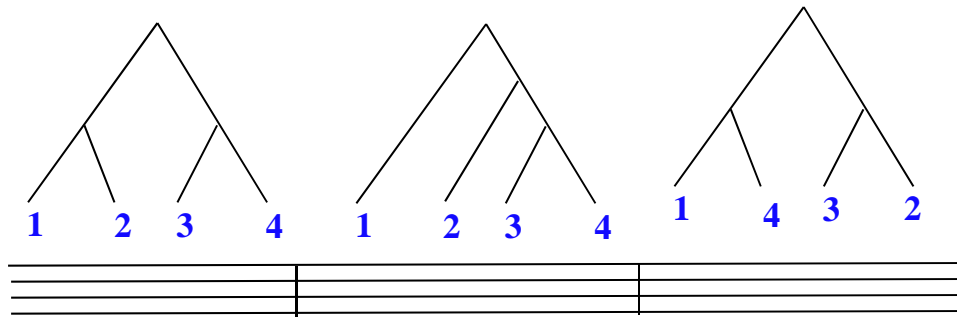
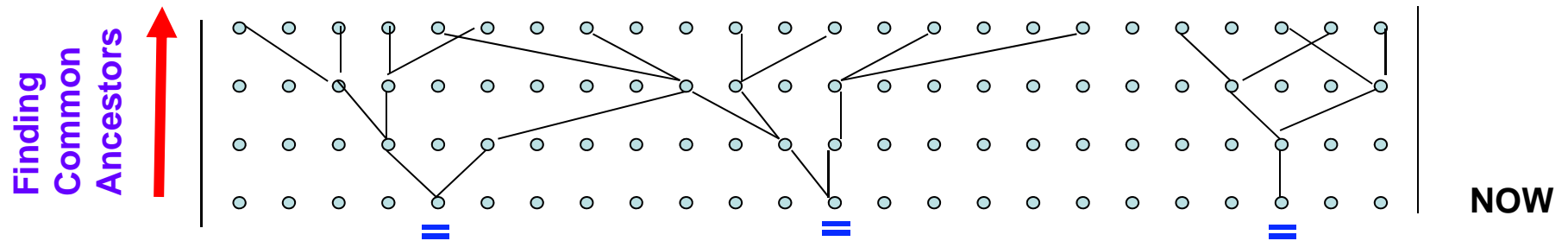


Recombination Histories & Global Pedigrees

Finding Minimal Recombination Histories



Global Pedigrees

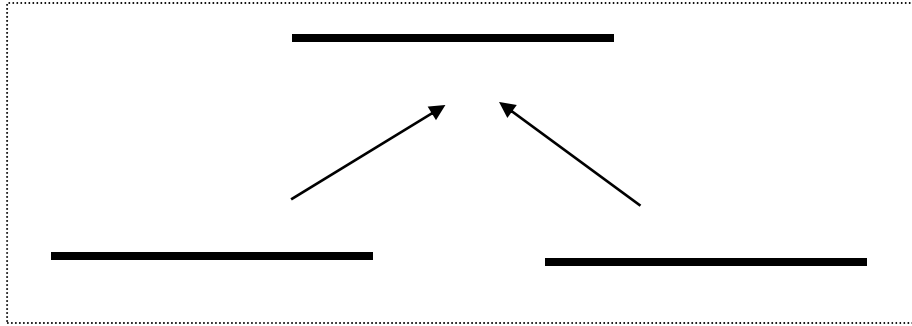


Acknowledgements

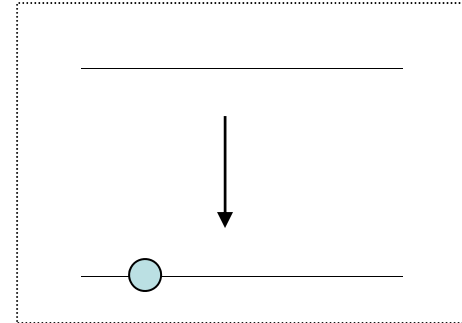
Yun Song - Rune Lyngsø - Mike Steel

Basic Evolutionary Events

Coalescent/Duplication

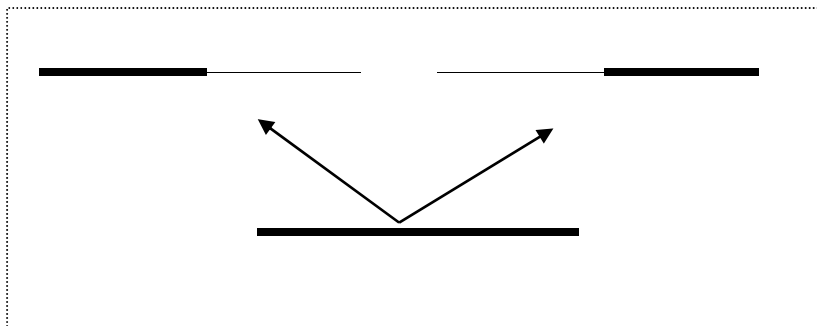


Mutation

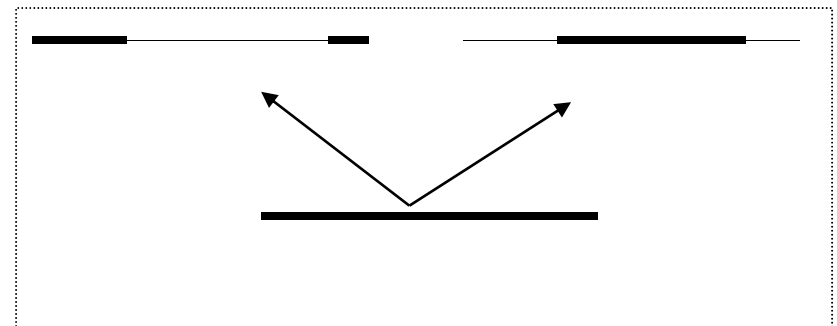


Infinite site assumption ?

Recombination

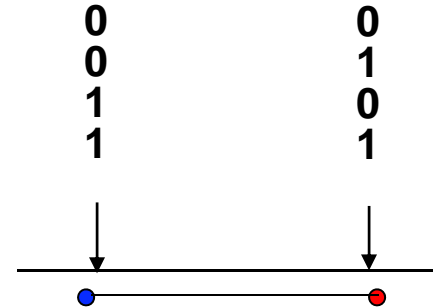
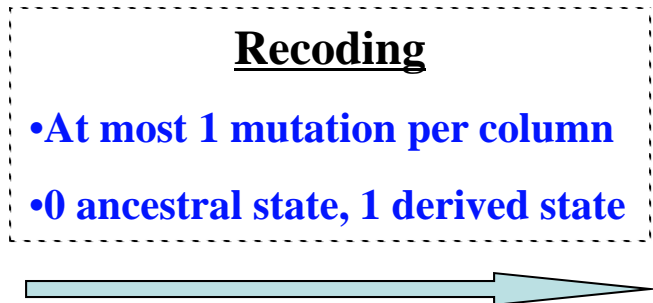


Gene Conversion

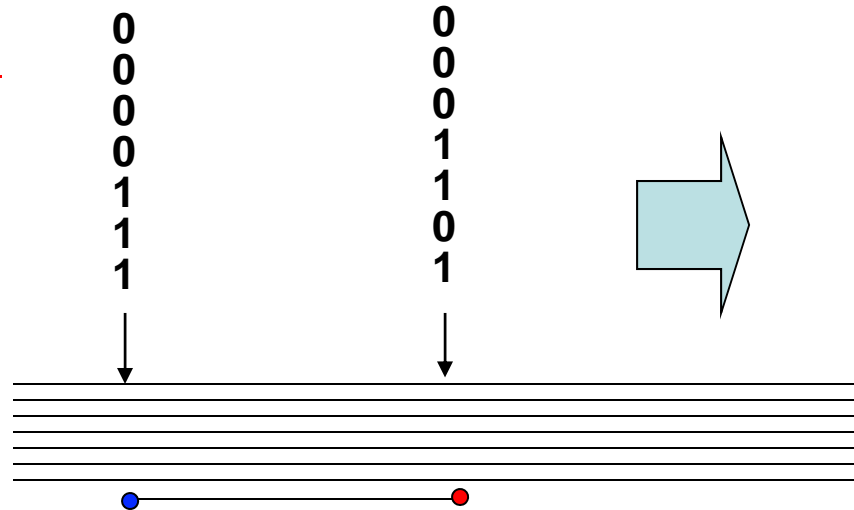


Local Inference of Recombinations

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



Incompatibility:



Four combinations

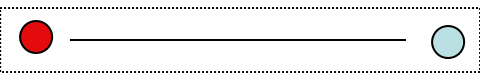
00
10
01
11

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

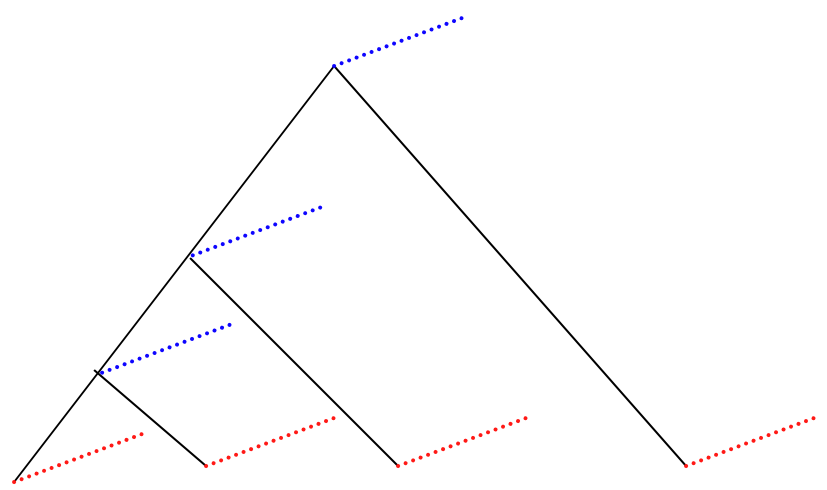
$$N_R \geq N_T - N_M - 1$$

Hudson & Kaplan's R_M

```
0 0 0 0 0 1 0 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0
0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 1 1 1
0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 1
1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 1
1 1 1 1 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 1
```



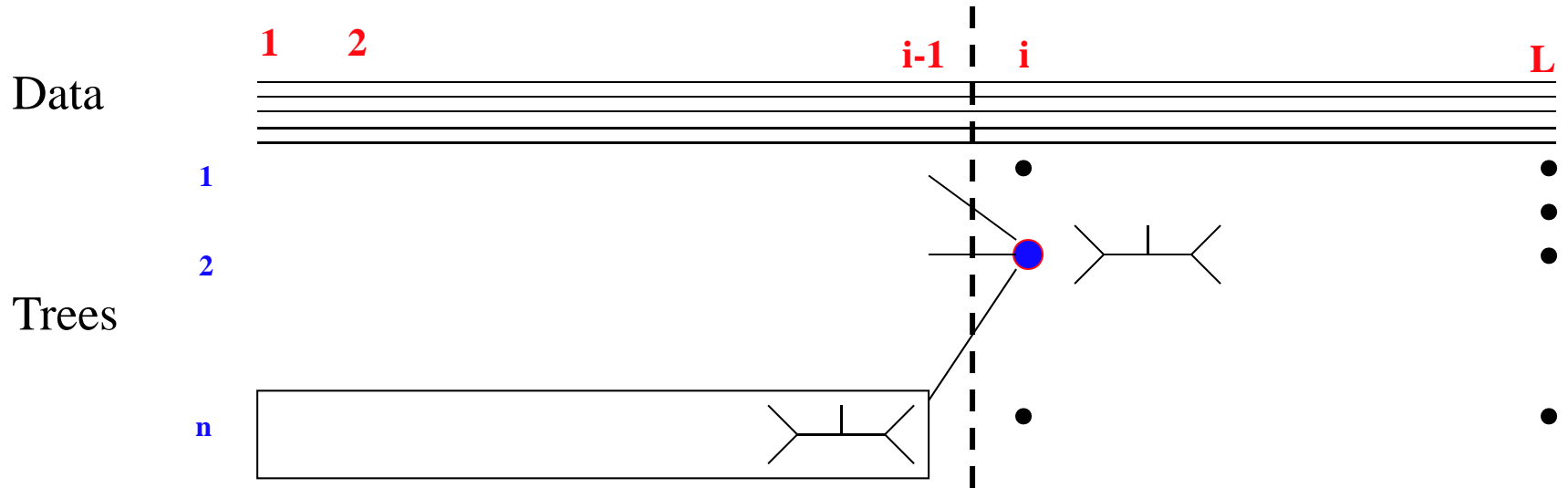
If you equate R_M with expected number of recombinations, this could be used as an estimator. Unfortunately, R_M is a gross underestimate of the real number of recombinations.



Minimal Number of Recombinations

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

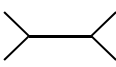
Last Local Tree Algorithm:



How many local trees?

How many neighbors?

Bi-partitions

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

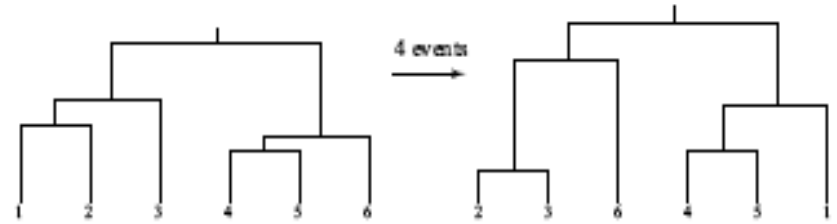
$$3n^2 - 13n + 14$$

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

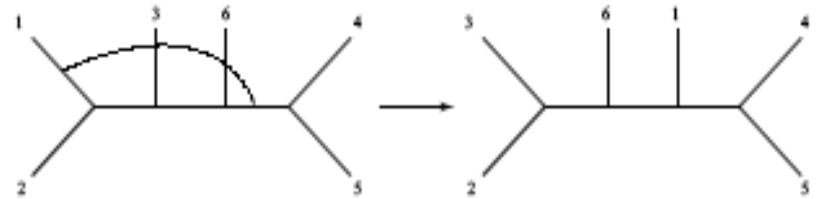
$$\sim n^3$$

Metrics on Trees based on subtree transfers.

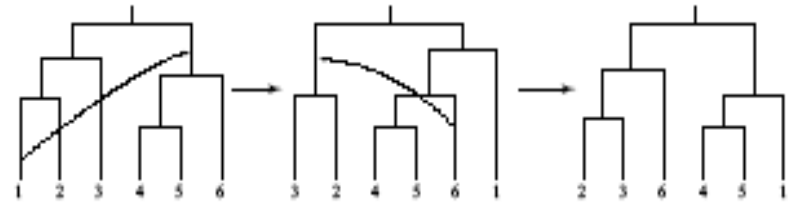
Trees including branch lengths



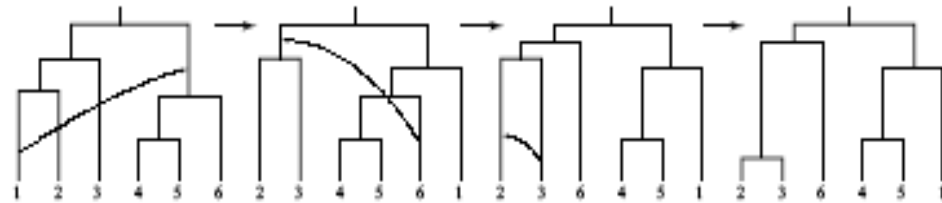
Unrooted tree topologies



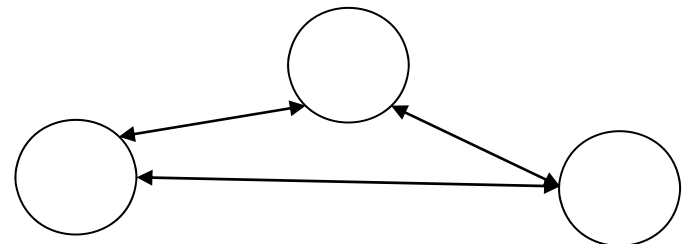
Rooted tree topologies



Tree topologies with age ordered internal nodes



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



Tree Combinatorics and Neighborhoods

Observe that the size of the unit-neighbourhood of a tree does not grow nearly as fast as the number of trees

$\delta(T)$:= number of trees one SPR operation away from a given tree T .

	Unrooted		Rooted			Dendrograms		
n	# of trees	δ	# of trees	δ_{\max}	δ_{\min}	# of trees	δ_{\max}	δ_{\min}
4	3	2	15	12	10	18	12	13
5	15	12	105	28	24	180	33	37
6	105	30	945	52	44	2,700	71	79
7	945	56	10,395	84	70	56,700	128	143
8	10,395	90	135,135	124	102	1,587,600	210	233
9	135,135	132	2,027,025	170	140	57,153,600	?	?
10	2,027,025	182	34,459,425	224	184	2,571,912,000	?	?

Due to Yun Song

$(2n-3)!! = \frac{(2n-2)!}{2^{n-1}(n-1)!}$
 $2(n-3)(2n-7)$
 $3n^2 - 13n + 14$
 $4(n-2)^2 - 2 \sum_{m=1}^{n-2} \lfloor \log_2(m+1) \rfloor$
 $\frac{n!(n-1)!}{2^{n-1}}$
 $\frac{1}{3} (2n^3 - 3n^2 - 20n + 39)$

Allen & Steel (2001)

Song (2003+)

The Minimal Recombination History for the Kreitman Data

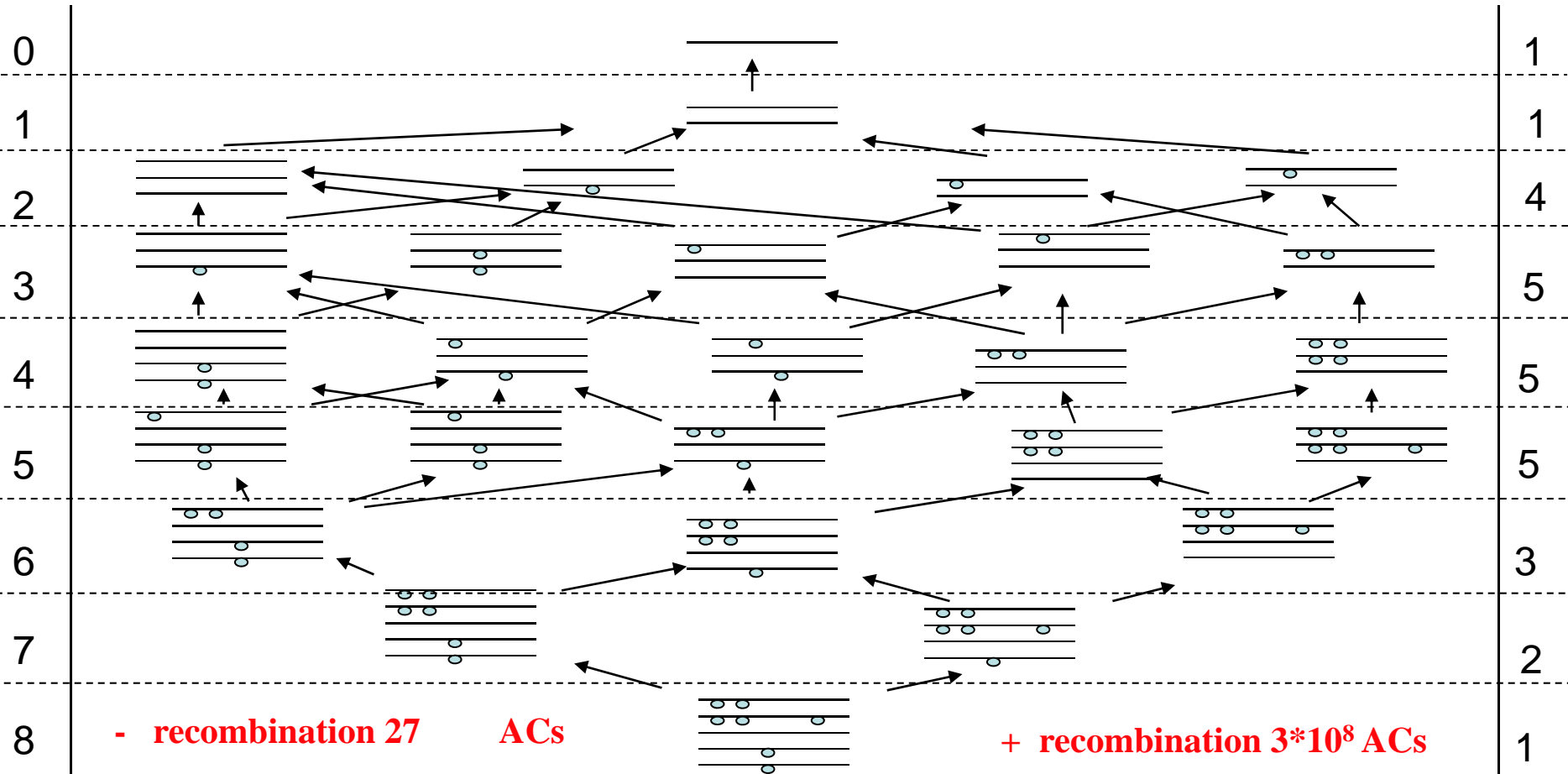
Methods	# of rec events obtained
Hudson & Kaplan (1985)	5
Myers & Griffiths (2003)	6
Song & Hein (2004). <i>Set theory based approach.</i>	7
Song & Hein (2003). Tree scanning using DP	7
Lyngsø, Song & Hein (2006). <i>Massive Acceleration using Branch and Bound Algorithm.</i>	7
Lyngsø, Song & Hein (2006). <i>Minimal number of Gene Conversions (in prep.)</i>	5-2/6-1

The Griffiths-Ethier-Tavare Recursions

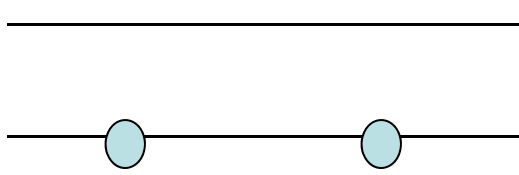
No recombination: Infinite Site Assumption
Ancestral State Known

History Graph: Recursions Exists
No cycles

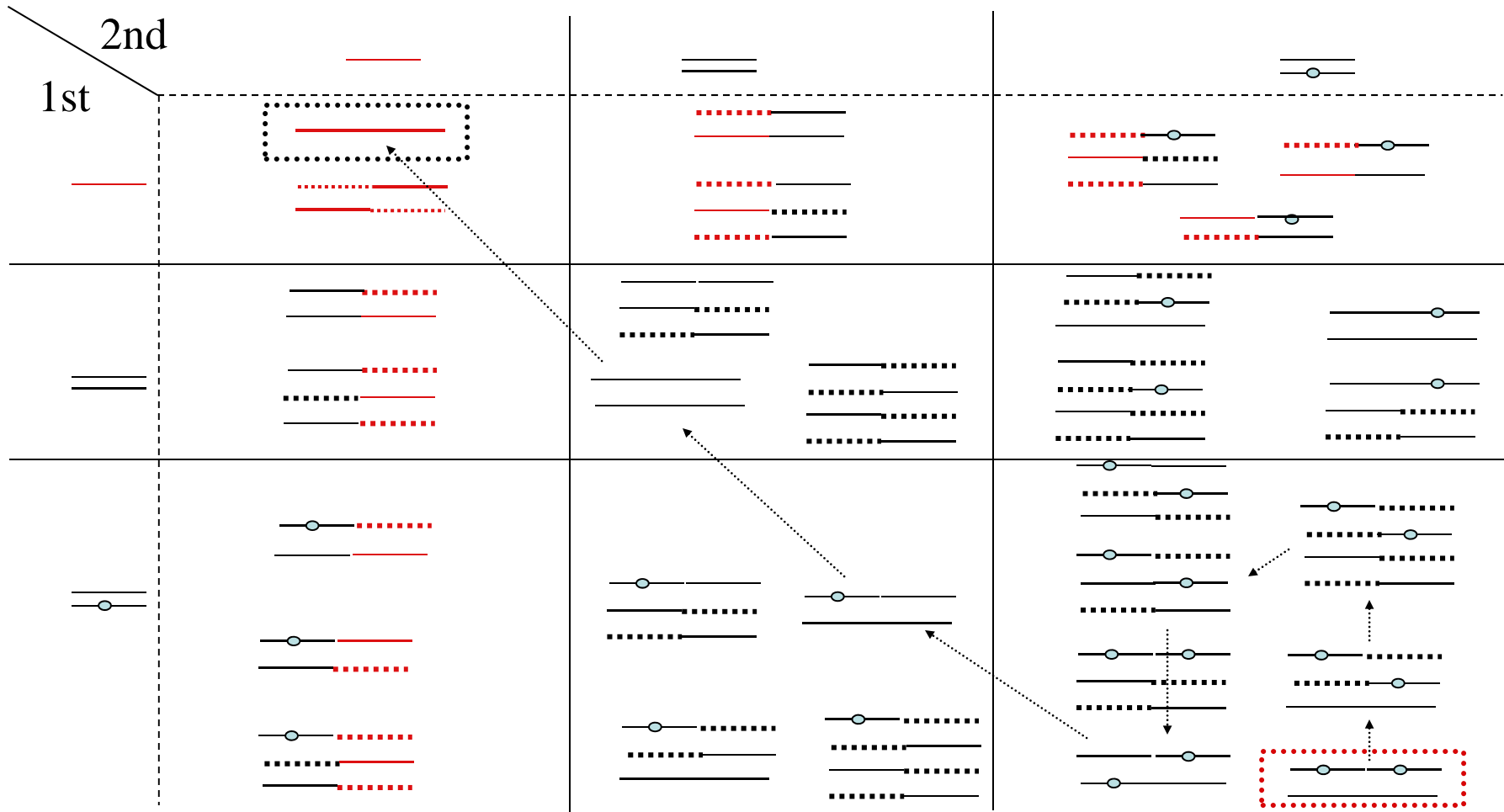
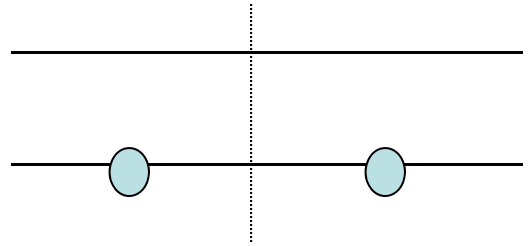
Possible Histories without Recombination for simple data example



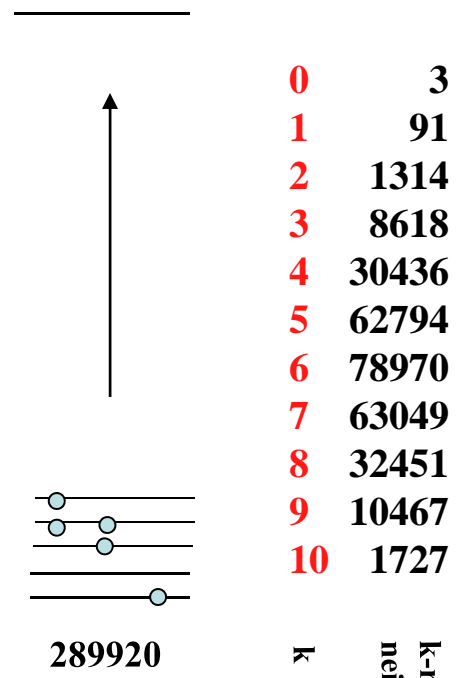
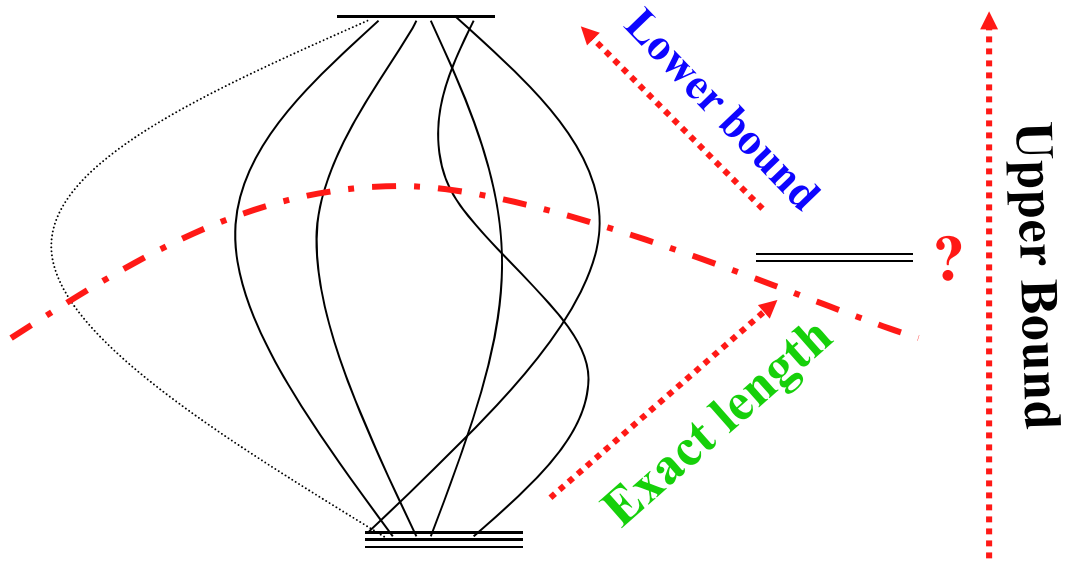
Ancestral configurations to 2 sequences with 2 segregating sites



mid-point heuristic



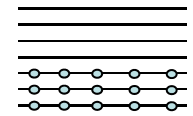
Counting + Branch and Bound Algorithm



0	3
1	91
2	1314
3	8618
4	30436
5	62794
6	78970
7	63049
8	32451
9	10467
10	1727

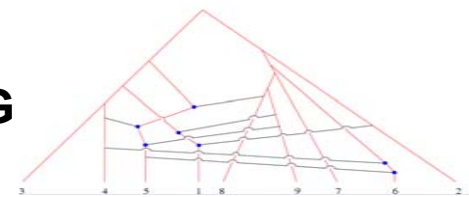
n	Number of segregating sites			
	2	3	4	5
2	30	573	16 875	689 175
3	108	6 286	743 387	149 861 079
4	330	62 589	32 482 009	35 523 729 489
5	866	445 137	893 479 326	4 938 627 635 669
6	2 143	3 302 506	29 521 615 942	962 962 451 049 968
7	4 611	17 409 443	568 860 072 916	91 812 561 254 804 105
8	9 728	98 432 218	13 273 296 248 617	
9	18 378	420 106 717	195 515 335 378 914	
10	34 552	1 917 604 869		
11	59 577	6 985 275 356		

k
k-recombination neighborhood

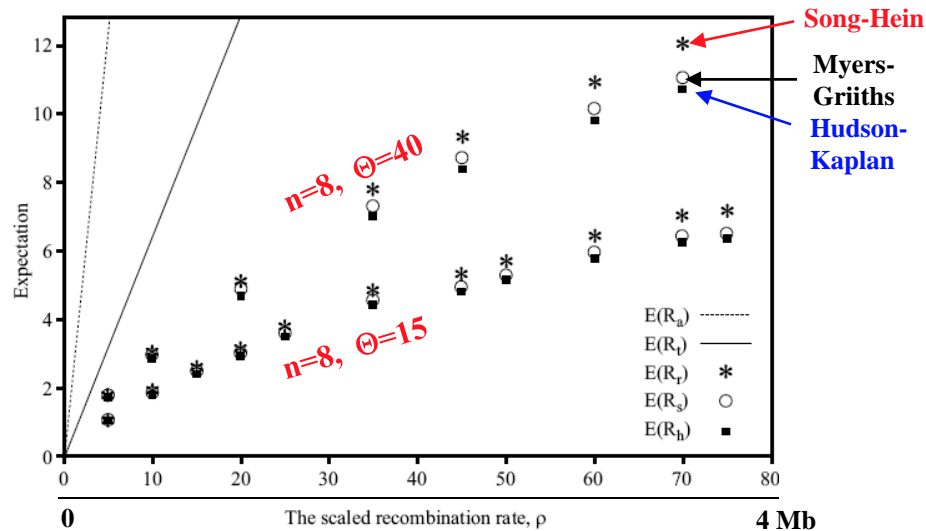
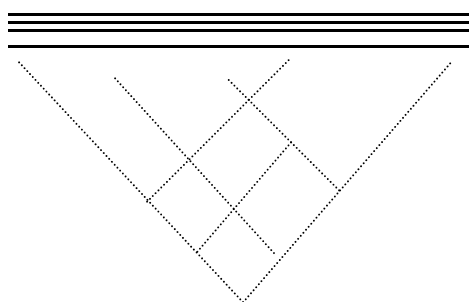


minARGs: Recombination Events & Local Trees

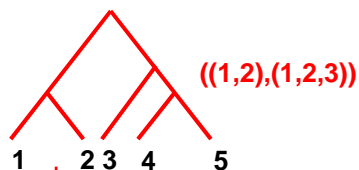
Minimal ARG



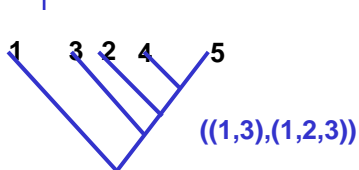
True ARG



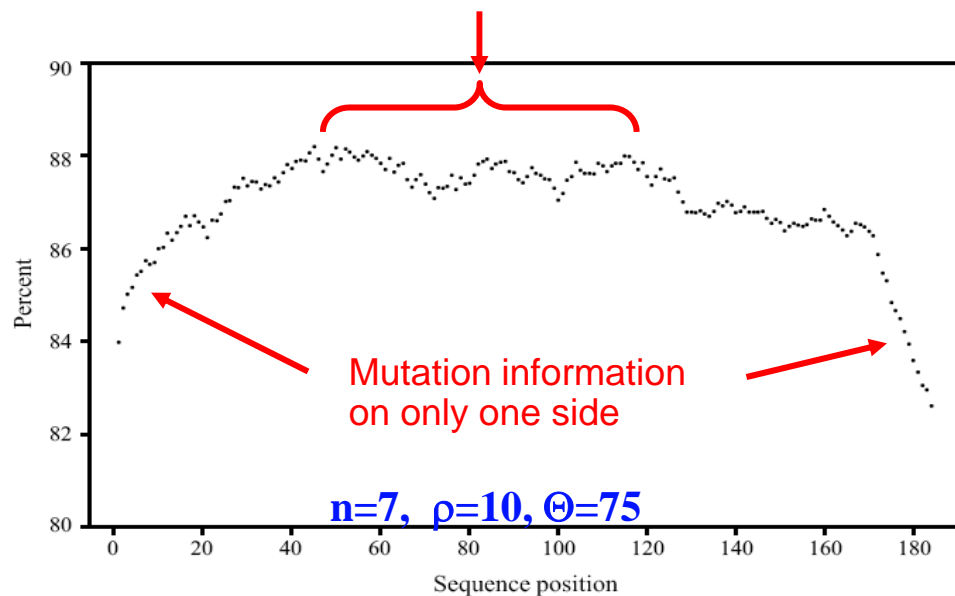
True ARG



Reconstructed ARG



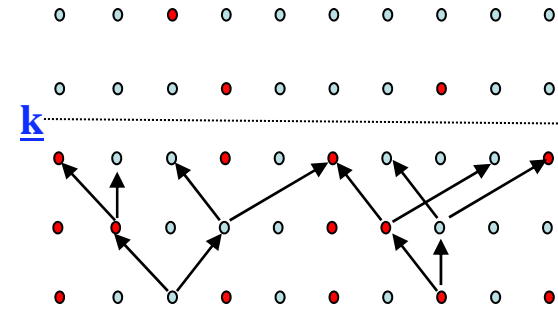
Mutation information on both sides



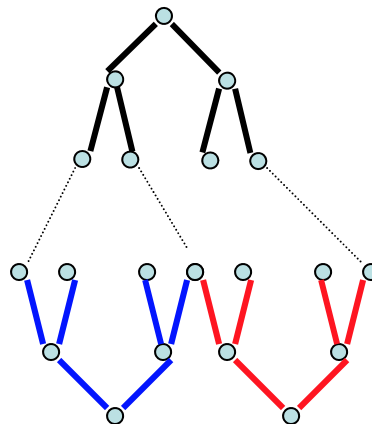
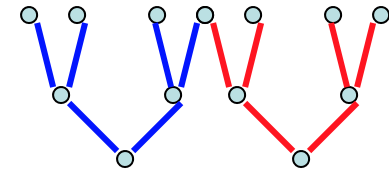
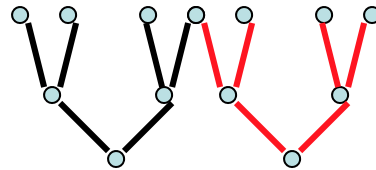
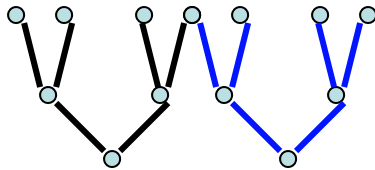
Reconstructing global pedigrees: Superpedigrees

Steel and Hein, 2005

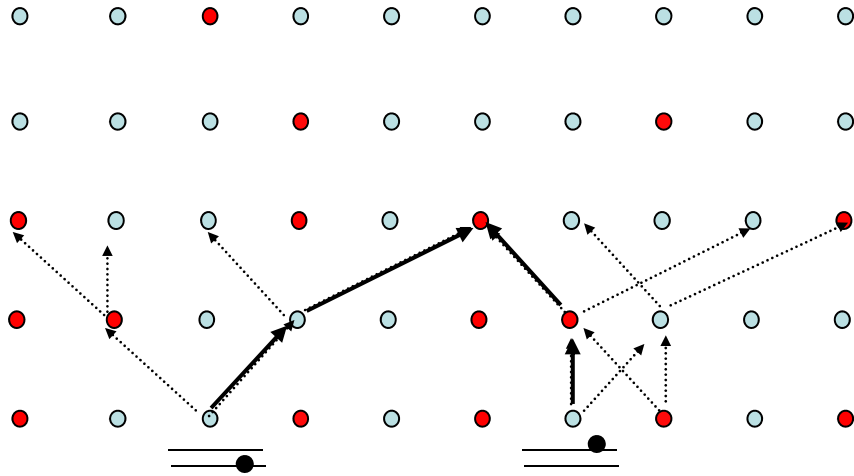
The gender-labeled pedigrees for all pairs, defines global pedigree



Gender-unlabeled pedigrees doesn't!!

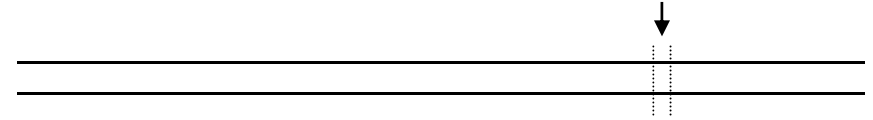


Benevolent Mutation and Recombination Process



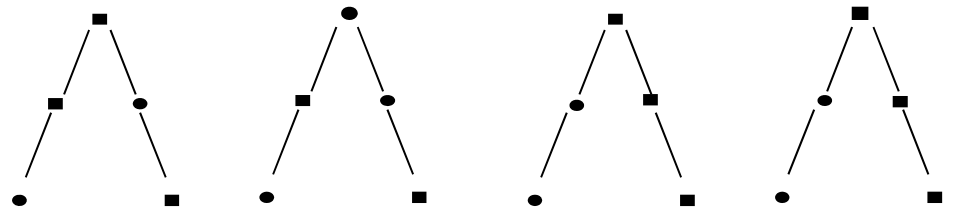
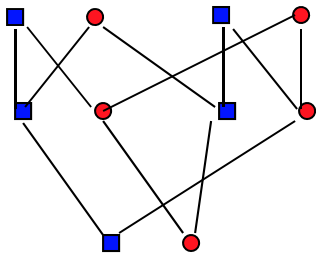
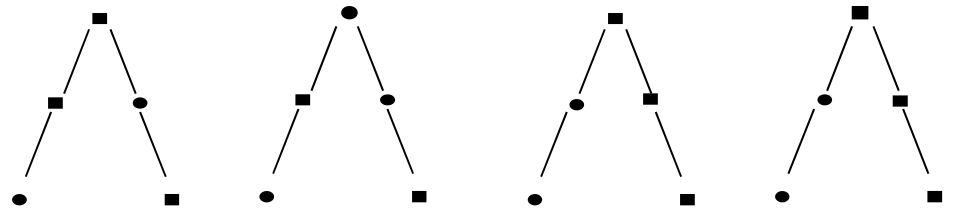
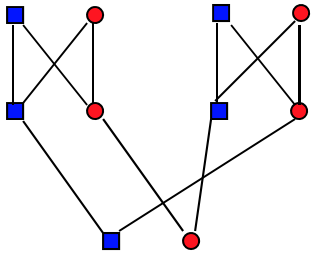
Genomes with ρ and $\mu/\rho \rightarrow \infty$

ρ – recombination rate, μ – mutation rate



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

Counter example:



Embedded phylogenies: