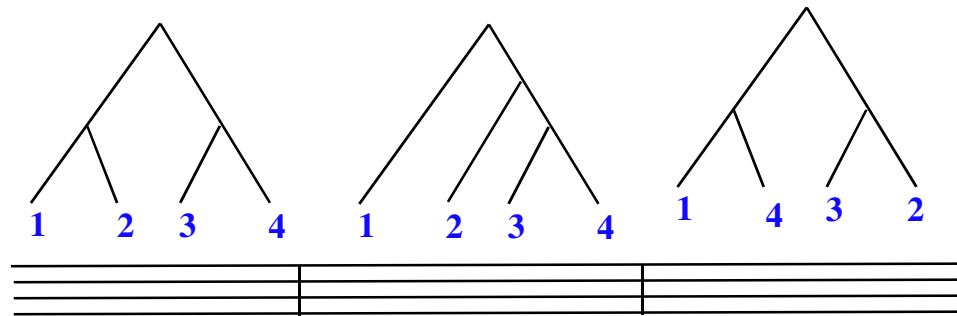
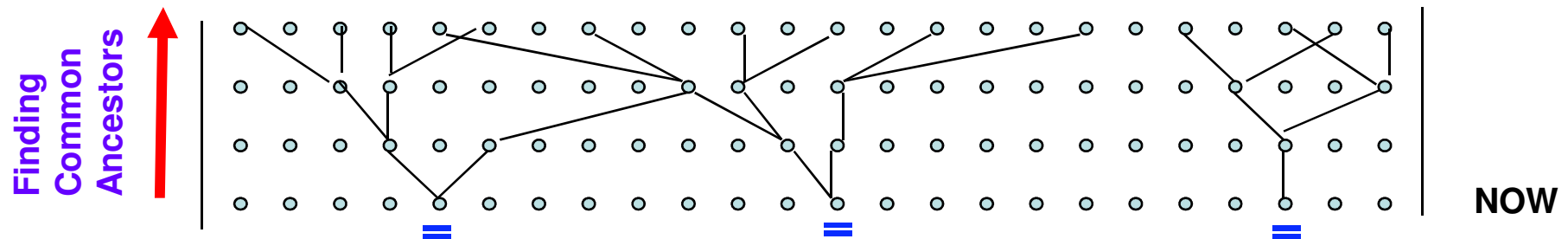


Recombination Histories & Global Pedigrees

Finding Minimal Recombination Histories



Global Pedigrees

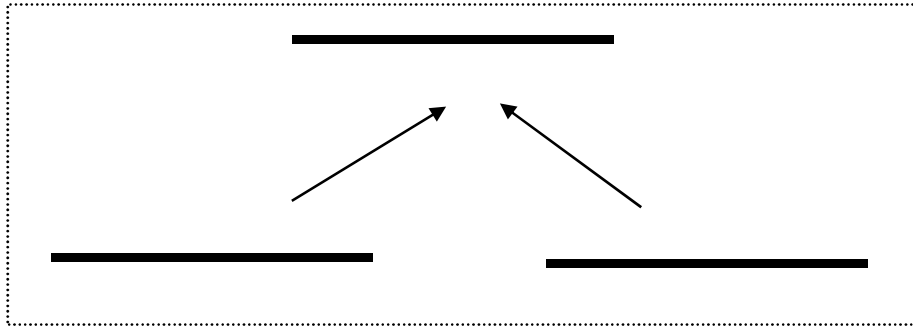


Acknowledgements

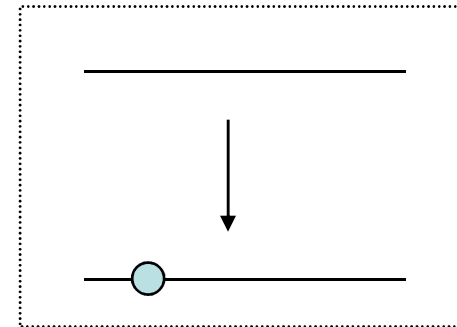
Yun Song - Rune Lyngsø - Mike Steel

Basic Evolutionary Events

Coalescent/Duplication

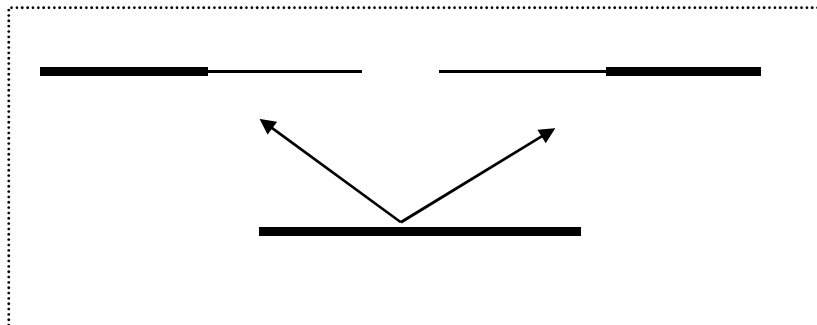


Mutation

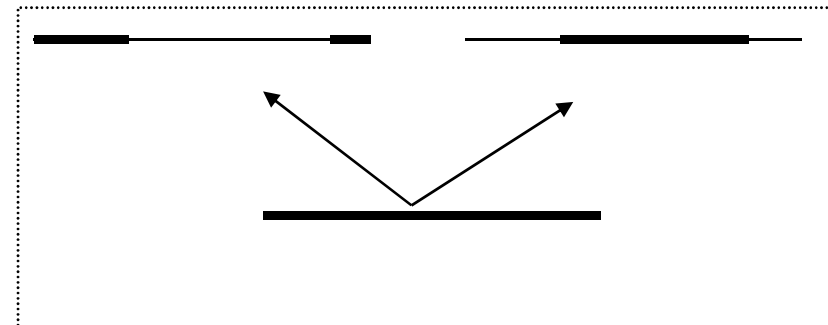


Infinite site assumption ?

Recombination

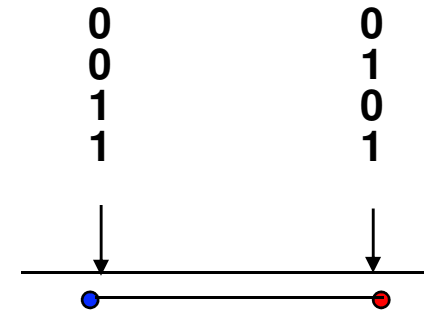
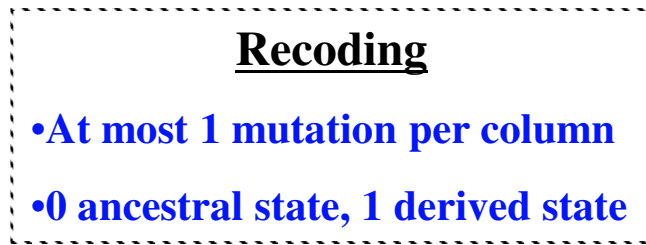


Gene Conversion

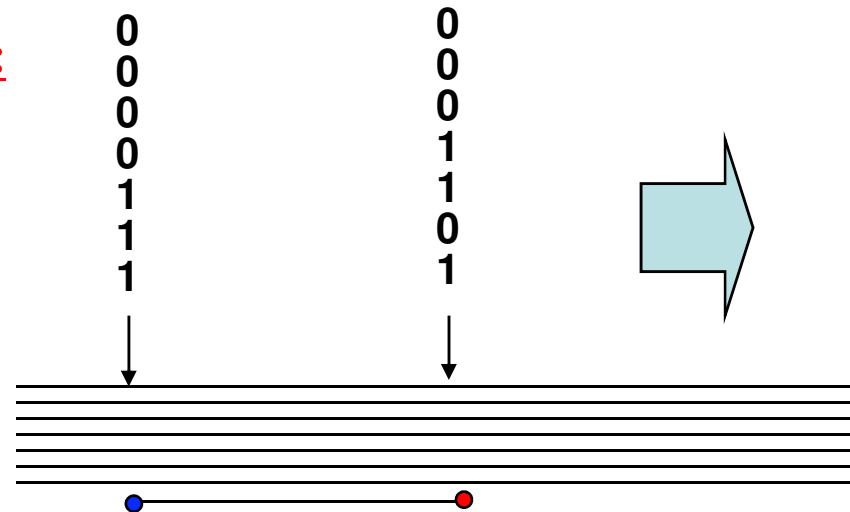


Local Inference of Recombinations

T...G
T...C
A...G
A...C



Incompatibility:



Four combinations

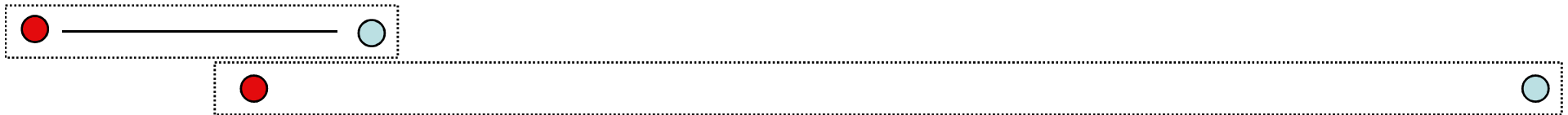
00
10
01
11

Myers-Griffiths (2002): Number of Recombinations in a sample, N_R , number of types, N_T , number of mutations, N_M obeys:

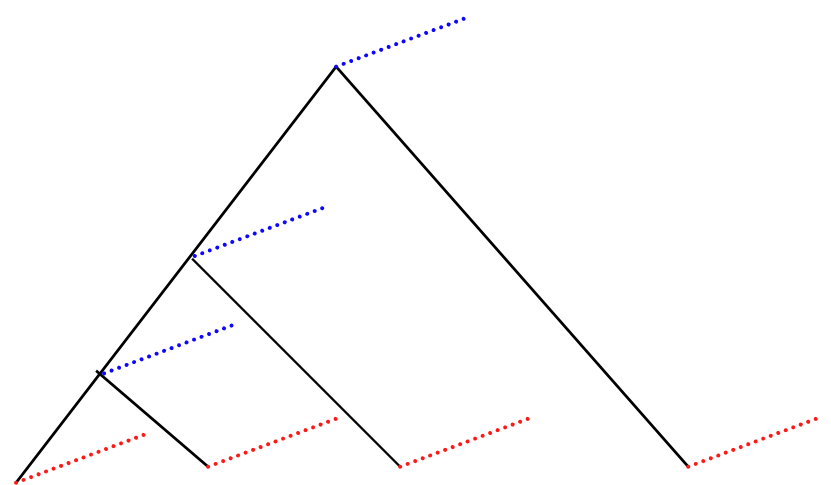
$$N_R \geq N_T - N_M - 1$$

”Observing” Recombinations: Hudson & Kaplan’s R_M

```
00000100000011111111110000000000
00100000000011111111110000000000
0000000000000000000000000000000010
0000000000110000000000000000000010
000111111111000000000000000000001
001000000100000000000000001010111
001000000100000000000000011111101
111111111000000000000000011111101
111111100100000000000000011111101
```



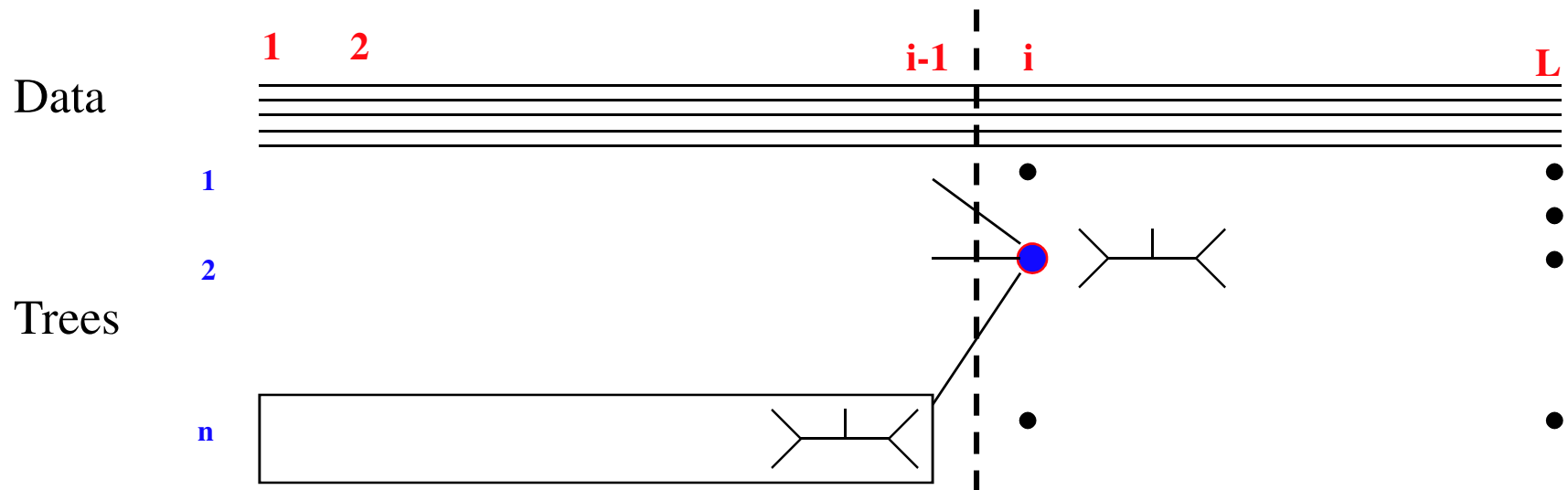
If you equate R_M with expected number of recombinations, this could be used as an estimator. Unfortunately, R_M is a gross underestimate of the real number of recombinations.



Minimal Number of Recombinations

The Kreitman data (1983): 11 sequences, 3200bp, 43(28) recoded, 9 different

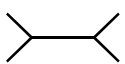
Last Local Tree Algorithm:




How many local trees?

How many neighbors?

Bi-partitions

• **Unrooted**  $\frac{(2n-2)!}{2^{n-1}(n-1)!}$

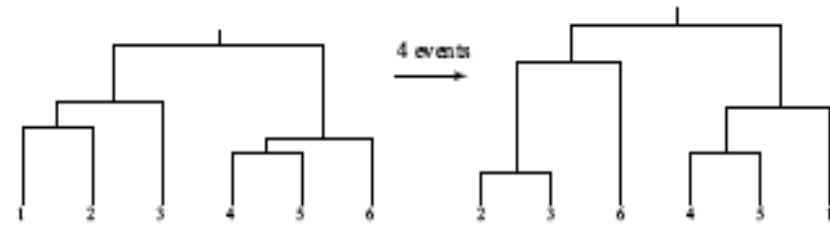
$$3n^2 - 13n + 14$$

• **Coalescent**  $\frac{n! (n-1)!}{2^{n-1}}$

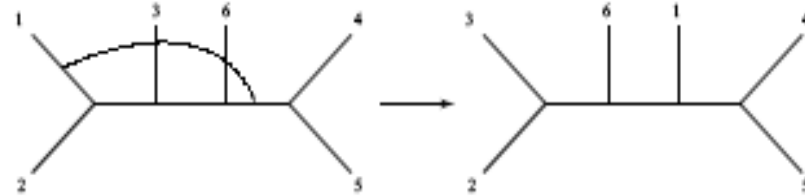
$$\sim n^3$$

Metrics on Trees based on subtree transfers.

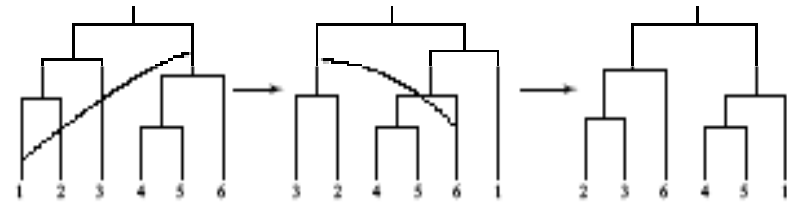
Trees including branch lengths



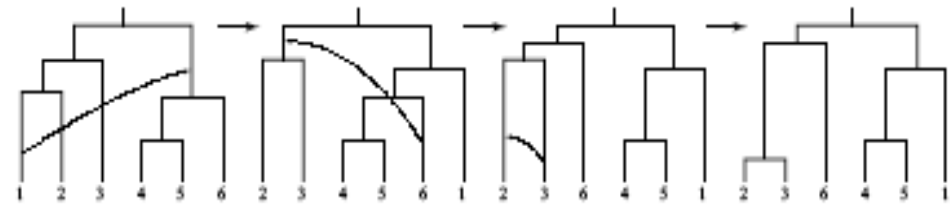
Unrooted tree topologies



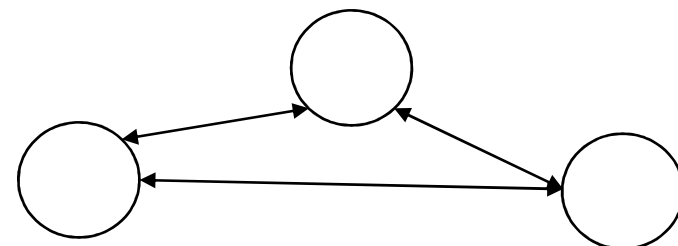
Rooted tree topologies



Tree topologies with age ordered internal nodes

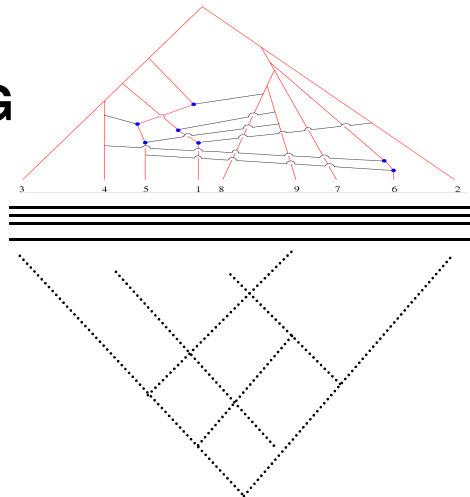


Pretending the **easy** problem (unrooted) is the **real** problem (age ordered), causes violation of the triangle inequality:

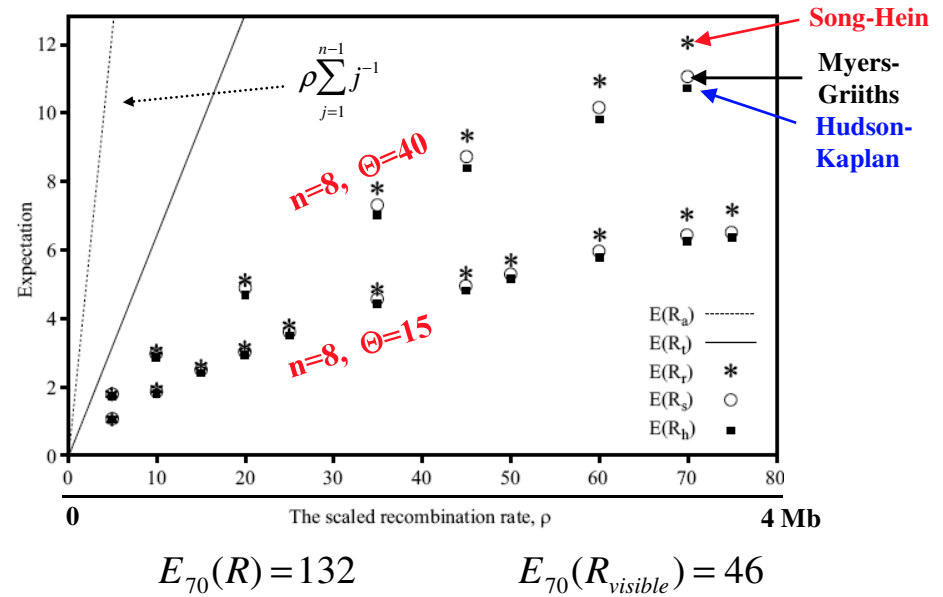


minARGs: Recombination Events & Local Trees

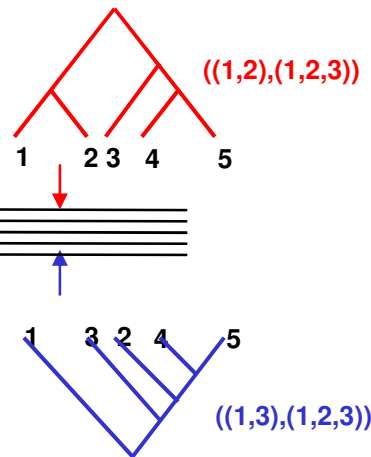
Minimal ARG



True ARG

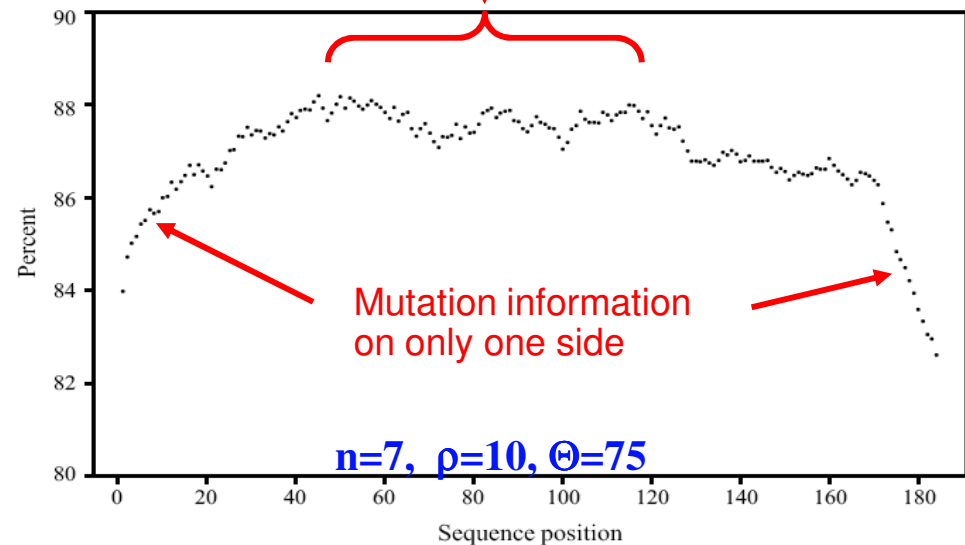


True ARG



Reconstructed ARG

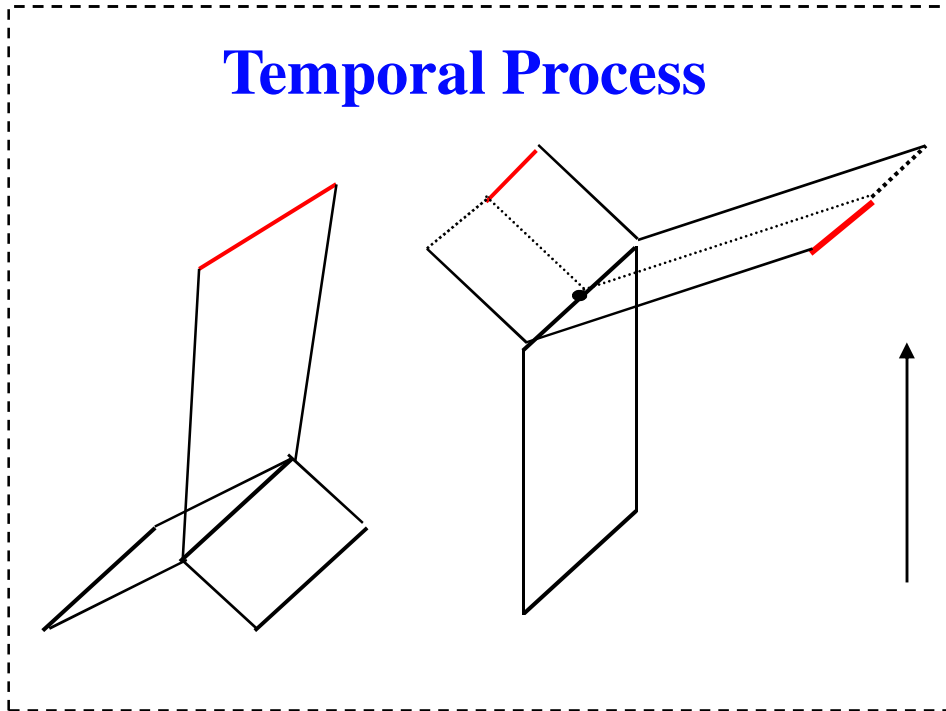
Mutation information on both sides



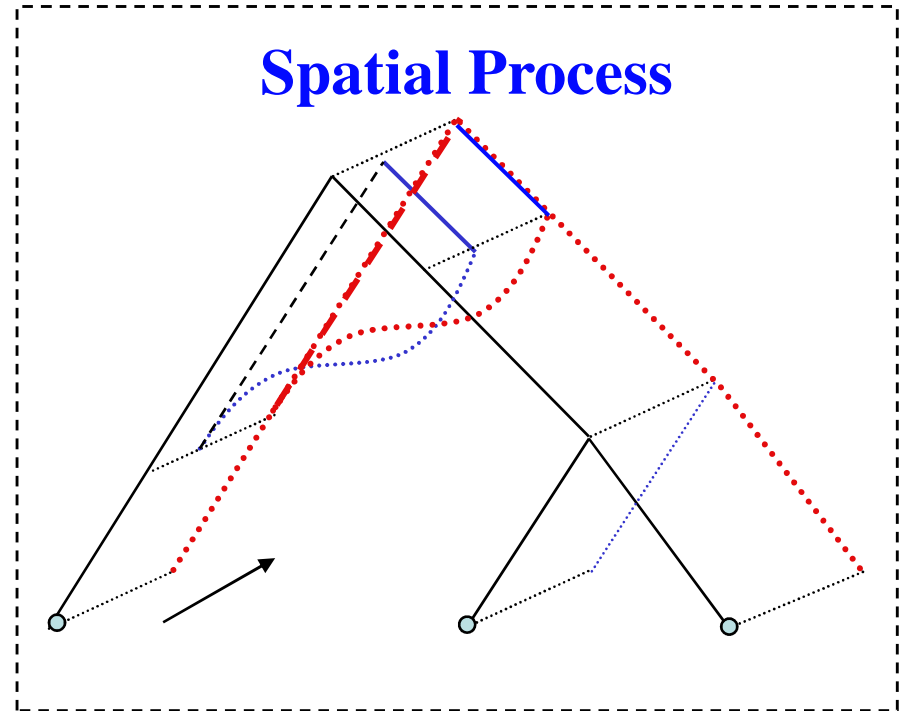
Time versus Spatial 1: Coalescent-Recombination

(Griffiths, 1981; Hudson, 1983 - Wiuf & Hein, 1999)

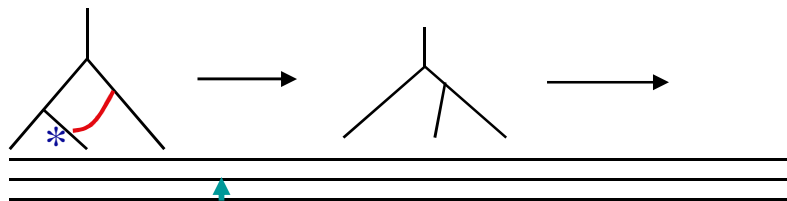
Temporal Process



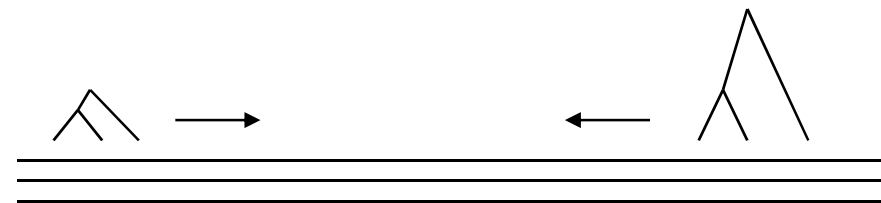
Spatial Process



i. The process is non-Markovian

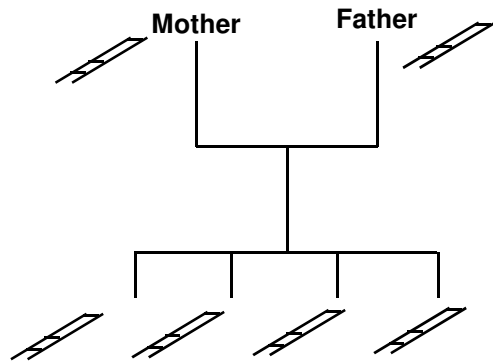


ii. The trees cannot be reduced to Topologies



Time versus Spatial 2: Pedigrees

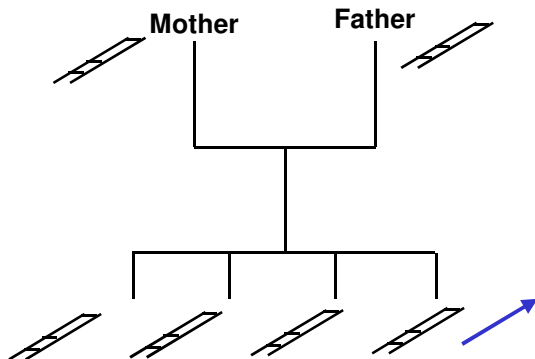
Elston-Stewart (1971) - Temporal Peeling Algorithm:



Condition on parental states

Recombination and mutation are Markovian

Lander-Green (1987) - Genotype Scanning Algorithm:



Condition on paternal/maternal inheritance

Recombination and mutation are Markovian

Time versus Spatial 3: Phylogenetic Alignment

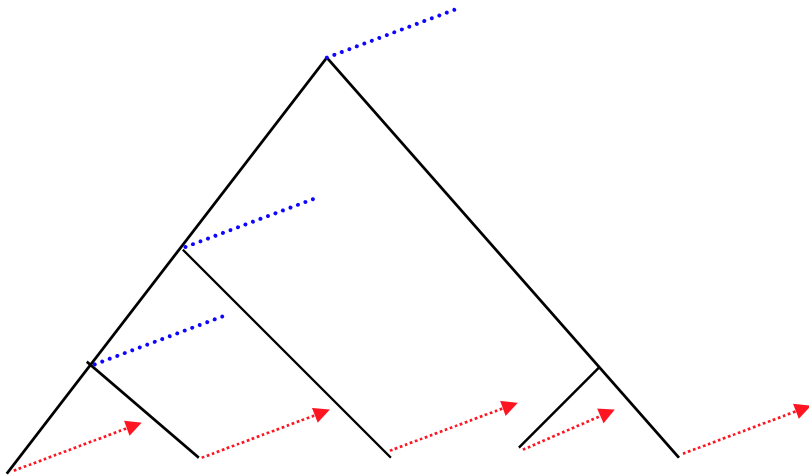
- **Optimisation Algorithms**

 - indels of length 1 (David Sankoff, 1973) **Spatial**

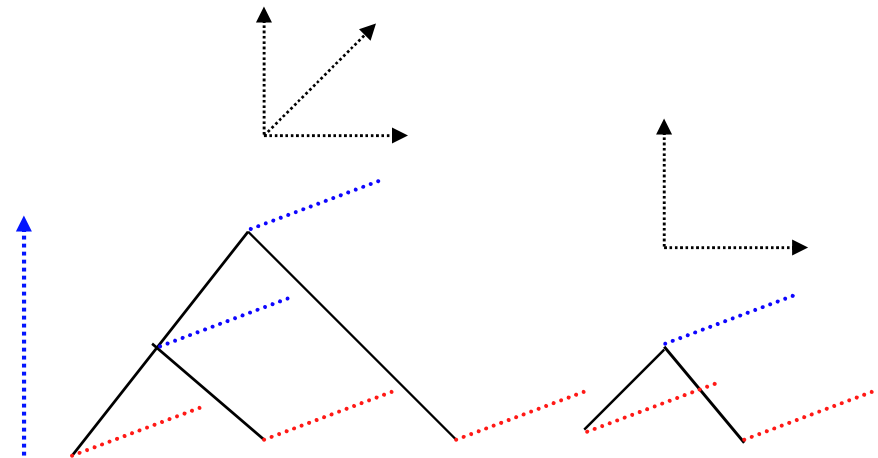
 - indels of length k (Bjarne Knudsen, 2003) **Temporal**

- **Statistical Alignment**

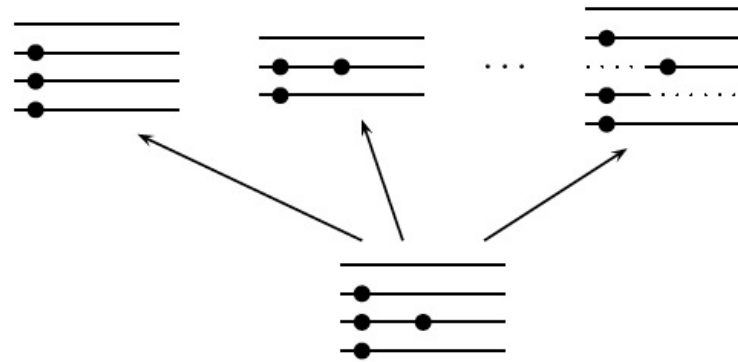
Spatial:



Temporal:



BB & Heuristic minimal ancestral recombination graphs



Beagle

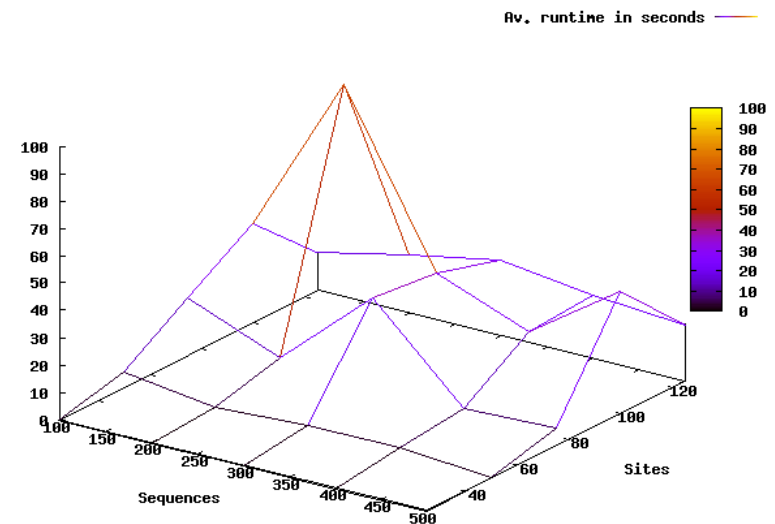
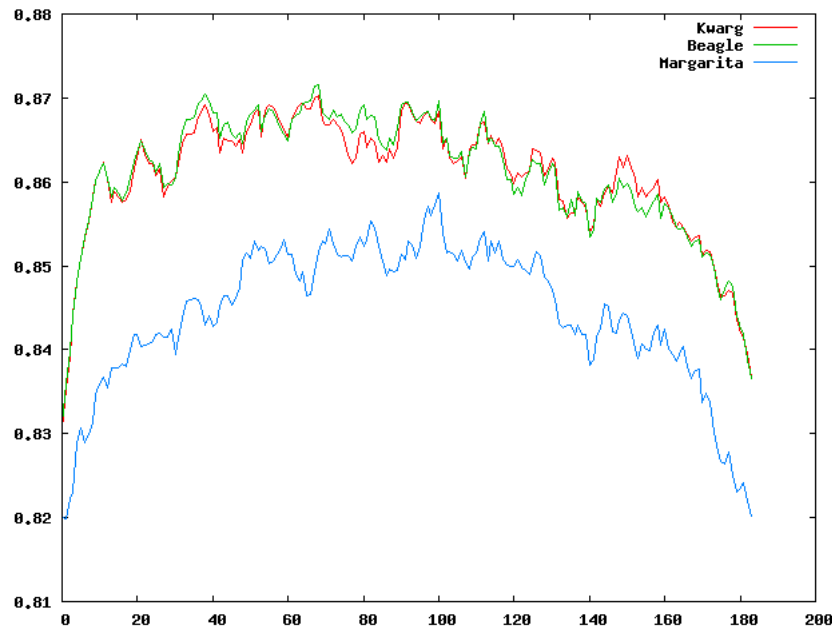
Try each in turn
until shortest route
is determined

Margarita

Just follow road
seeming to lead in
the right direction

Kwarg

Choice based on
location of next
crossroads

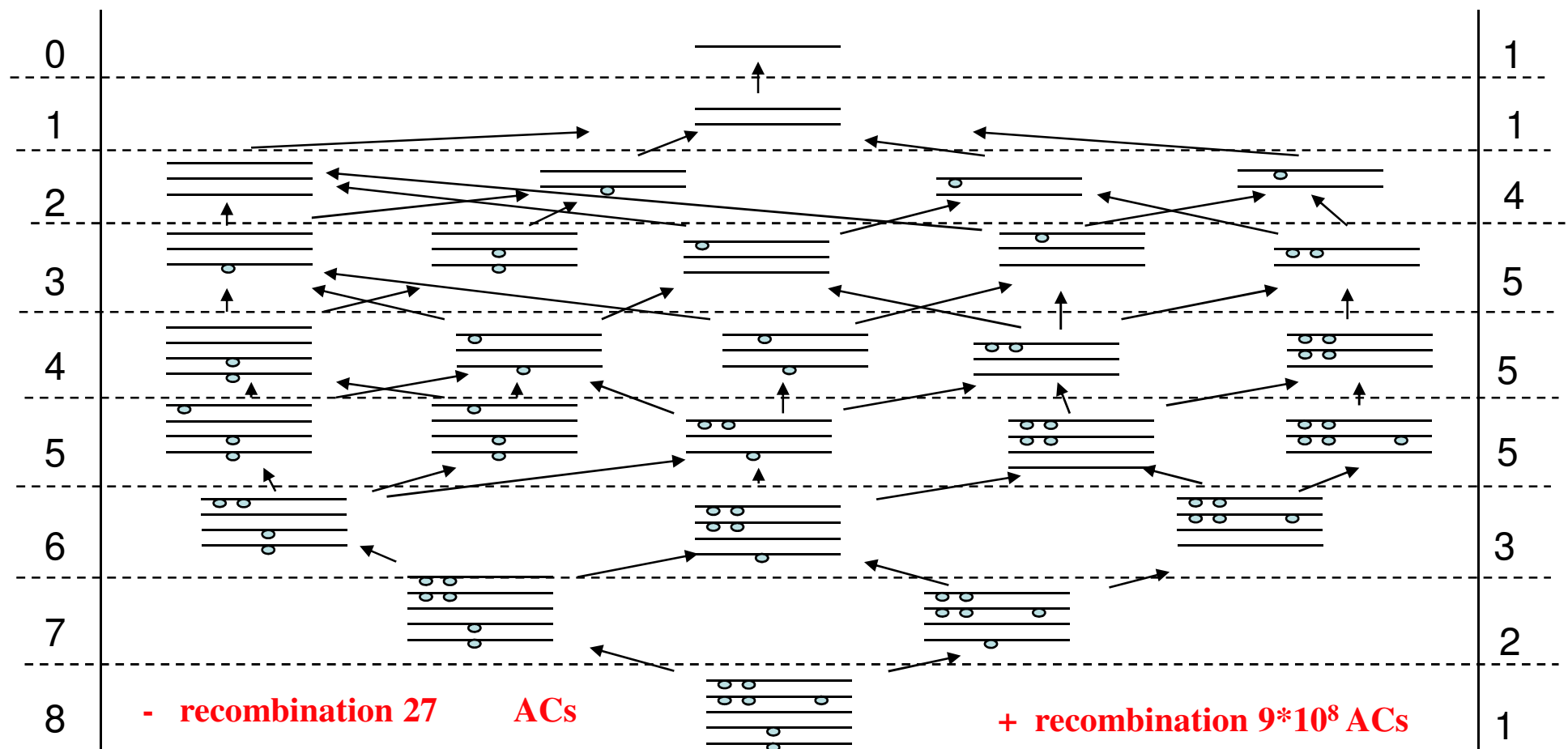


The Griffiths-Ethier-Tavare Recursions

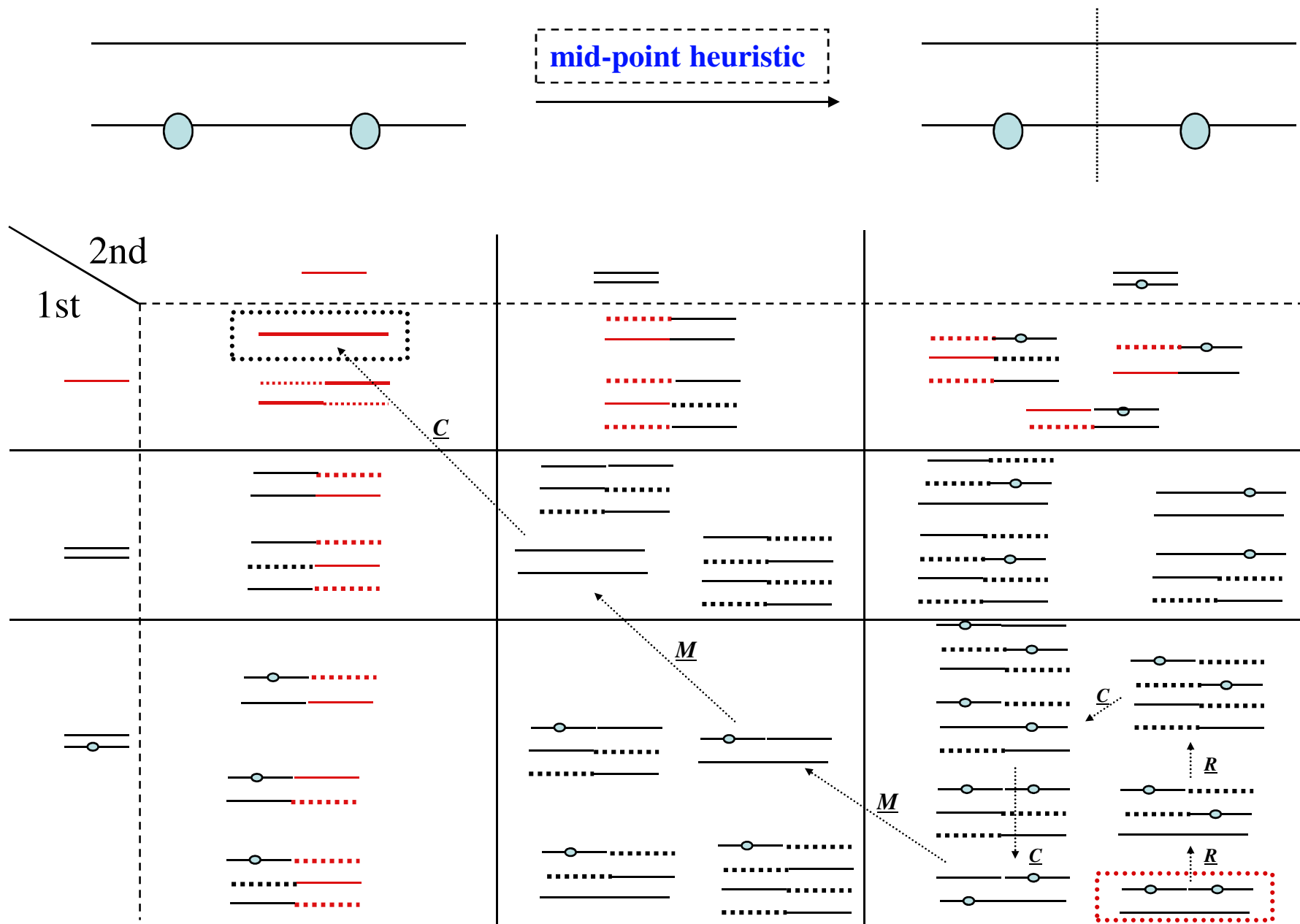
No recombination: Infinite Site Assumption
Ancestral State Known

History Graph: Recursions Exists
No cycles

Possible Histories without Recombination for simple data example



Ancestral configurations to 2 sequences with 2 segregating sites



0-ARG

1-ARG

2-ARG

5 states

+15 states

+10 states

$\Theta, \rho = 2, 0$

0.148

0.148

0.148

$\Theta, \rho = 1, 1$

0.037

0.082

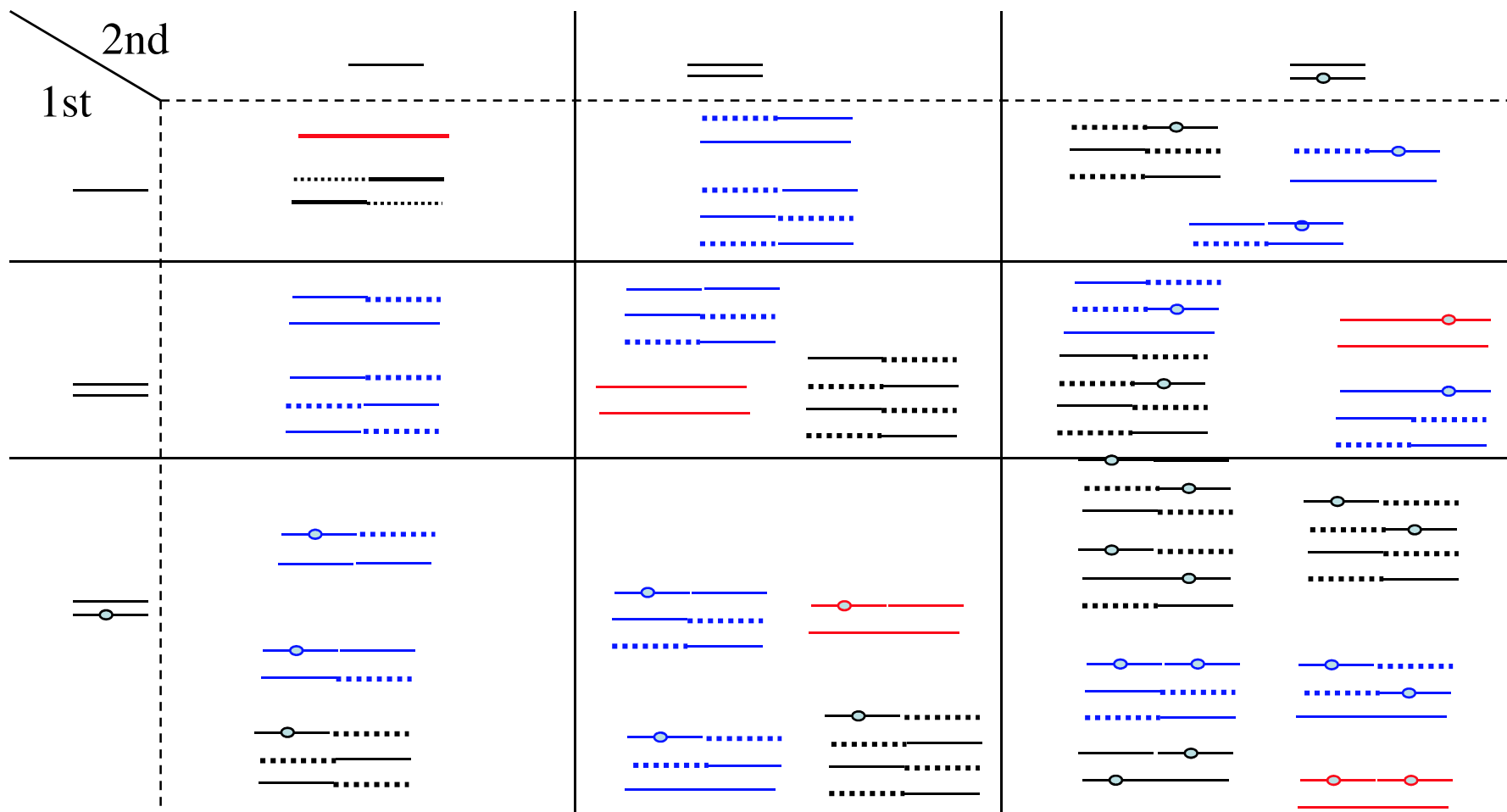
0.090

$\Theta, \rho = 2, 2$

0.032

0.074

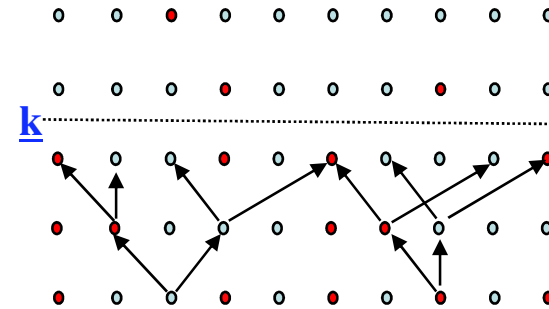
0.085



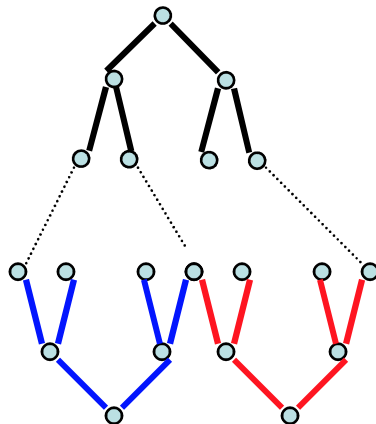
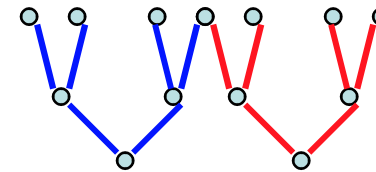
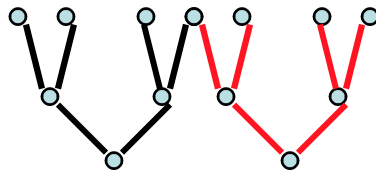
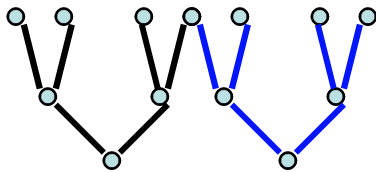
Reconstructing global pedigrees: Superpedigrees

Steel and Hein, 2005

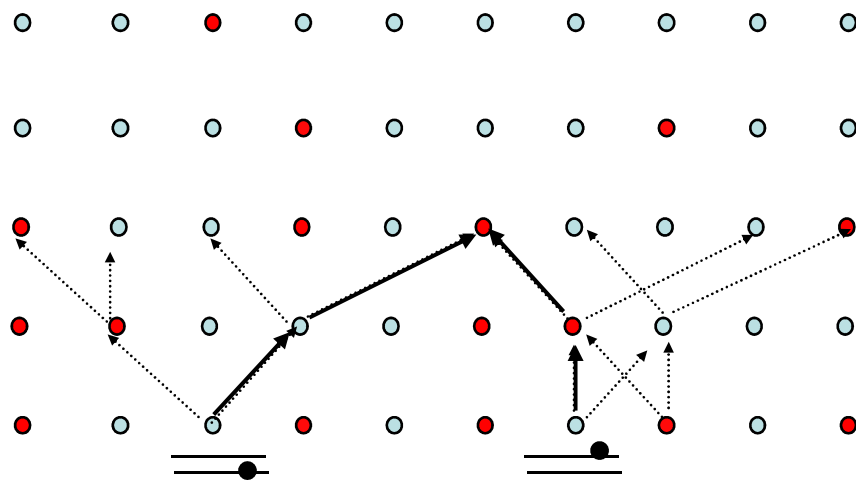
The gender-labeled pedigrees for all pairs defines global pedigree



Gender-unlabeled pedigrees don't!!

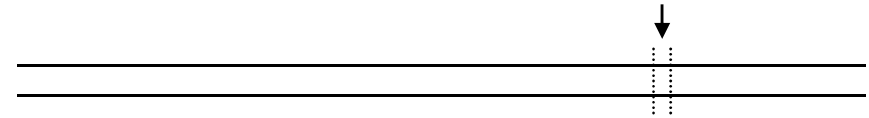


Benevolent Mutation and Recombination Process



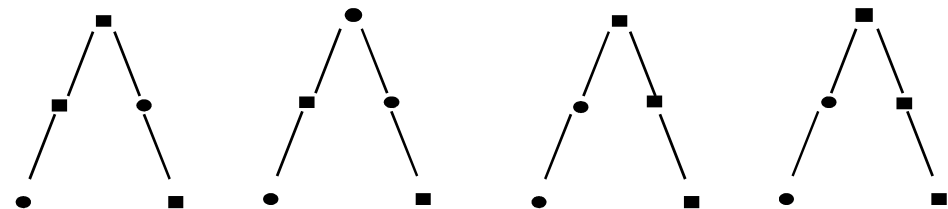
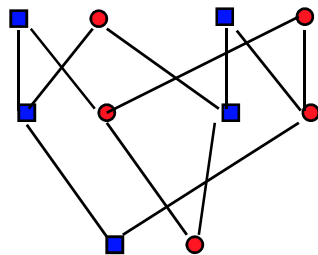
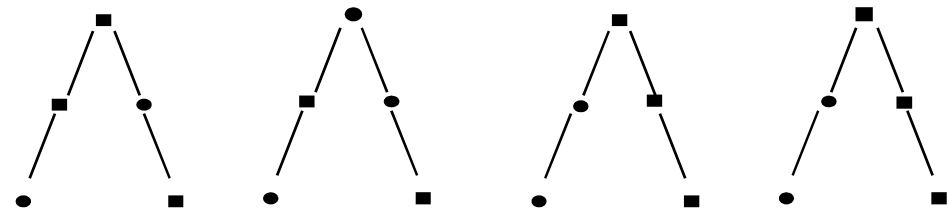
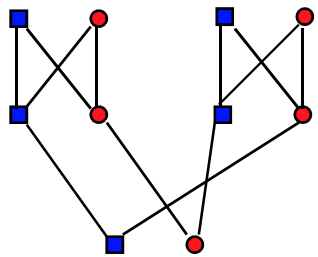
Genomes with ρ and $\mu/\rho \rightarrow \infty$

ρ – recombination rate, μ – mutation rate



- All embedded phylogenies are observable
- Do they determine the pedigree?

Counter example:



Embedded phylogenies:

Infinite Sequences: From ARG to Pedigree

What can you observe?

A. The ARG?

B. Sequence of local trees?

C. Sequence of pairs of local trees with recombination points

D. Set of local trees?

E. Set/Sequences of local unrooted tree topologies?

Given A/B/C/D/E above how much does that constrain set of pedigrees

*How **many** pedigrees are compatible varying over **data**?*

Motivation

- *Inferring/Reconstructing Evolutionary Histories of Sequences*
- *Inferring parameters in the Evolutionary Process Generating Sequences*
- *Understanding why the problem is computationally hard.*

Overview

- *Bounds on Recombinations in Data Sets*
- *Minimal Histories*
 - *Spatial Algorithm*
 - *Temporal Algorithm*
- *Digression: Other Spatial/Temporal Algorithms*
- *Number of Ancestral States*
- *Likelihood calculations on the ε -ARG.*
- *Pedigrees and the ARG.*

References

- Allen, B. and Steel, M., Subtree transfer operations and their induced metrics on evolutionary trees, *Annals of Combinatorics* 5, 1-13 (2001)
- Baroni, M., Grunewald, S., Moulton, V., and Semple, C. Bounding the number of hybridisation events for a consistent evolutionary history. *Journal of Mathematical Biology* 51 (2005), 171-182
- Bordewich, M. and Semple, C. On the computational complexity of the rooted subtree prune and regraft distance. *Annals of Combinatorics* 8 (2004), 409-423
- Griffiths, R.C. (1981). Neutral two-locus multiple allele models with recombination. *Theor. Popul. Biol.* **19**, 169-186.
- J.J.Hein: Reconstructing the history of sequences subject to Gene Conversion and Recombination. *Mathematical Biosciences*. (1990) 98.185-200.
- J.J.Hein: A Heuristic Method to Reconstruct the History of Sequences Subject to Recombination. *J.Mol.Evol.* 20.402-411. 1993
- Hein, J.J., T.Jiang, L.Wang & K.Zhang (1996): "On the complexity of comparing evolutionary trees" *Discrete Applied Mathematics* 71.153-169.
- Hein, J., Schierup, M. & Wiuf, C. (2004) *Gene Genealogies, Variation and Evolution*, Oxford University Press
- Hudson, 1993 Properties of a neutral allele model with intragenic recombination. *Theor Popul Biol.* 1983 23(2):183-2
- Kreitman, M. Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature*. 1983 304(5925):412-7.
- Lyngsø, R.B., Song, Y.S. & Hein, J. (2005) [Minimum Recombination Histories by Branch and Bound](#). *Lecture Notes in Bioinformatics: Proceedings of WABI 2005* 3692: 239–250.
- Lyngsø, R.B., Song, Y.S. & Hein, J. (2008) Accurate computation of likelihoods in the coalescent with recombination via parsimony. *Lecture Notes in Bioinformatics: Proceedings of RECOMB 2008* 463–477
- Song, Y.S. (2003) On the combinatorics of rooted binary phylogenetic trees. *Annals of Combinatorics*, 7:365–379
- Song, Y.S., Lyngsø, R.B. & Hein, J. (2005) Counting Ancestral States in Population Genetics. Submitted.
- Song, Y.S. & Hein, J. (2005) [Constructing Minimal Ancestral Recombination Graphs](#). *J. Comp. Biol.*, 12:147–169
- Song, Y.S. & Hein, J. (2004) [On the minimum number of recombination events in the evolutionary history of DNA sequences](#). *J. Math. Biol.*, 48:160–186.
- Song, Y.S. & Hein, J. (2003) Parsimonious reconstruction of sequence evolution and haplotype blocks: finding the minimum number of recombination events, *Lecture Notes in Bioinformatics, Proceedings of WABI'03*, 2812:287–302.
- [Song YS, Wu Y, Gusfield D](#). Efficient computation of close lower and upper bounds on the minimum number of recombinations in biological sequence evolution. *Bioinformatics*. 2005 Jun 1;21 Suppl 1:i413-i422.
- Wiuf, C. Inference on recombination and block structure using unphased data. *Genetics*. 2004 Jan;166(1):537-45.