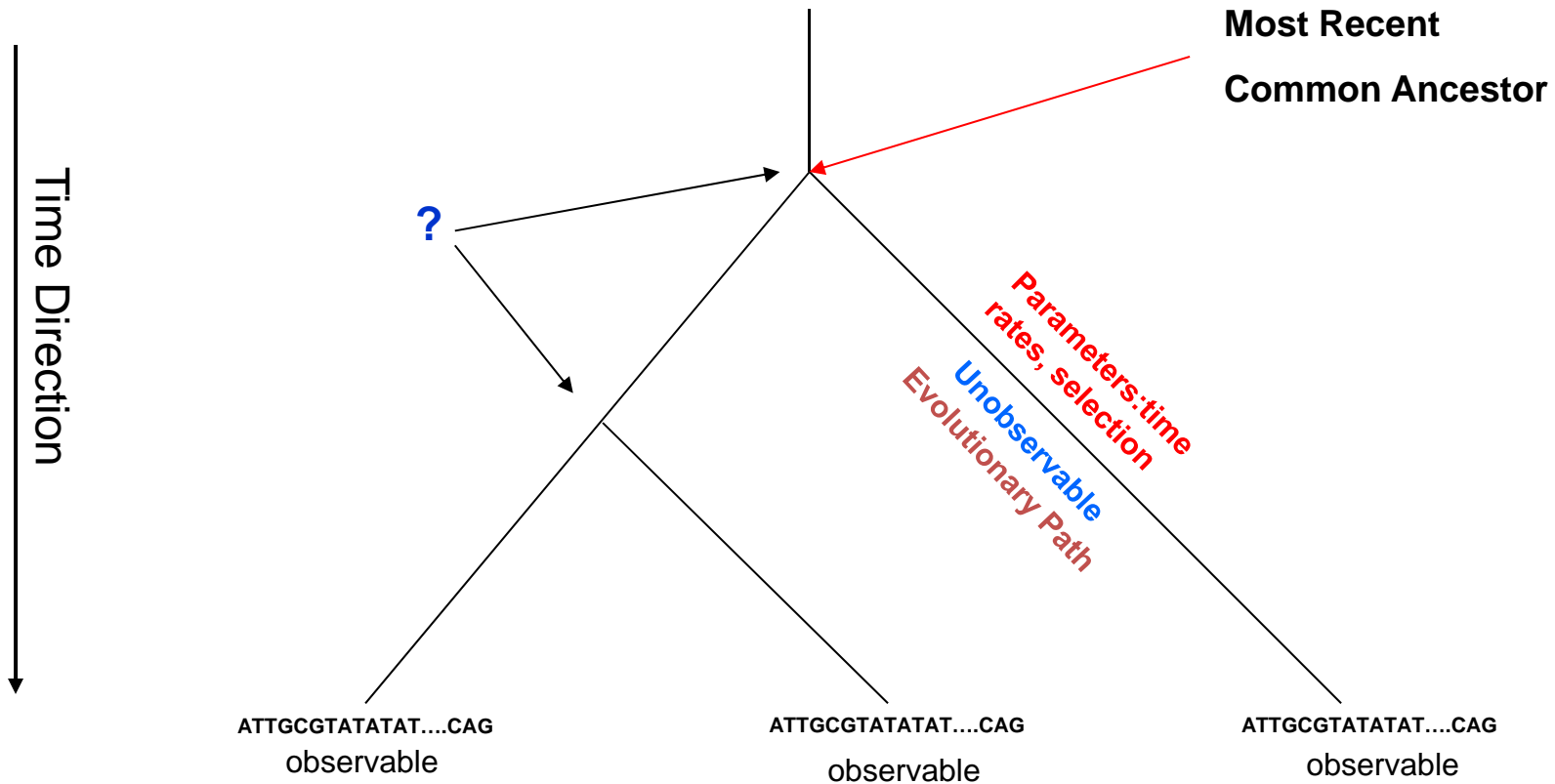


Comparative Biology with focus on 8 examples

- *Comparative Biology*
- *The Domain of Comparative Biology*
- *Co-modeling in Comparative Biology*
- *The purpose of Comparative Biology*
- *Examples of Stochastic Comparative Modeling*
 - ***Gene Frequencies in Populations***
 - ***Genome Structure Evolution***
 - ***Stemmatology: Manuscript Evolution***
 - ***RNA Secondary Structure Evolution***
 - ***Protein Structure Evolution***
 - ***Movement Evolution***
 - ***Shape Evolution***
 - ***Pattern 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	CTNS continuous time continuous states	Felsenstein 68 + 50 others
Sequences	CTUS continuous time countable states	Thorne, Kishino Felsenstein,91 + 40others
Gene Structure	Matching	DeGroot, 07
Genome Structure	CTCS MM	Miklos,
Population	Brownian Motion/Diffusion	Fisher, Wright, Haldane, Kimura,
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	CTFS	Mithani, 2009a,b
Protein Interaction	CTCS	Stumpf, Wiuf, Ideker
Regulatory Pathways	CTCS	Quayle and Bullock, 06, Teichmann
Signal Transduction	CTCS	Soyer et al.,06
Macromolecular Assemblies	?	
Motors	?	
Shape	- (non-evolutionary models)	Dryden and Mardia, 1998, Bookstein, Jones & Moriarty
Patterns	- (non-evolutionary models)	Turing, 52;
Tissue/Organs/Skeleton/....	- (non-evolutionary models)	Grenander,
Dynamics		
MD movements of proteins	-	Biggins 05, Munz 10,
Locomotion	-	
Culture	analogues to genetic models	Cavalli-Sforza & Feldman, 83
Manuscripts (stemmatology)	analogous to sequence models	Chris J Howe, http://www.cs.helsinki.fi/u/ttonteri/case/
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	-	

The Purpose of Comparative Biology

To describe evolution:

- *Make realistic model (pass goodness-of-fit (GOF) test)*
- *Estimate Parameters*
- *Make statements about the path of evolution – **ancestral analysis***

Analyse homologous pairs or sets

- *What is the equilibrium distribution*
- *Integrate over histories*

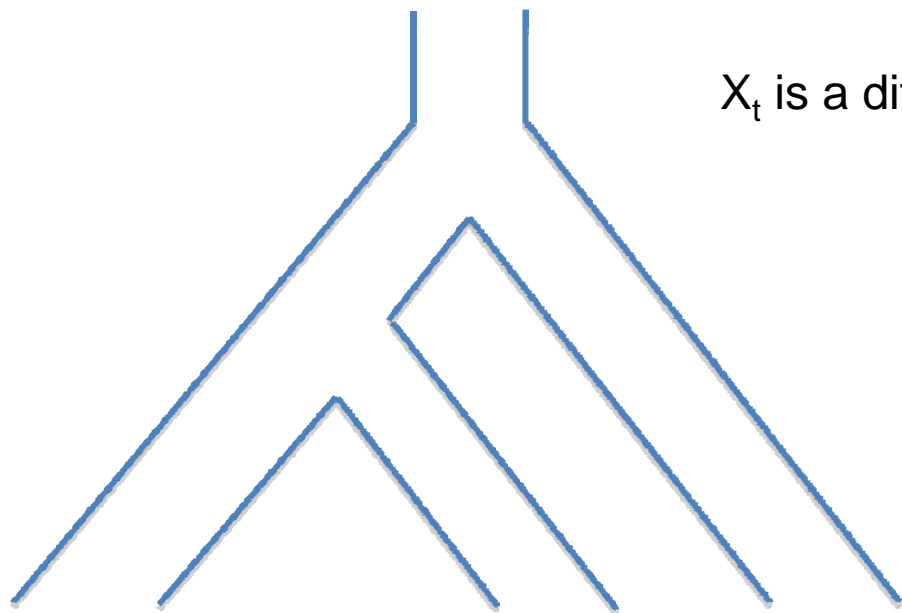
Biological Questions:

- *Rate of Evolution*
- *Heterogeneity*
 - Time*
 - State Space*
- *Selection*
- *Co-Evolution of different components within a level*
- *Dependence among different levels (co-modelling)*

Most of these questions have not been addressed beyond the sequence level:

- *Primarily due to lack of data*
- *Secondarily due to lack of models*

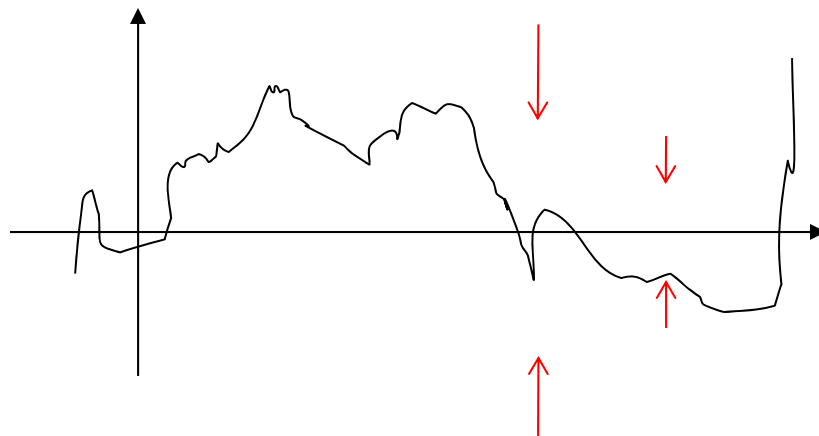
Population Gene Frequencies



X_t is a diffusion with $\mu(x)=0$ and $\sigma(x)=x(1-x)$

Famous Models:

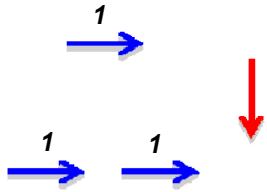
- *Continuous Time Continuous States Markov Process - specifically Diffusion.*
- *For instance Ornstein-Uhlenbeck, which has Gaussian equilibrium distribution*



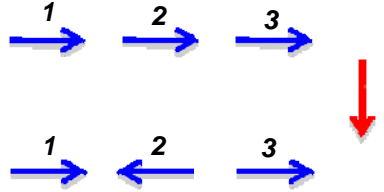
Genome Structure Evolution

- **Evolutionary events:**

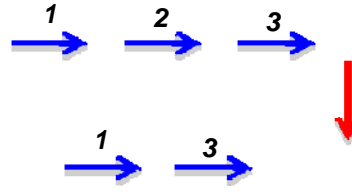
Duplication



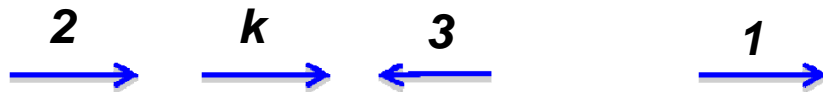
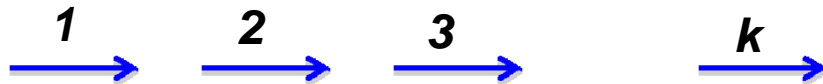
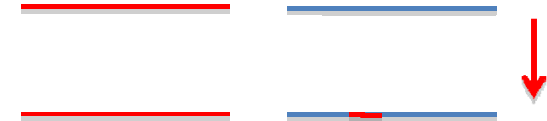
Inversion



Deletion



Transposition



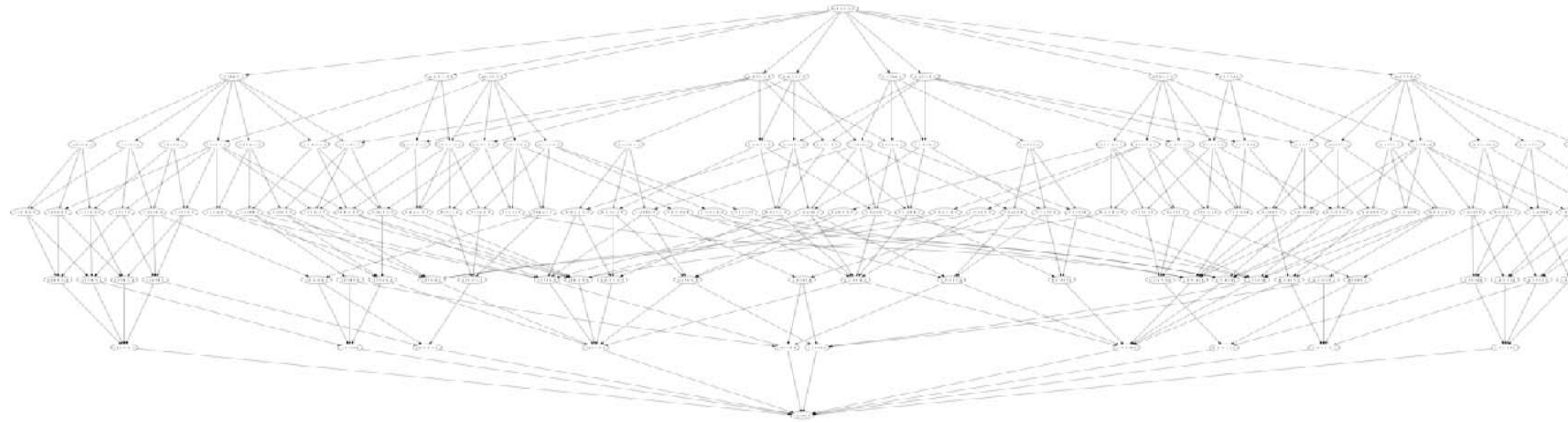
- **Inference Principles**

- *Shortest Path (Parsimony)*
- *Sum over paths with probabilities (ML)*

- **Extensions:**

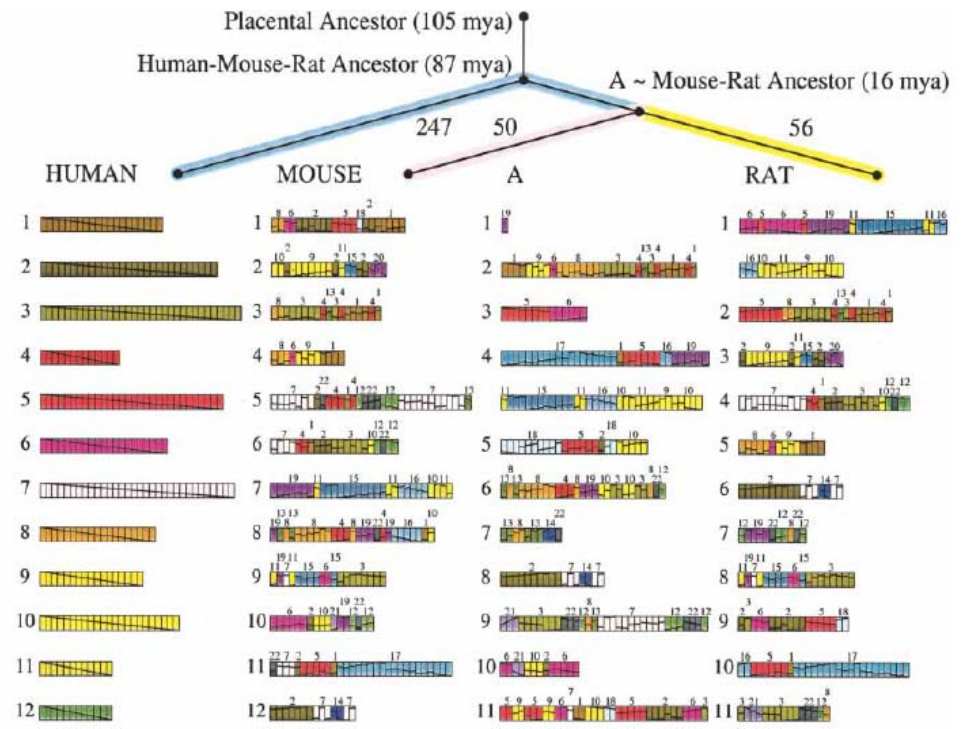
- *Directions of Genes Unknown*
- *A set of chromosomes related by a phylogeny*

Genome Structure Evolution



- **Full graph for 5 genes**

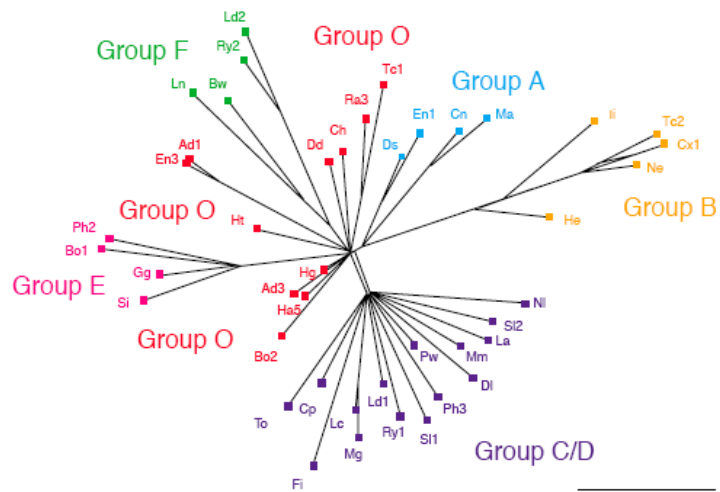
- **Genomic reconstruction for human, mouse and rat.**



Stemmatology: Evolution of Manuscripts

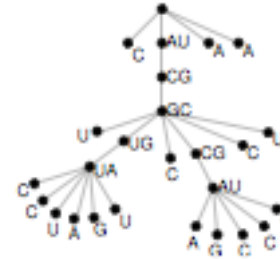
Ashmole 59	Buryed at Caane thus seythe the Croniculer
Digby 186	Beryed att Cane & thus says the cronyclere
BL Ad 31042	Beryed at caene so seyth the cronyclere
Lansd. 762	Buried at cane this saith the croneclere
de Worde	And is buried at Cane as the Cronycle sayes
R. Wyer	And buryed at cane as the Cronycle sayes

Phylogeny of "Canterbury Tales":

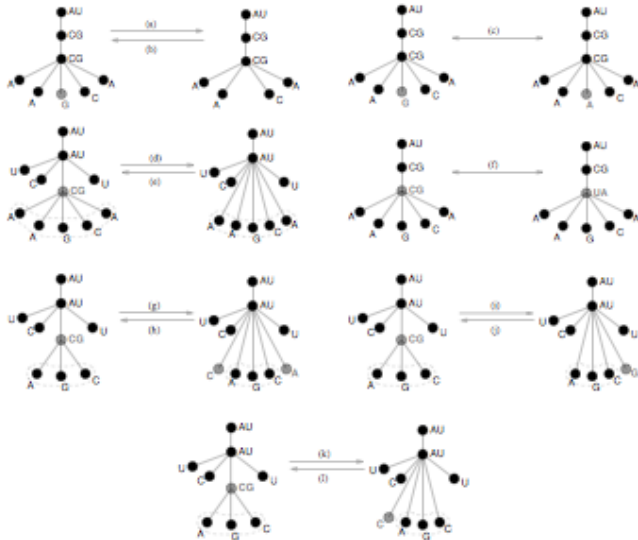


RNA Structure Evolution

Tree Representations of RNA Structure



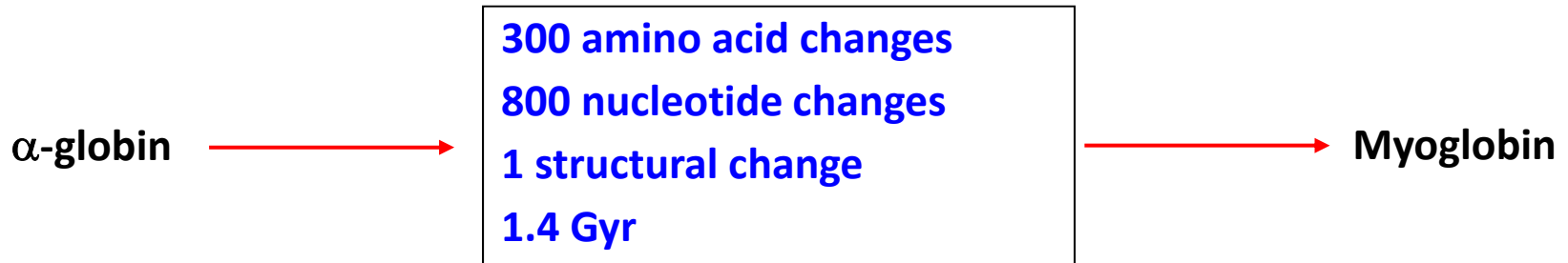
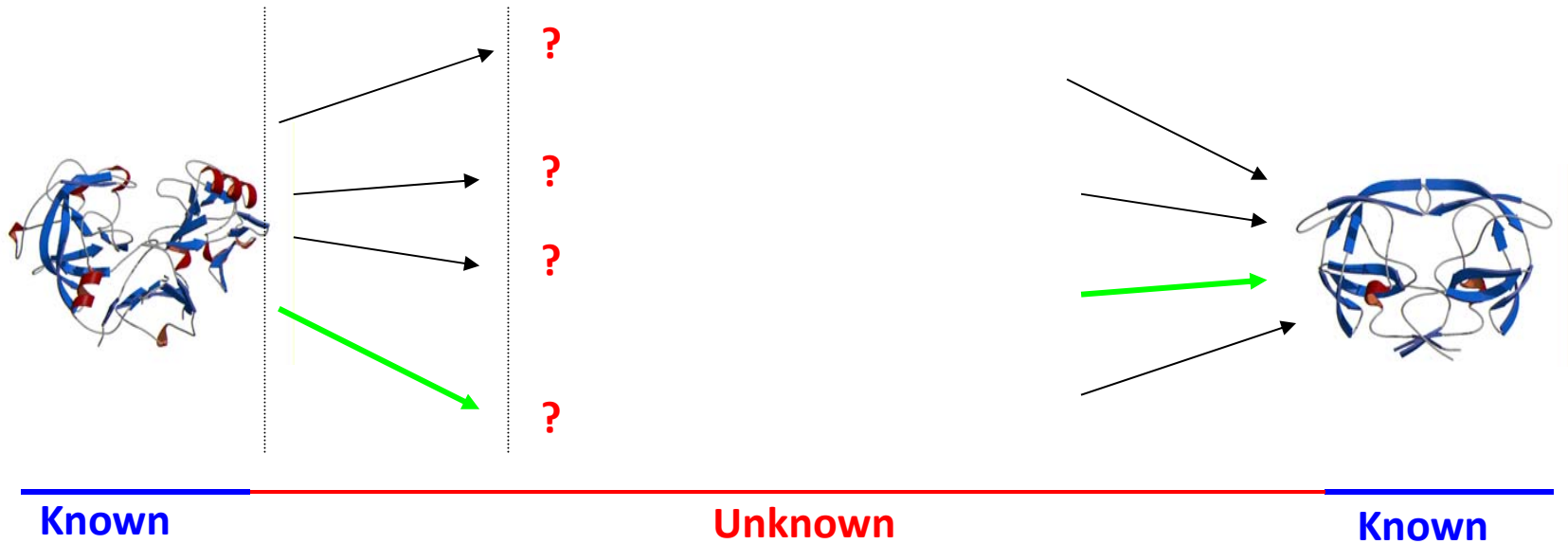
Basic Edit Operations



A Tree Distance Pairwise Edit Algorithm



Protein Structure Evolution



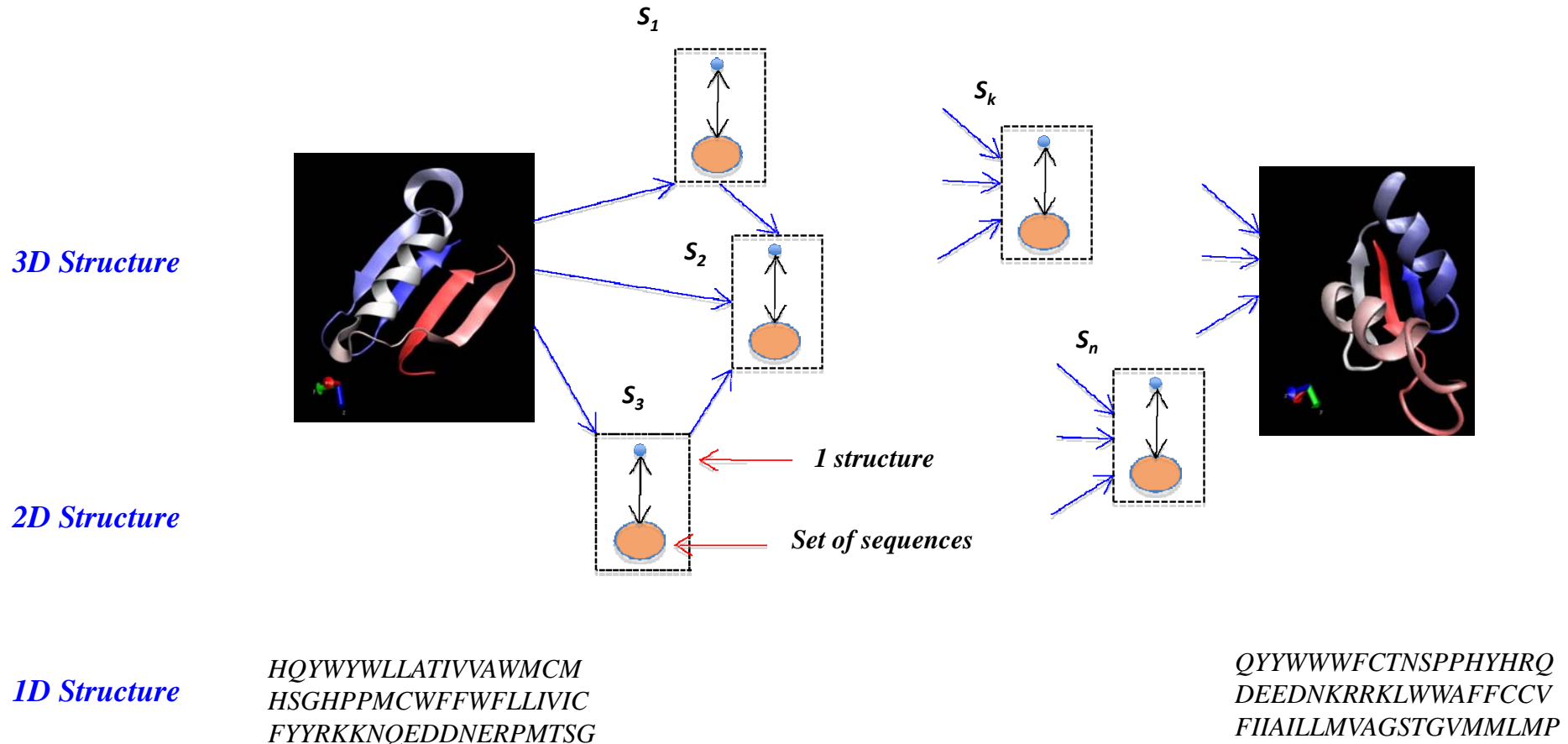
1. Given Structure what are the possible events that could happen?
2. What are their probabilities? Old fashioned substitution + indel process with bias.

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

3. Summation over all paths.

Trajectories between two Secondary Structures

- *Observation: two structures with sequence and secondary structure information*
- *Space of Protein Structures is large and complicated – both continuous and discrete*
- *Approximated by a series of stepping stones and a continuous time markov chain*



The Evolution/Comparison of Molecular Movements

Molecular Movements of Homologous Proteins are themselves homologous

The full problem: 2 times 1000 atoms observed at 10^6 time points.

Reductions:

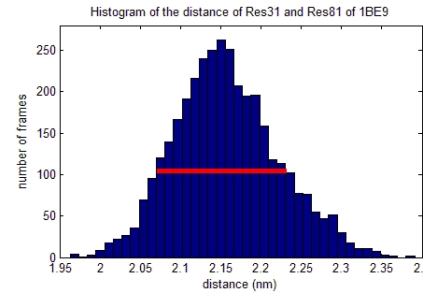
- i. only α -carbons \rightarrow 100 space points
- ii. Only correlated pairwise movements \rightarrow 1 dimensional summary for each aa pair



$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$$

$$D = \begin{pmatrix} \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & d_{ij} & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}$$

Dynamic Fingerprint Matrix (DFM)

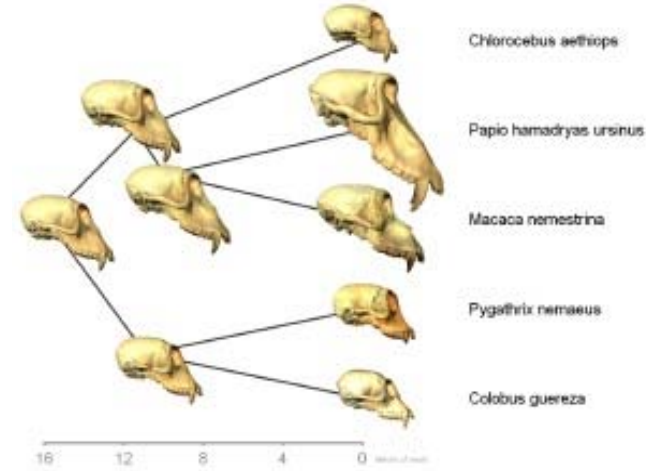
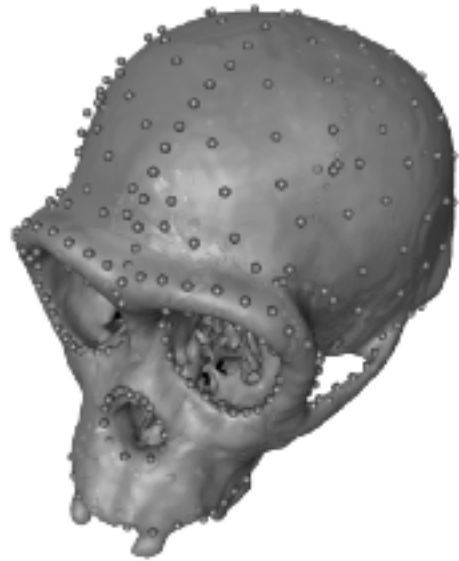


$$A_{ij} = \sigma\{d_{ij}(t)\}$$

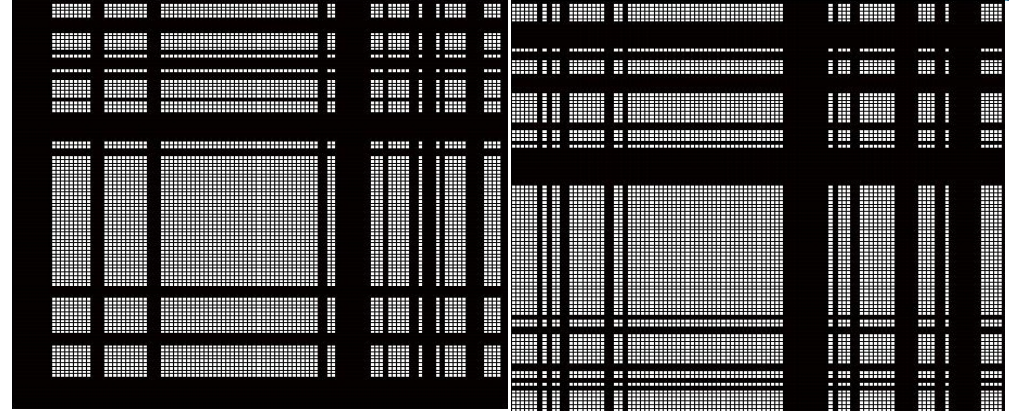
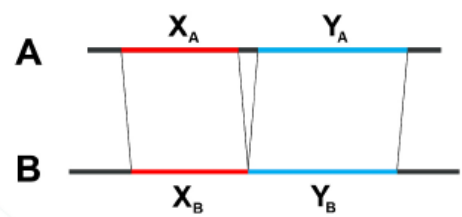
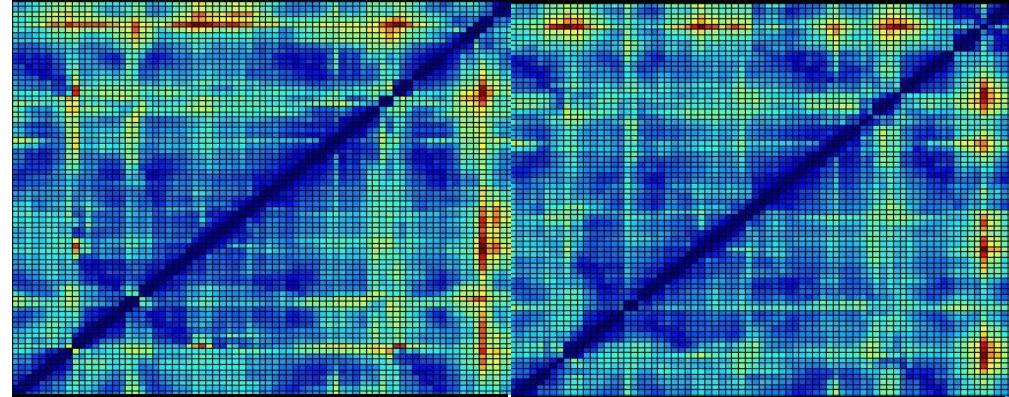
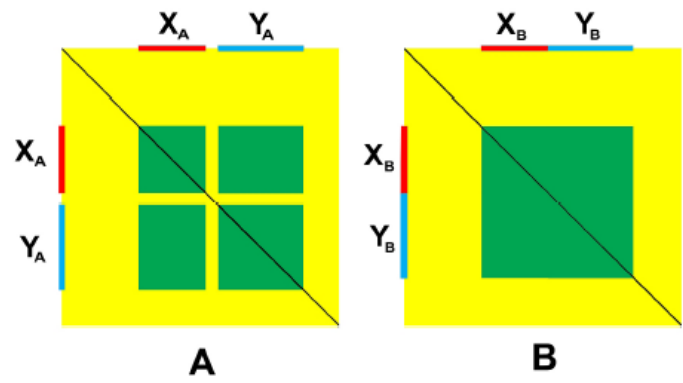
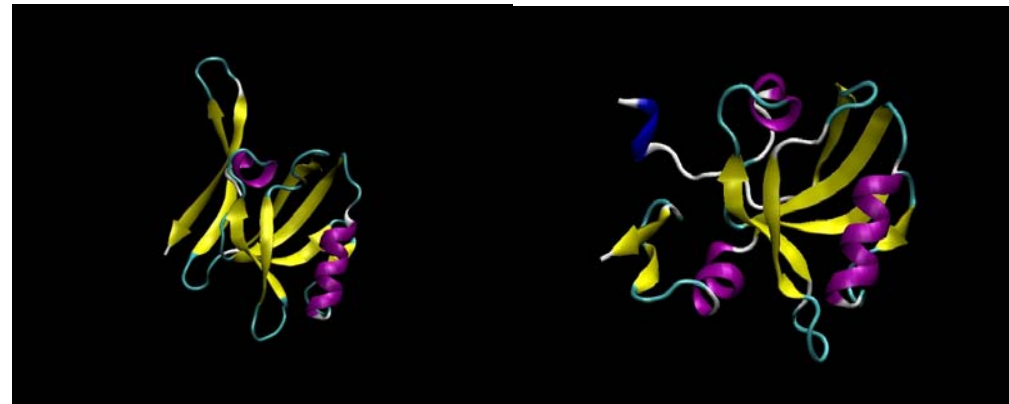
$$A = \begin{pmatrix} \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & A_{ij} & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}$$

Shapes and Shape Evolution

- Landmarks
- Semilandmarks



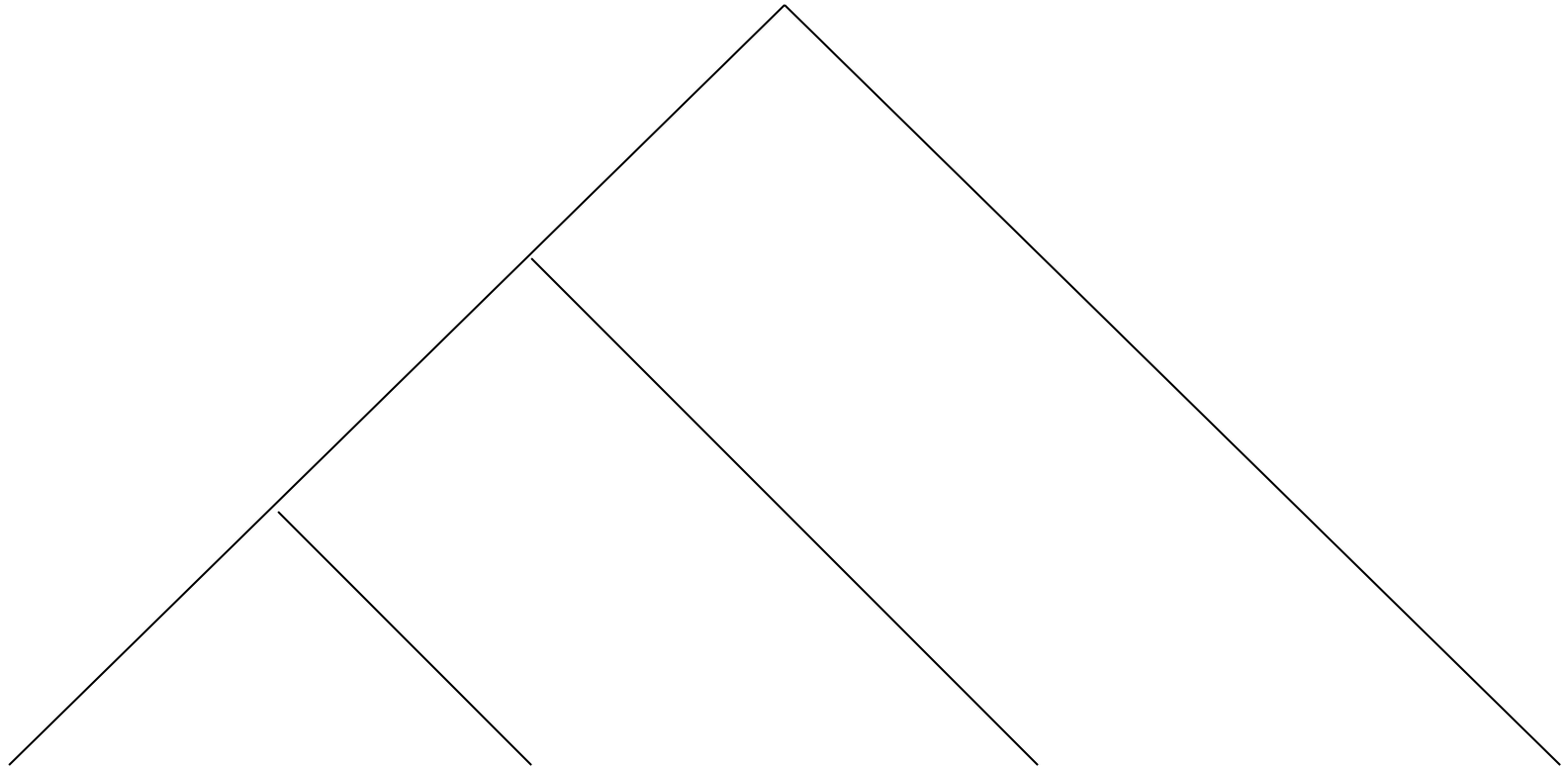
The Evolution/Comparison of Molecular Movements



$$S_{\alpha\beta}^{AB} = - \sum_{i \in \alpha} \sum_{j \in \beta} s(i, j)$$

1be9/1-83 R P R R I V I X R T G L G F N I I G G E G E G I F I S F I L A G G P A D L S G E L R X G D Q I L S V N G V D L R N A S X E Q G Q T V T I I A X P S F E A N S R V N S S
 1gau/1-83 N V I S V R F R V G L G F L V X E R V S X P V I I S D L I R G G A A E Q S G L I Q A G D I I L A V N D R L D S Y D R G I A S T X V L I L R G T X L E X T I R V T Q

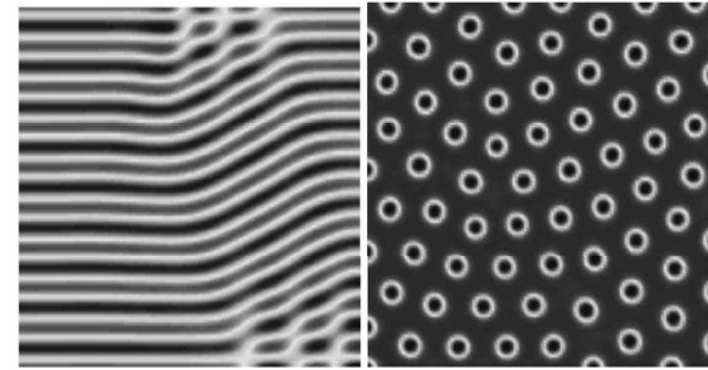
The Phylogenetic Turing Patterns I



The Phylogenetic Turing Patterns II

Reaction-Diffusion Equations:

$$\begin{cases} u_t = D_u \nabla^2 u + \gamma(u + av - uv^2 - puv) \\ v_t = D_v \nabla^2 v + \gamma(bv + hu + uv^2 + puv) \end{cases}$$



Stripes: p small

Spots: p large

Analysis Tasks:

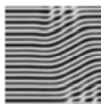
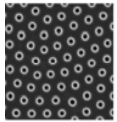
1. Choose Class of Mechanisms
2. Observe Empirical Patterns
3. Choose Closest set of Turing Patterns T_1, T_2, \dots, T_k ,
4. Choose parameters p_1, p_2, \dots, p_k (sets?) behind $T_{1..}$

Evolutionary Modelling Tasks:

1. $p(t_1) - p(t_2) \sim N(0, (t_1 - t_2)\Sigma)$
 2. Non-overlapping intervals have independent increments
- I.e. Brownian Motion

Scientific Motivation:

1. Is there evolutionary information on pattern mechanisms?
2. How does patterns evolve?



Summary

- *Comparative Biology*
- *The Domain of Comparative Biology*
- *Co-modeling in Comparative Biology*
- *The purpose of Comparative Biology*
- *Examples of Stochastic Comparative Modeling*
 - *Gene Frequencies in Populations*
 - *Genome Structure Evolution*
 - *Stemmatology: Manuscript Evolution*
 - *RNA Secondary Structure Evolution*
 - *Protein Structure Evolution*
 - *Movement Evolution*
 - *Shape Evolution*
 - *Pattern Evolution*