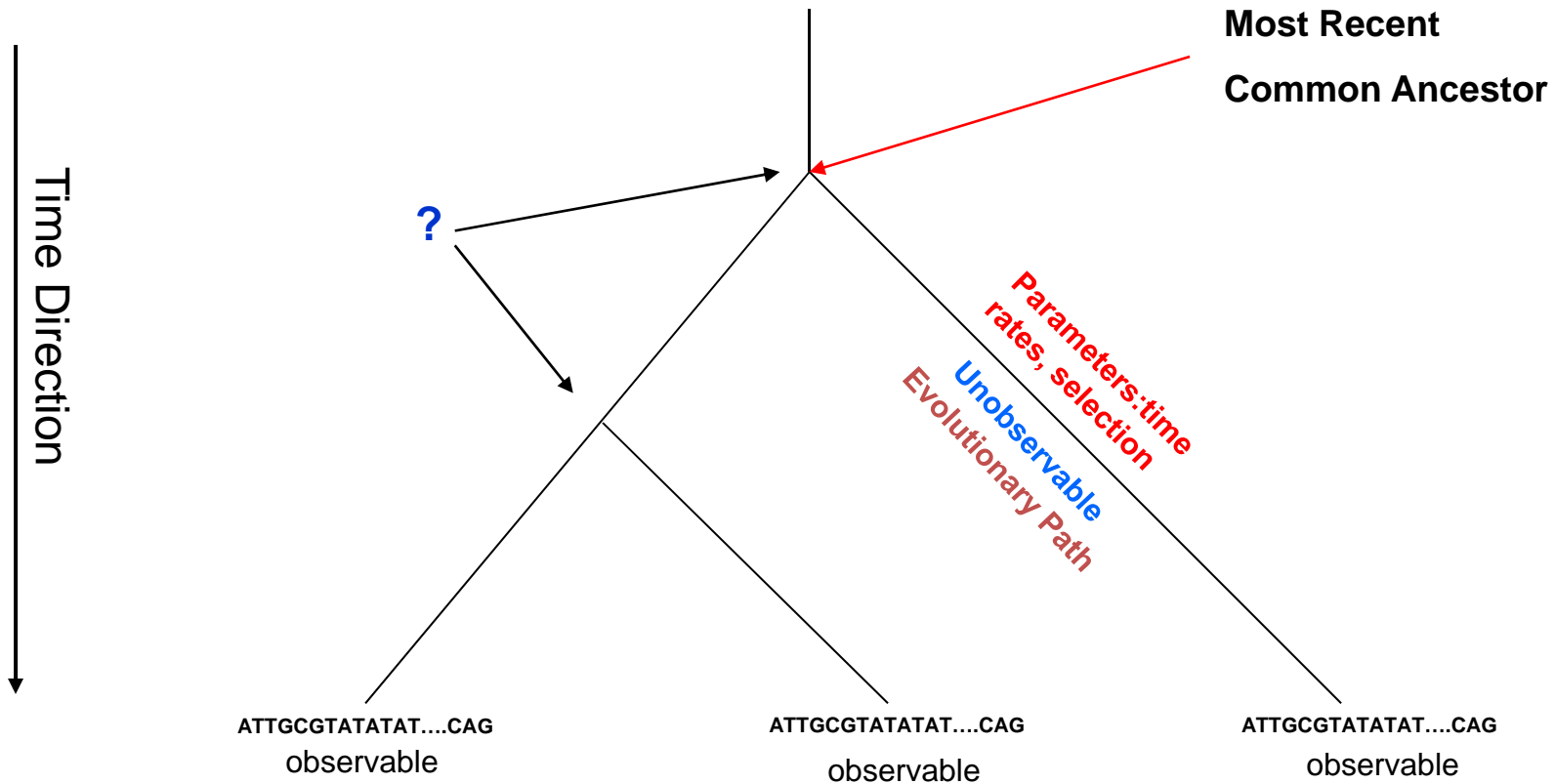
*Protein Structure Evolution*

*Movement Evolution*

*Metabolic Pathway Evolution*

*Dynamic System Evolution*

# Comparative Biology



## Key Questions:

- Which phylogeny?
- Which ancestral states?
- Which process?

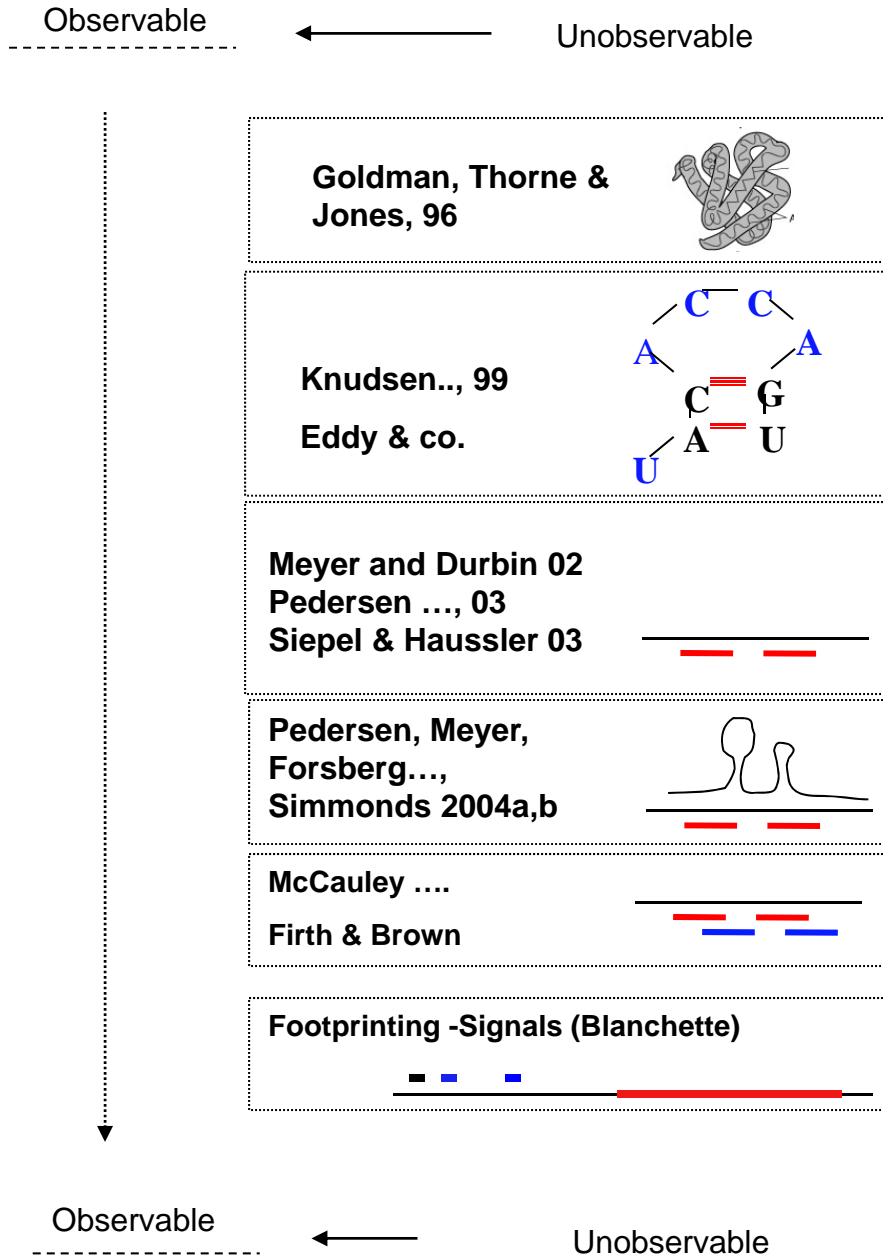
## Key Generalisations:

- Homologous objects
- Co-modelling
- Genealogical Structures?

# Comparative Biology: Evolutionary Models

| <u>Object</u>                  | <u>Type</u>                           | <u>Reference</u>                               |
|--------------------------------|---------------------------------------|--|
| Nucleotides/Amino Acids/codons | CTFS continuous time finite states    | Jukes-Cantor 69 +500 others                    |
| Continuous Quantities          | CTCS continuous time countable states | Felsenstein 68 + 50 others                     |
| Sequences                      | CTCS                                  | Thorne, Kishino Felsenstein,91 + 40others      |
| Gene Structure                 | Matching                              | DeGroot, 07                                    |
| Genome Structure               | CTCS MM                               | Miklos,  |
| Structure                      |                                       |  |
| RNA                            | SCFG-model like                       | Holmes, I. 06 + few others                     |
| Protein                        | non-evolutionary: extreme variety     | Lesk, A;Taylor, W.                             |
| Networks                       | CTCS                                  | Snijder, T (sociological networks)             |
| Metabolic Pathways             | ?                                     |  |
| Protein Interaction            | CTCS                                  | Stumpf, Wiuf, Ideker                           |
| Regulatory Pathways            | CTCS                                  | Quayle and Bullock, 06                         |
| Signal Transduction            | CTCS                                  | Soyer et al.,06                                |
| Macromolecular Assemblies      | ?                                     |  |
| Motors                         | ?                                     |  |
| Shape                          | - (non-evolutionary models)           | Dryden and Mardia, 1998                        |
| Patterns                       | - (non-evolutionary models)           | Turing, 52;                                    |
| Tissue/Organs/Skeleton/....    | - (non-evolutionary models)           | Grenander,                                     |
| Dynamics                       |                                       |  |
| MD movements of proteins       | -                                     |  |
| Locomotion                     | -                                     |  |
| Culture                        | analogues to genetic models           | Cavalli-Sforza & Feldman, 83                   |
| Language                       |                                       |  |
| Vocabulary                     | “Infinite Allele Model” (CTCS)        | Swadesh,52, Sankoff,72, Gray & Aitkinson, 2003 |
| Grammar                        |                                       | Dunn 05  |
| Phonetics                      |                                       | Bouchard-Côté 2007                             |
| Semantics                      |                                       | Sankoff,70                                     |
| Phenotype                      | Brownian Motion/Diffusion             |  |
| Dynamical Systems              | -                                     |  |

# Co-Modelling and Conditional Modelling



AGGTATATA**ATGCG**.....  $P_{\text{coding}}\{\text{ATG} \rightarrow \text{GTG}\}$  or  
AGCCATTTA**GTGCG**.....  $P_{\text{non-coding}}\{\text{ATG} \rightarrow \text{GTG}\}$



## • Conditional Modelling

$$P(\text{Sequence} \mid \text{Structure})P(\text{Structure}) =$$

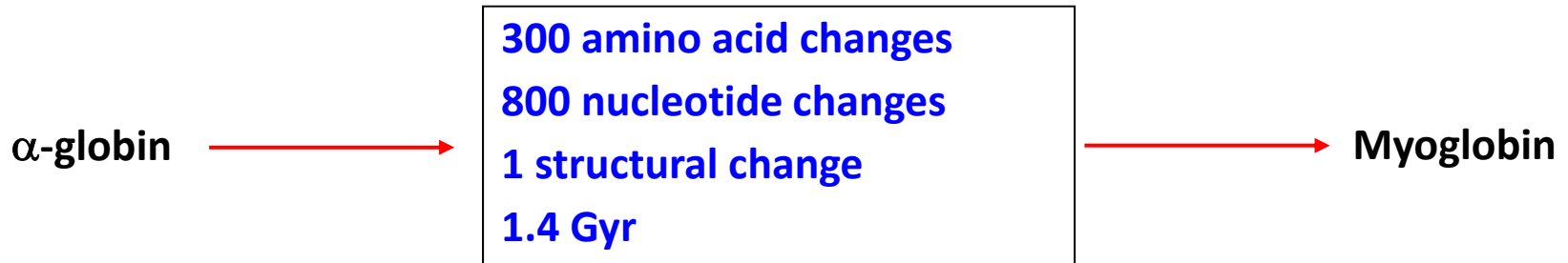
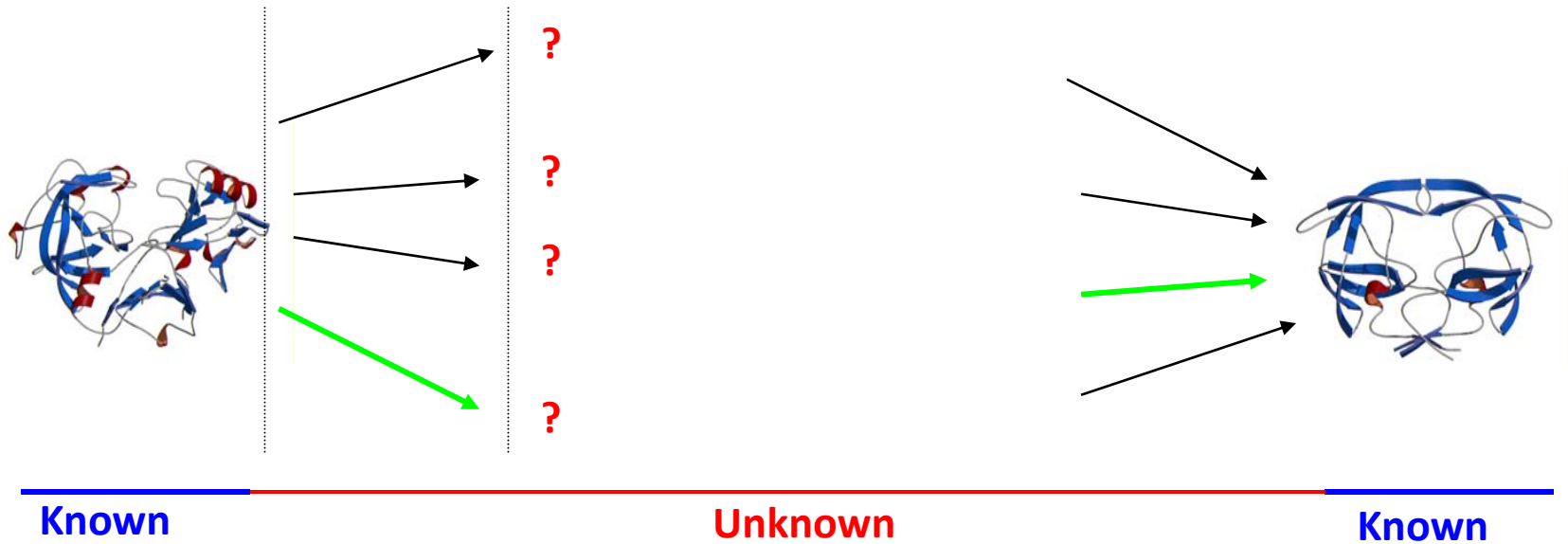
$$P(\text{Structure} \mid \text{Sequence})P(\text{Sequence})$$

## Needs:

i.  $P(\text{Sequence} \mid \text{Structure})$

ii.  $P(\text{Structure})$

# Protein Structure

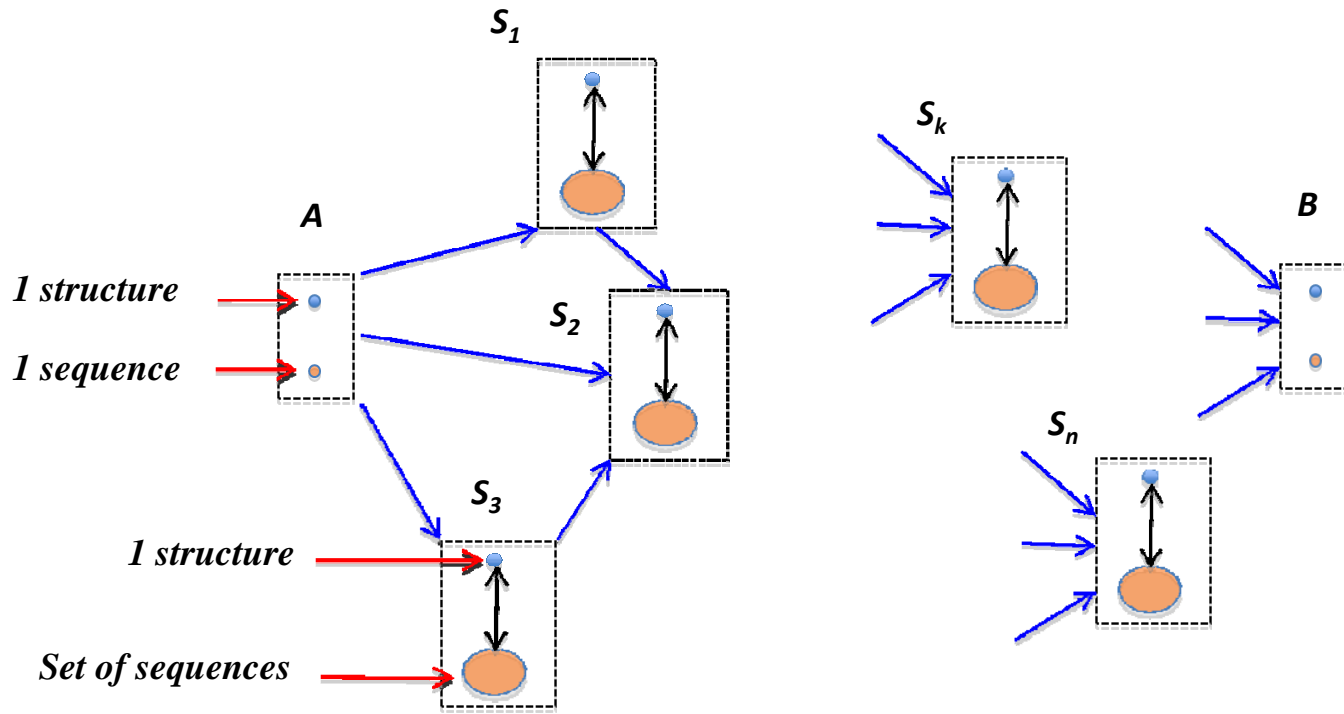


1. Given Structure what are the possible events that could happen?
2. What are their probabilities? Old fashioned substitution + indel process with bias.

Bias: Folding(**Sequence**  $\rightarrow$  **Structure**) & Fitness of Structure

3. Summation over all paths.

# The Mill Hill – Oxford Model



*Stepping Stones*

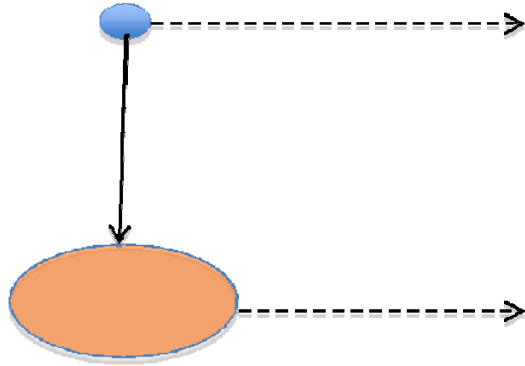
*Structures*

*Sequences Sets*

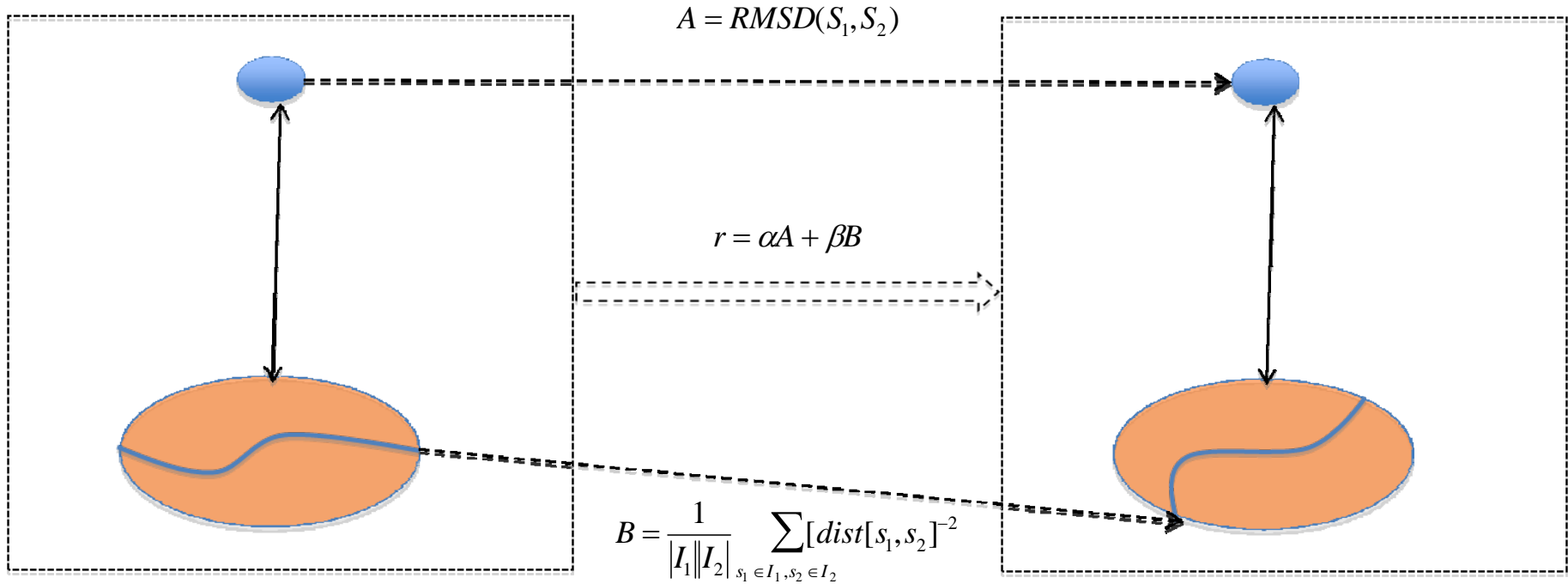
*Jump rates:*

# *Trajectories between two Secondary Structures*

# *The Structure of an Island*



# Jump rates between Islands



## *The probability of $A \rightarrow B$ , $P_{\Theta}(A \rightarrow B)$*

*Two structures have been observed and in principle  $P_{\Theta}(A, B)$  [=  $P_{\Theta}(A) * P_{\Theta}(A \rightarrow B)$ ] should be calculated. However, calculating  $P_{\Theta}(A)$  [the probability of randomly picked protein] is very difficult and present we must focus on  $P_{\Theta}(A \rightarrow B)$ . This can still give us information about different parameters. Especially time,  $t$ , and the relative importance of the sequence process and the structure constraints are interesting to estimate by maximum likelihood.*

# *A random walk in Structure Space fra A towards B*

# *The Evolution of Molecular Movements*

# *A Model for the Evolution of Metabolisms*

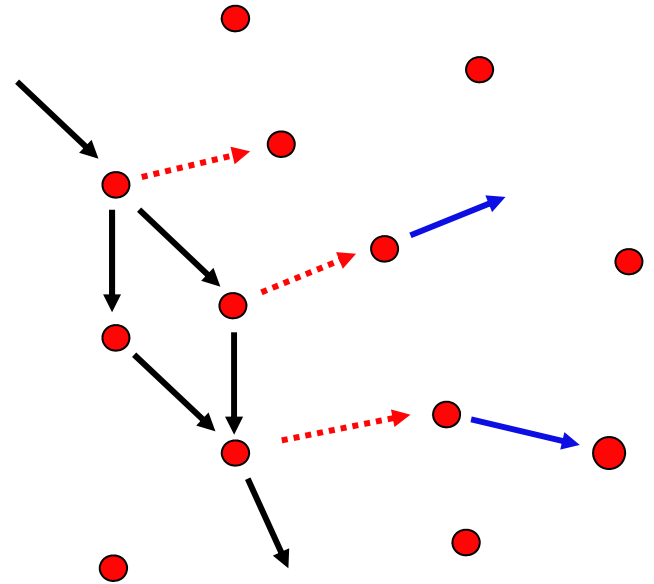
- A given set of metabolites: ●
- A given set of possible reactions -  
arrows not shown.
- A core metabolism: →
- A set of present reactions - **M**  
black and red arrows

## Restriction R:

A metabolism must define a connected graph

**M** + **R** defines

1. a set of deletable (dashed) edges **D(M)**: ⋯→
2. and a set of addable edges **A(M)**: →



Let  $\mu$  be the rate of deletion  
 $\lambda$  the rate of insertion

Then

$$\frac{dP(M)}{dt} = \lambda \sum_{M' \in D(M)} P(M') + \mu \sum_{M'' \in A(M)} P(M'') - P(M)[\lambda|D(M)| + \mu|A(M)|]$$

# A Toy Example

(by Aziz Mithani)

## Equilibrium Probability

### Metabolic Universe

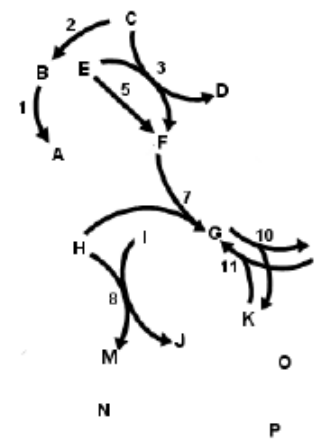
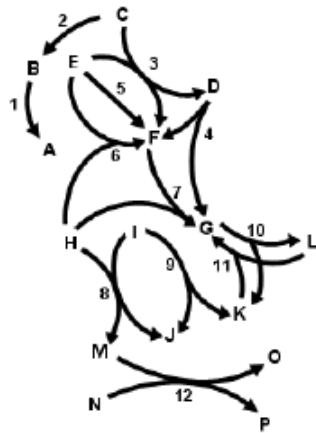
### 12 possible edges

1i 1u 3

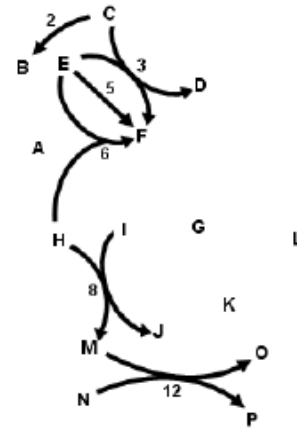
1i 2u 3

2u 1i 3

2i 2u 3



## Transition Probability



*dist=6*

## Transition Probability:

Full Exponentiation ( $2^{12}$  states 4096)

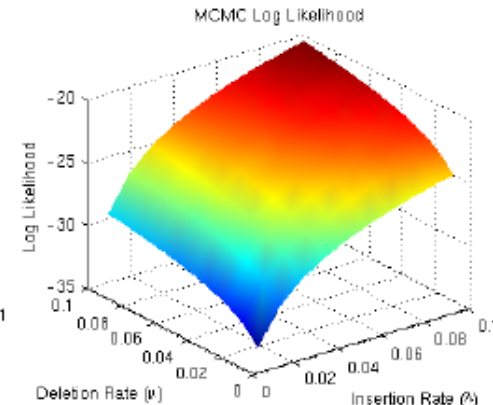
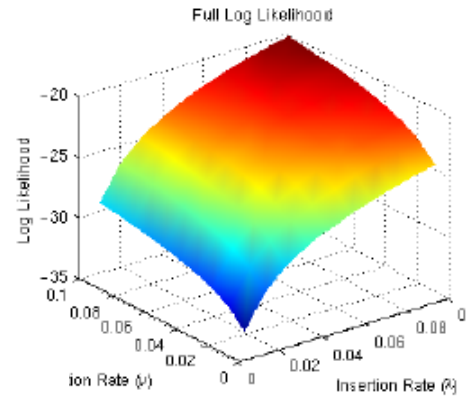
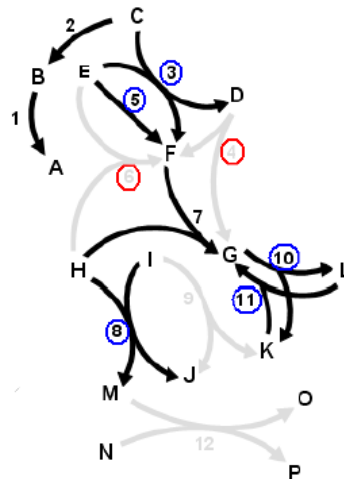
Exponentiation with corner cutting

$2^6 - 64, 384, 960, 1280, 960, 384, 64$

MCMC Integration

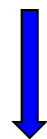
## Adding Connectedness

Favouring insertions connecting



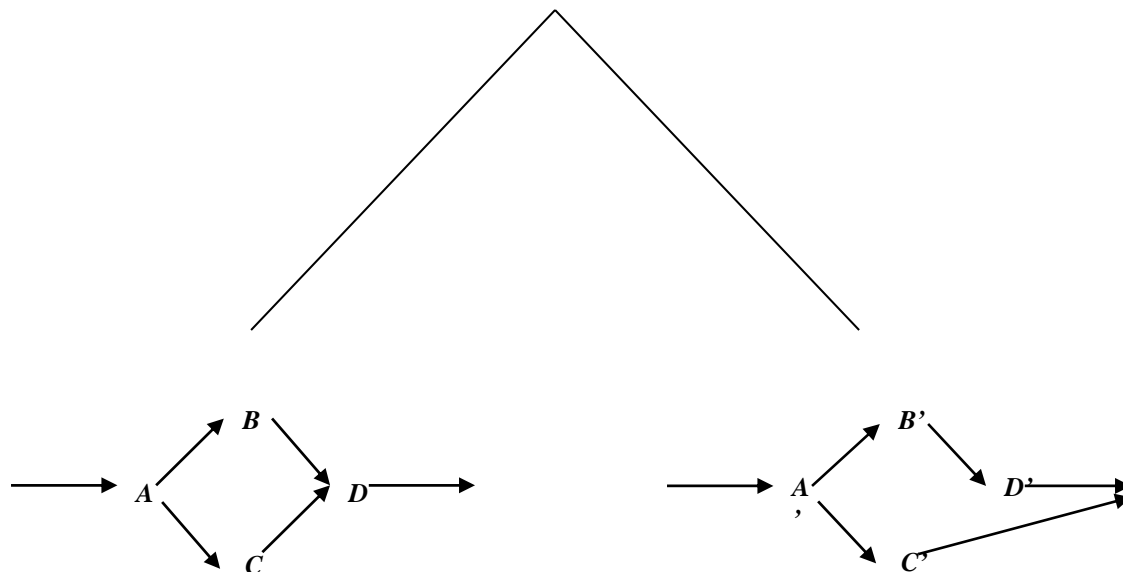
The proportion present:  $\frac{5}{7} = 0.714$

# Inference and Evolution



$$P(D_{mouse}, D_{human}) = P(D_{human} | N_{human}) P(D_{mouse} | N_{mouse}) P(N_{human}, N_{mouse})$$

Evolve



Infer network



Observe (data)

Human

Mouse

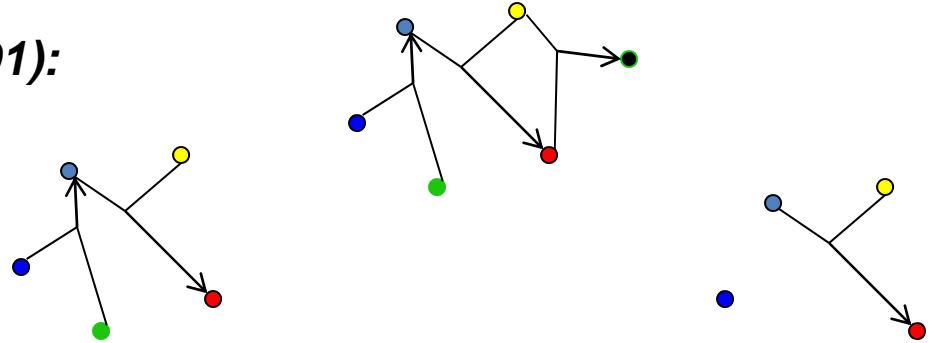
# Suggestion: Evolving Dynamical Systems

- *Goal: a time reversible model with sparse mass action system of order three!!*

**Adding/Deleting components (TKF91):**

**Add rate:  $(k+1)\lambda$**

**Delete rate:  $k\mu$**

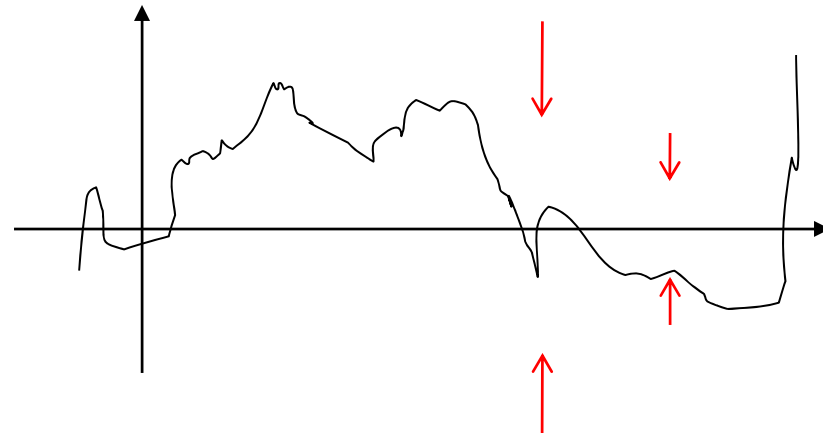


**Adding reactions with birth of component:**

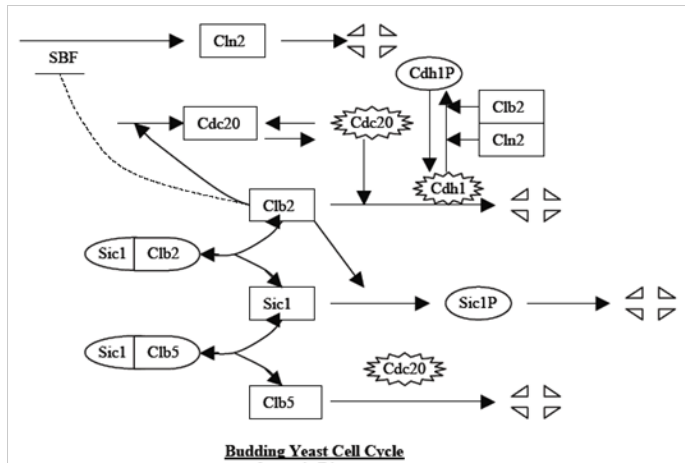
**There are  $3k(k-1)$  possible reactions involving a new-born**

**Reaction Coefficients:**

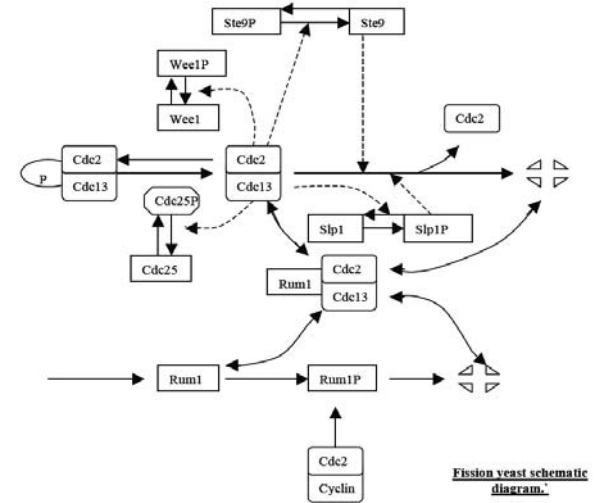
- *Continuous Time Continuous States Markov Process - specifically Diffusion.*
- *For instance Ornstein-Uhlenbeck, which has Gaussian equilibrium distribution*



# Network Example: Cell Cycle



*Evolve!*



|                           |      |       |      |       |      |       |       |       |       |
|---------------------------|------|-------|------|-------|------|-------|-------|-------|-------|
| Budding (D <sub>1</sub> ) | Clb5 | Clb2  | Cdh1 | Cdc20 | Sic1 | Cln2  | SBF   | (N/A) | (N/A) |
| Fission (D <sub>2</sub> ) | Cig2 | Cdc13 | Ste9 | Slp1  | Rum1 | (N/A) | (N/A) | Cdc25 | Wee1  |

- *What is the edit distance?*
- *Which properties are conserved?*
- *If you only knew Budding Yeast, how much would you know about Fission Yeast?*
- *As N1 starts to evolve, you can only add reactions. Isn't that strange?*
- *On a path from N1 to N2 how close to the minimal has evolution travelled?*
- *What is the number of equation systems possible for N1?*

# *Metabolisms on a phylogeny*

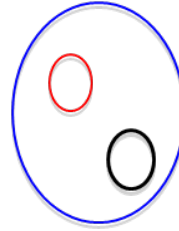
# MCMC and Gibbs Sampler

Choose neighborhood  $N$ .  $M-N$  will be fixed.

Calculate  $P(\{M_1 \rightarrow M_2\} \text{ intersection } N | \{M_1 \rightarrow M_2\} \text{ intersection } M-N)$  using MCMC.

Choose new  $N$  etc

**Metabolism 1**



*Evolutionary trajectory*



*MCMC integration over trajectories*

**Metabolism 2**

