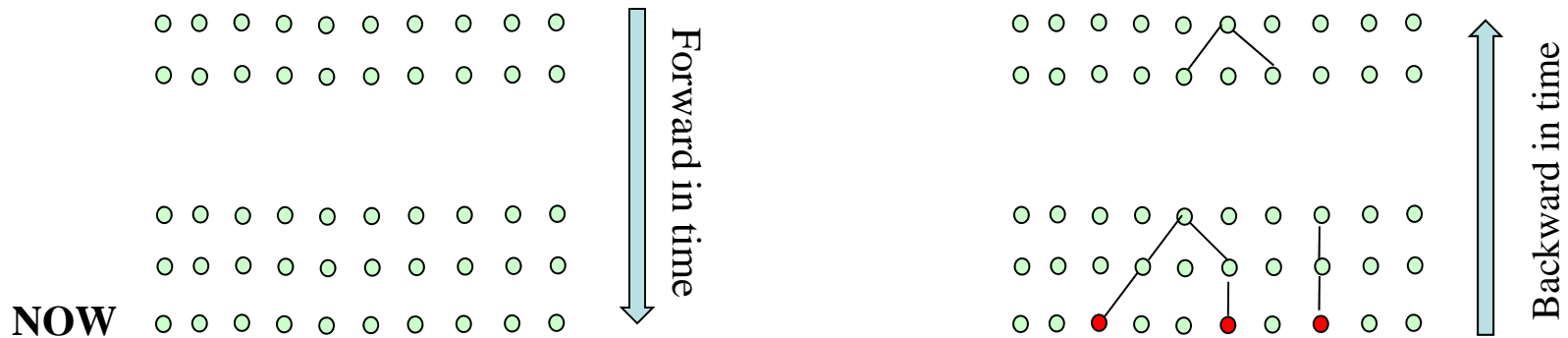
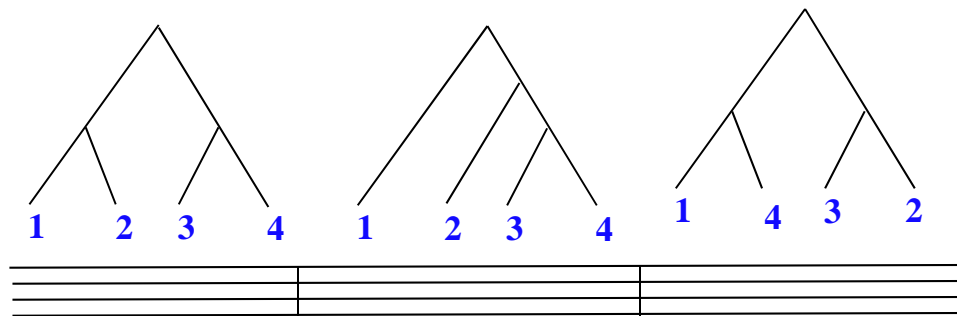


# Population Genetics and Recombination Histories

## Population Genetics and Genealogies

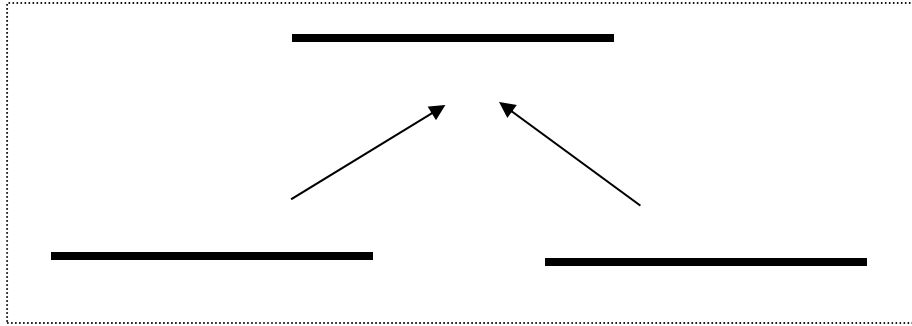


## Finding Minimal Recombination Histories

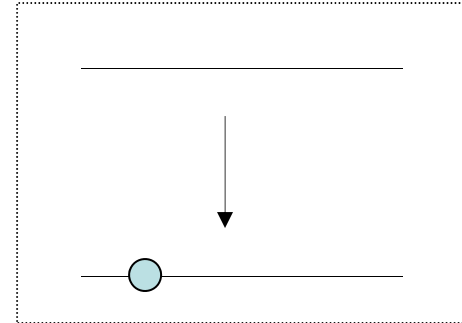


# Basic Evolutionary Events

## Coalescent/Duplication

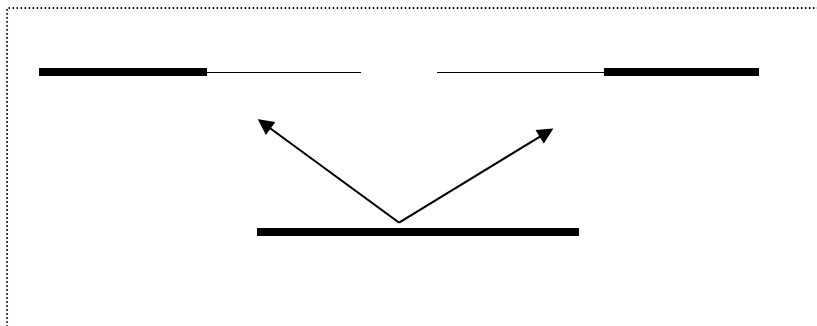


## Mutation

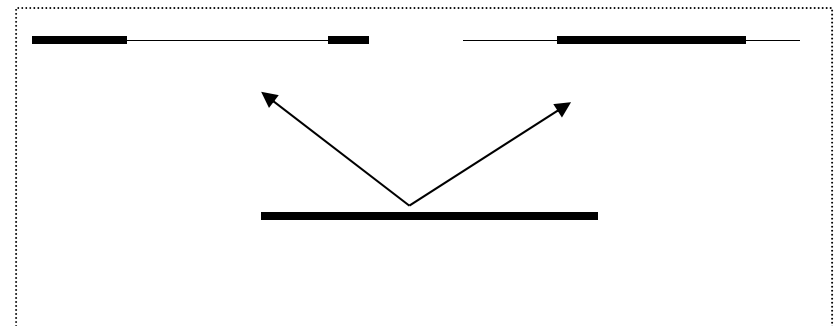


*Infinite site assumption ?*

## Recombination



## Gene Conversion



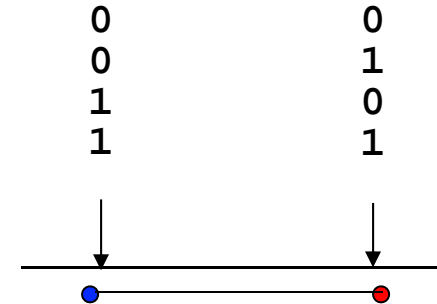


# Local Inference of Recombinations

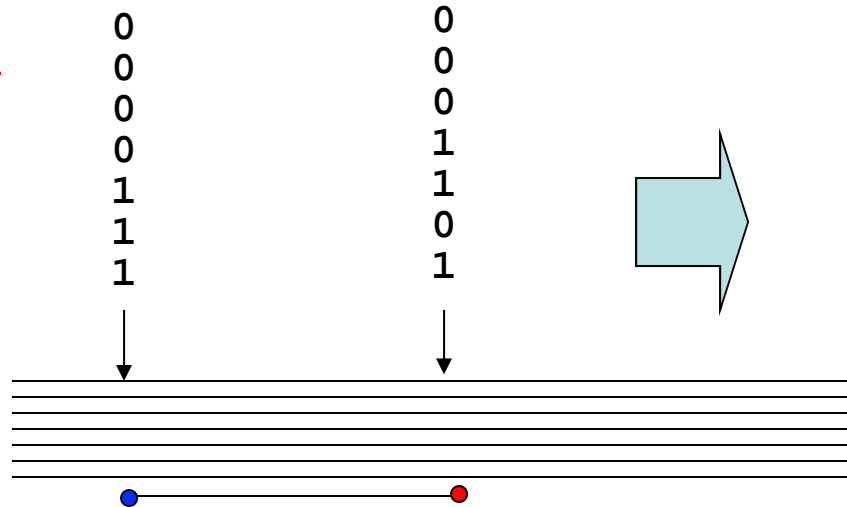
T . . . G  
 T . . . C  
 A . . . G  
 A . . . C

Recoding

- At most 1 mutation per column
- 0 ancestral state, 1 derived state



Incompatibility:



**Four combinations**

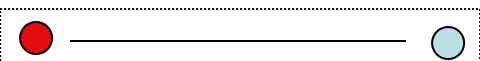
00  
 10  
 01  
 11

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

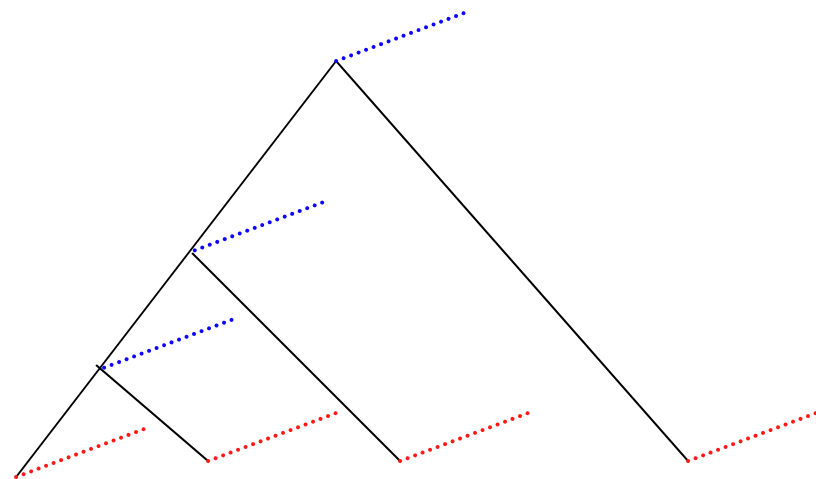
$$N_R \geq N_T - N_M - 1$$

# "Observing" Recombinations: Hudson & Kaplan's $R_M$

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



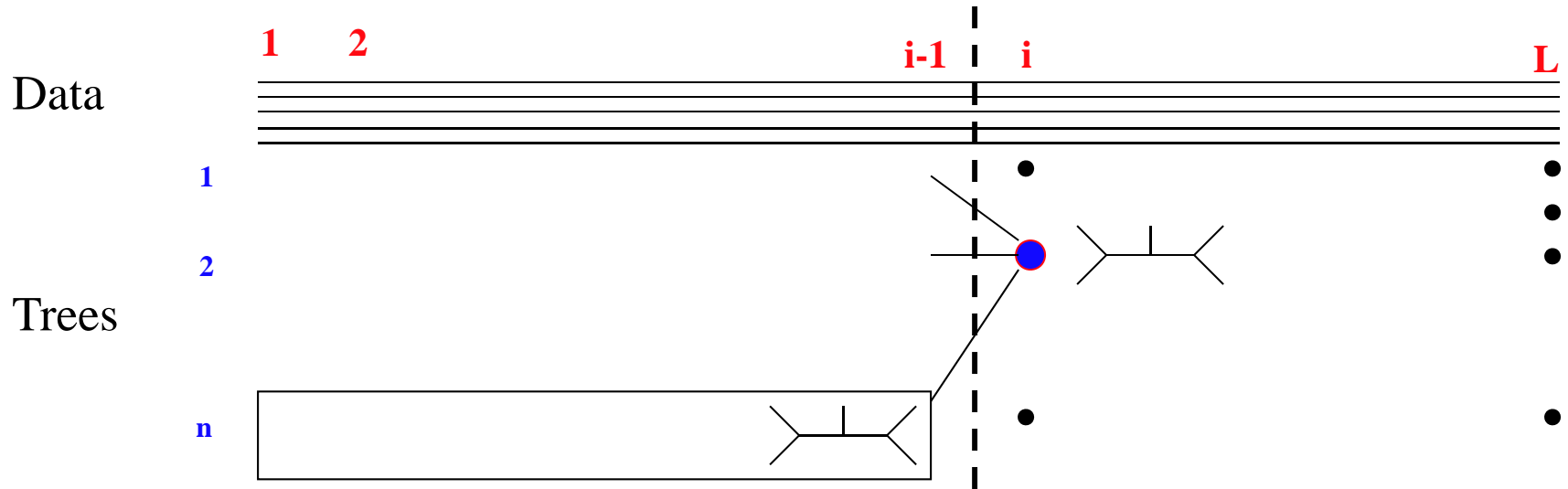
If you equate  $R_M$  with expected number of recombinations, this could be used as an estimator. Unfortunately,  $R_M$  is a gross underestimate of the real number of recombinations.



# Minimal Number of Recombinations

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

## Last Local Tree Algorithm:

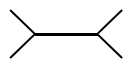


How many local trees?

How many neighbors?

Bi-partitions

• Unrooted



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

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

• Coalescent



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

$$\sim n^3$$

# Tree Combinatorics and Neighborhoods

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

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

	Unrooted		Rooted			Dendrograms		
$n$	# of trees	$\delta$	# of trees	$\delta_{\max}$	$\delta_{\min}$	# of trees	$\delta_{\max}$	$\delta_{\min}$
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	?	?

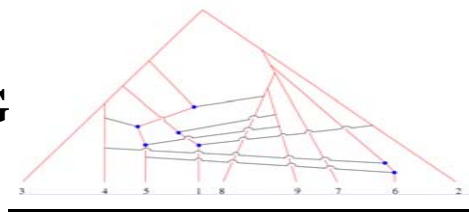
Due to Yun Song

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

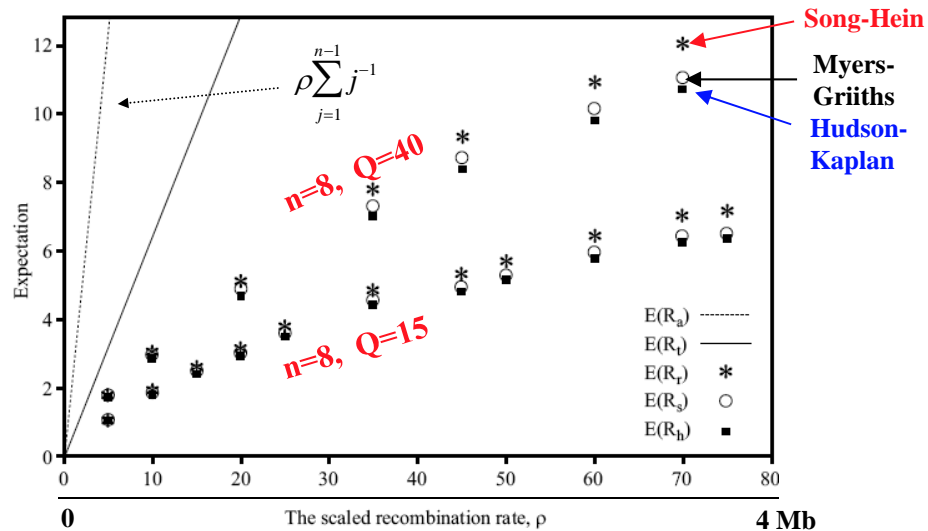
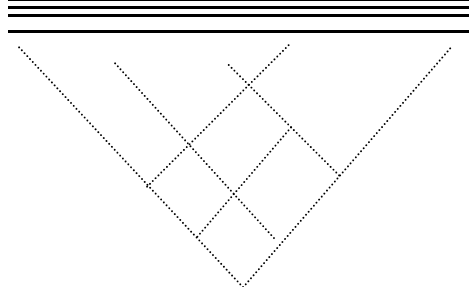


# minARGs: Recombination Events & Local Trees

Minimal ARG



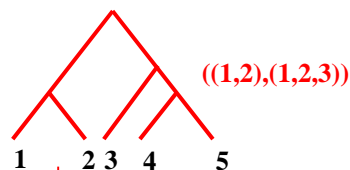
True ARG



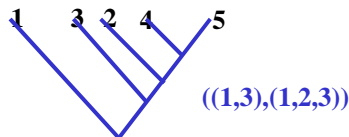
$$E_{70}(R) = 132$$

$$E_{70}(R_{\text{visible}}) = 46$$

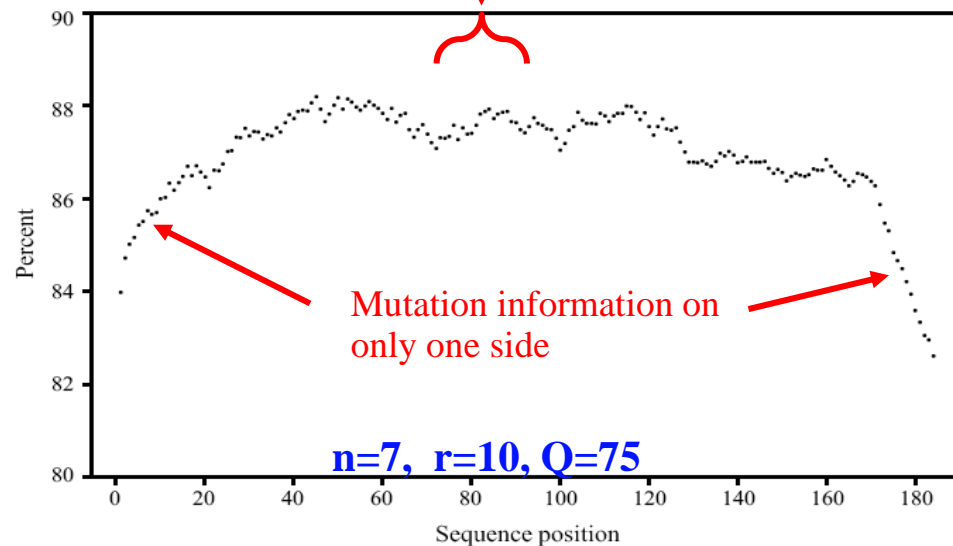
True ARG



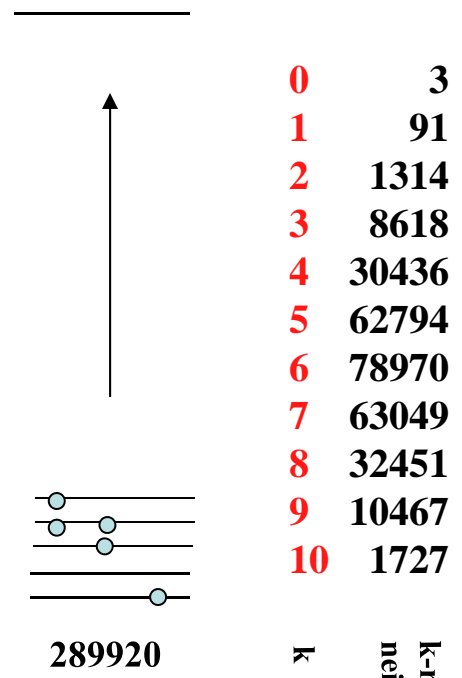
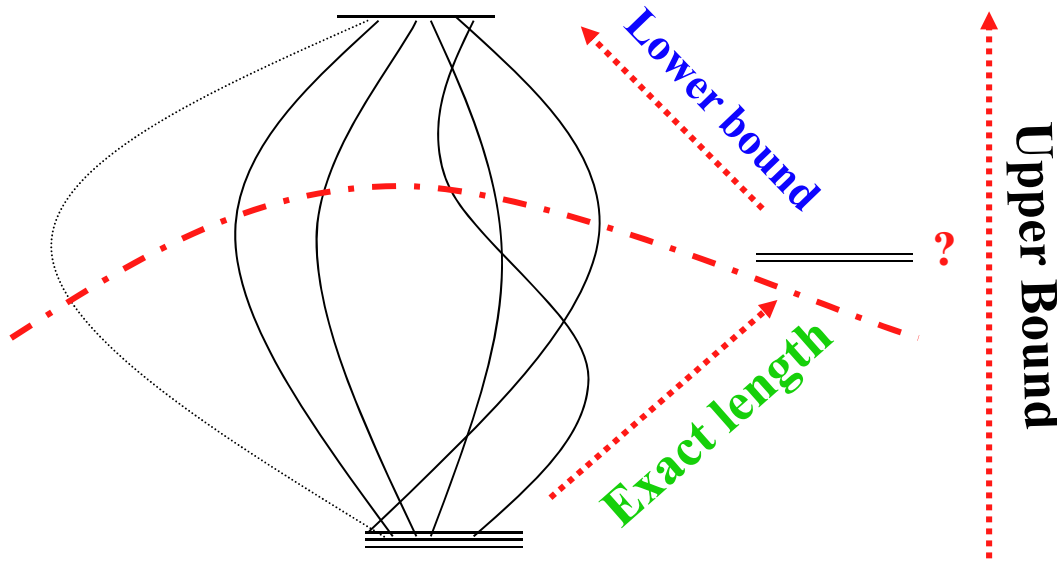
Reconstructed ARG



Mutation information on both sides



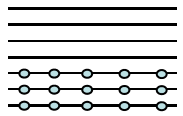
# Counting + Branch and Bound Algorithm



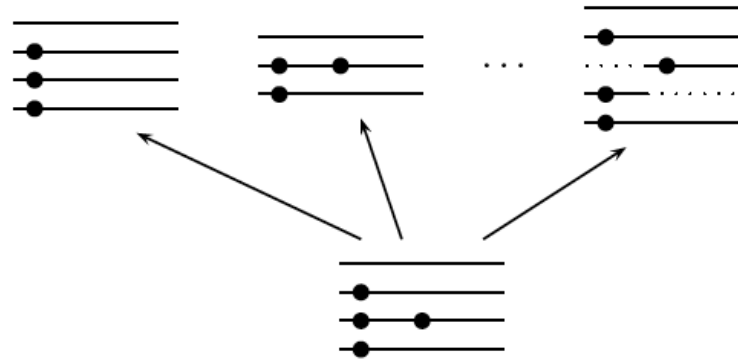
0	3
1	91
2	1314
3	8618
4	30436
5	62794
6	78970
7	63049
8	32451
9	10467
10	1727

k-recombination neighborhood

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		



# BB & Heuristic minimal ancestral recombination graphs



## Beagle

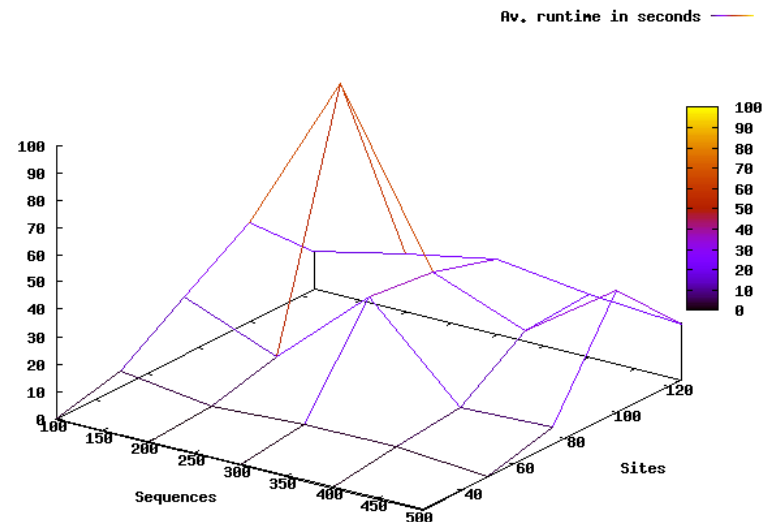
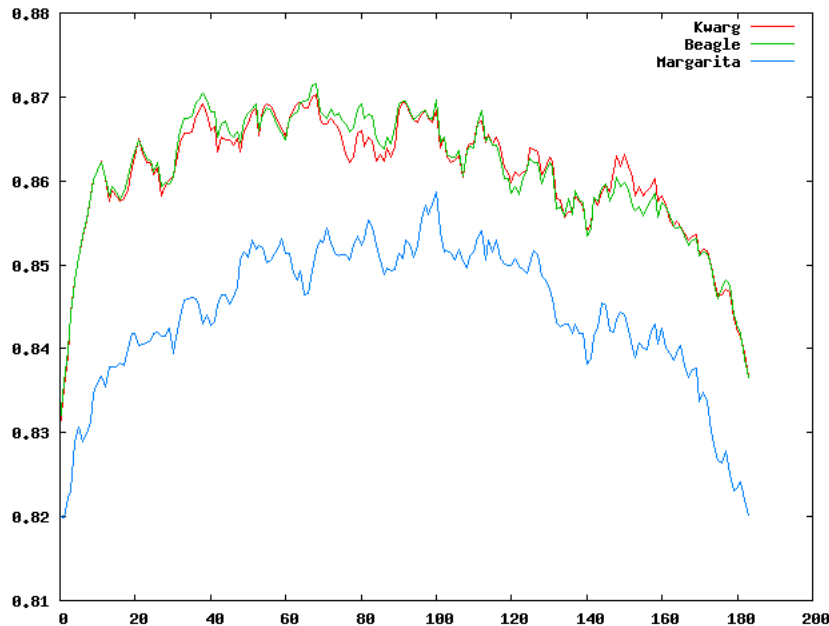
Try each in turn until shortest route is determined

## Margarita

Just follow road seeming to lead in the right direction

## Kwarg

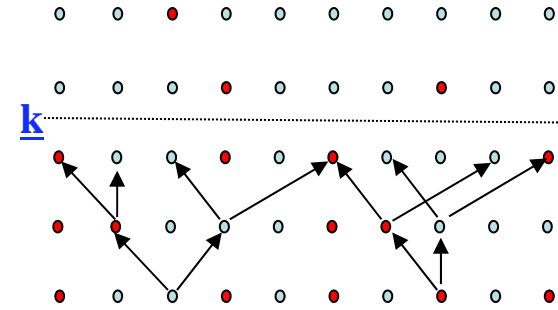
Choice based on location of next crossroads



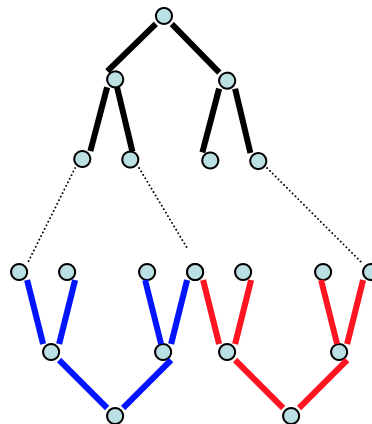
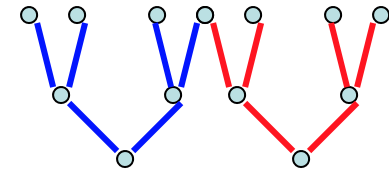
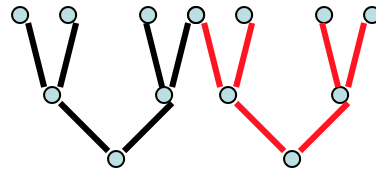
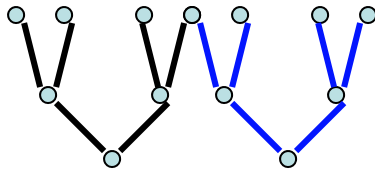
# Reconstructing global pedigrees: Superpedigrees

Steel and Hein, 2005

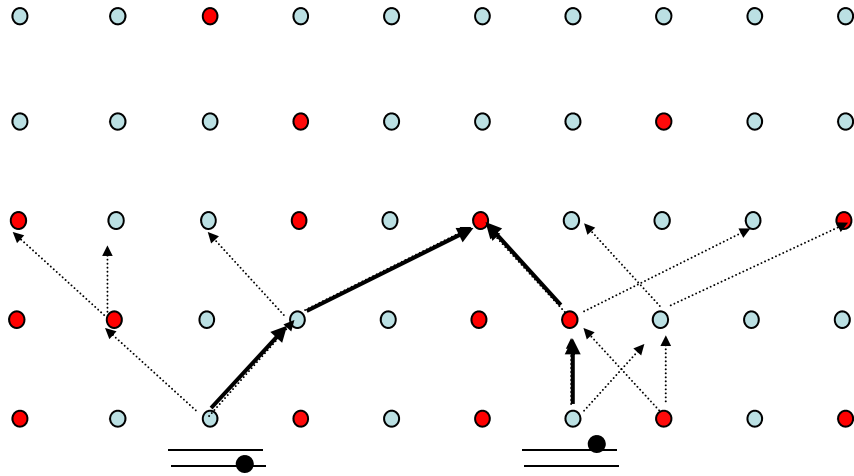
*The gender-labeled pedigrees for all pairs defines global pedigree*



*Gender-unlabeled pedigrees don't!!*

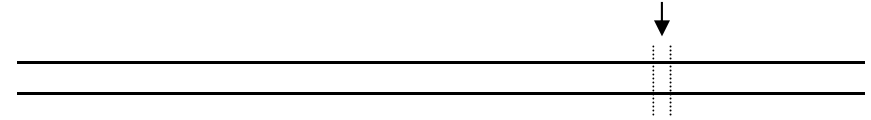


# Benevolent Mutation and Recombination Process



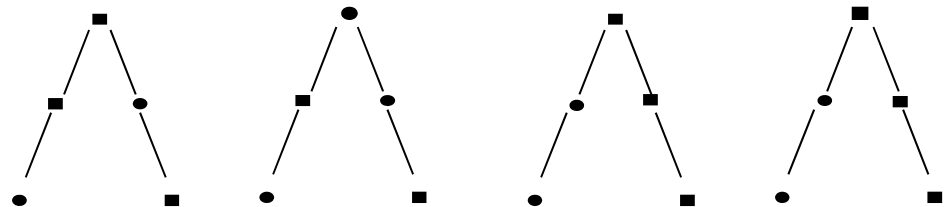
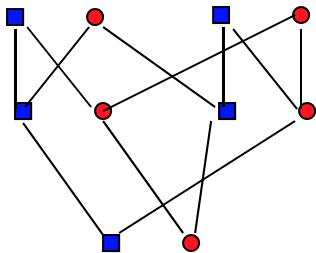
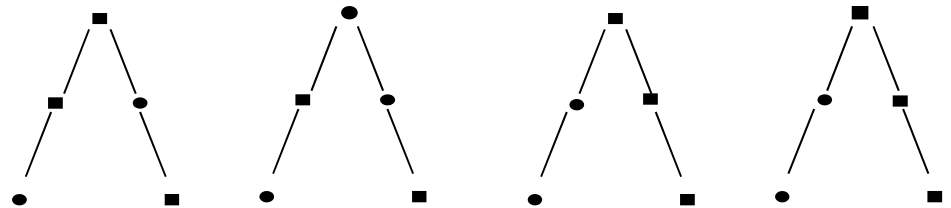
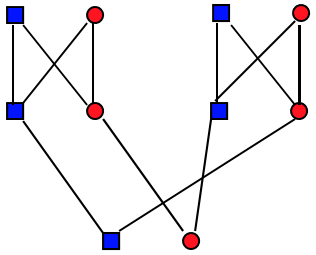
Genomes with  $r$  and  $m/r \rightarrow \infty$

$r$  - recombination rate,  $m$  - mutation rate



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

Counter example:



Embedded phylogenies:

# 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.