Loss of information at deeper divergences

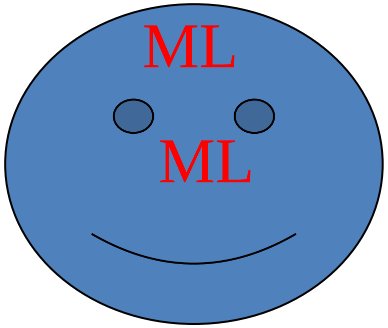
David Penny
Phylomania
Nov 2013

The mathematicos caused the problem!!!
Now they should solve it!

Okay, maybe we could help them,
Here are some ideas

Need relative – not absolute - information

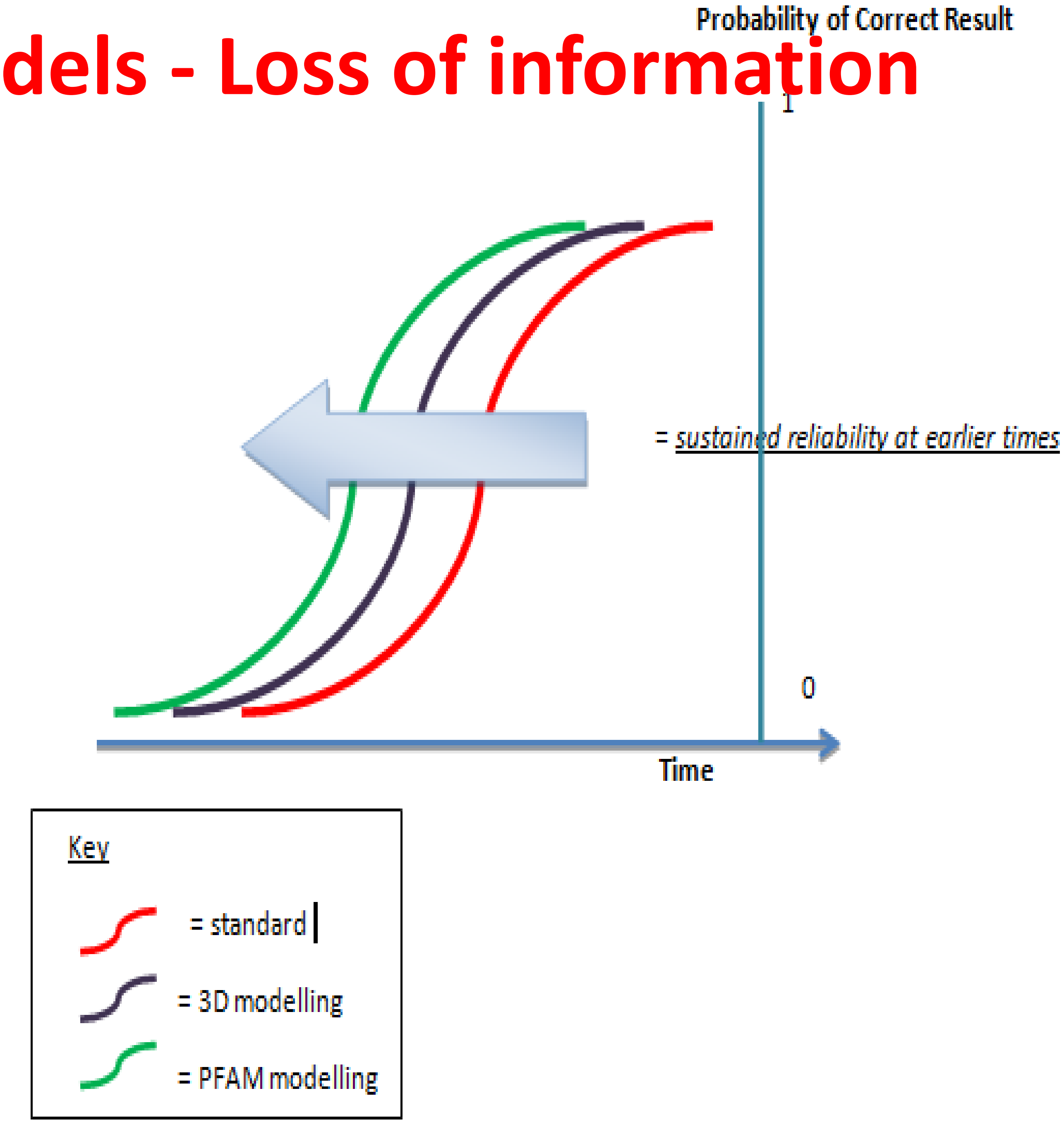
the comfort zone



ML Int			
ML Rel			
Ml_{av}		ML	
ML_{an}	MP	ML	
ML_{ep}	MP		MP
	popn phylogeny	classic phylogeny	deep

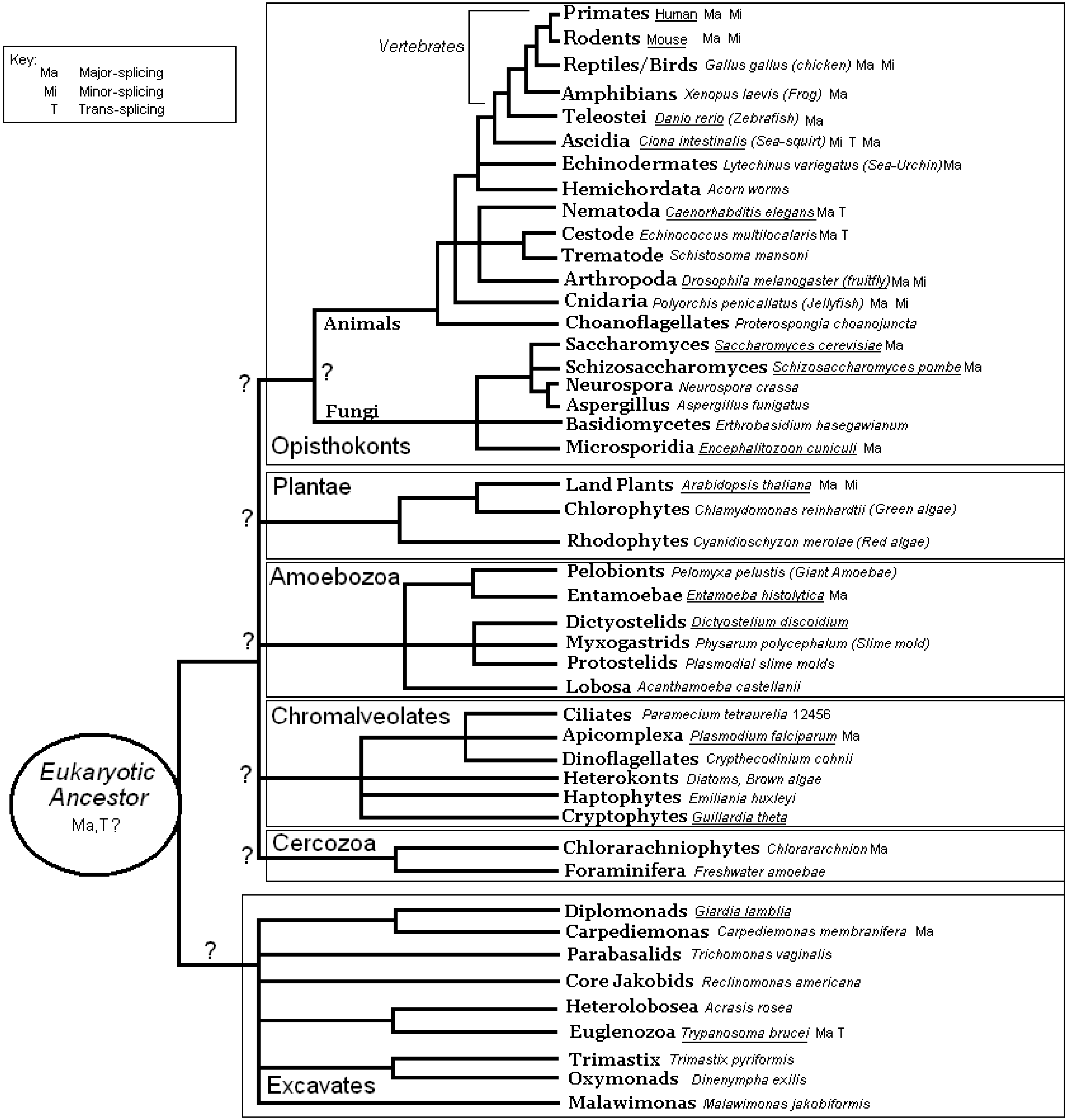
Markov models - Loss of information

can we go further back in time?

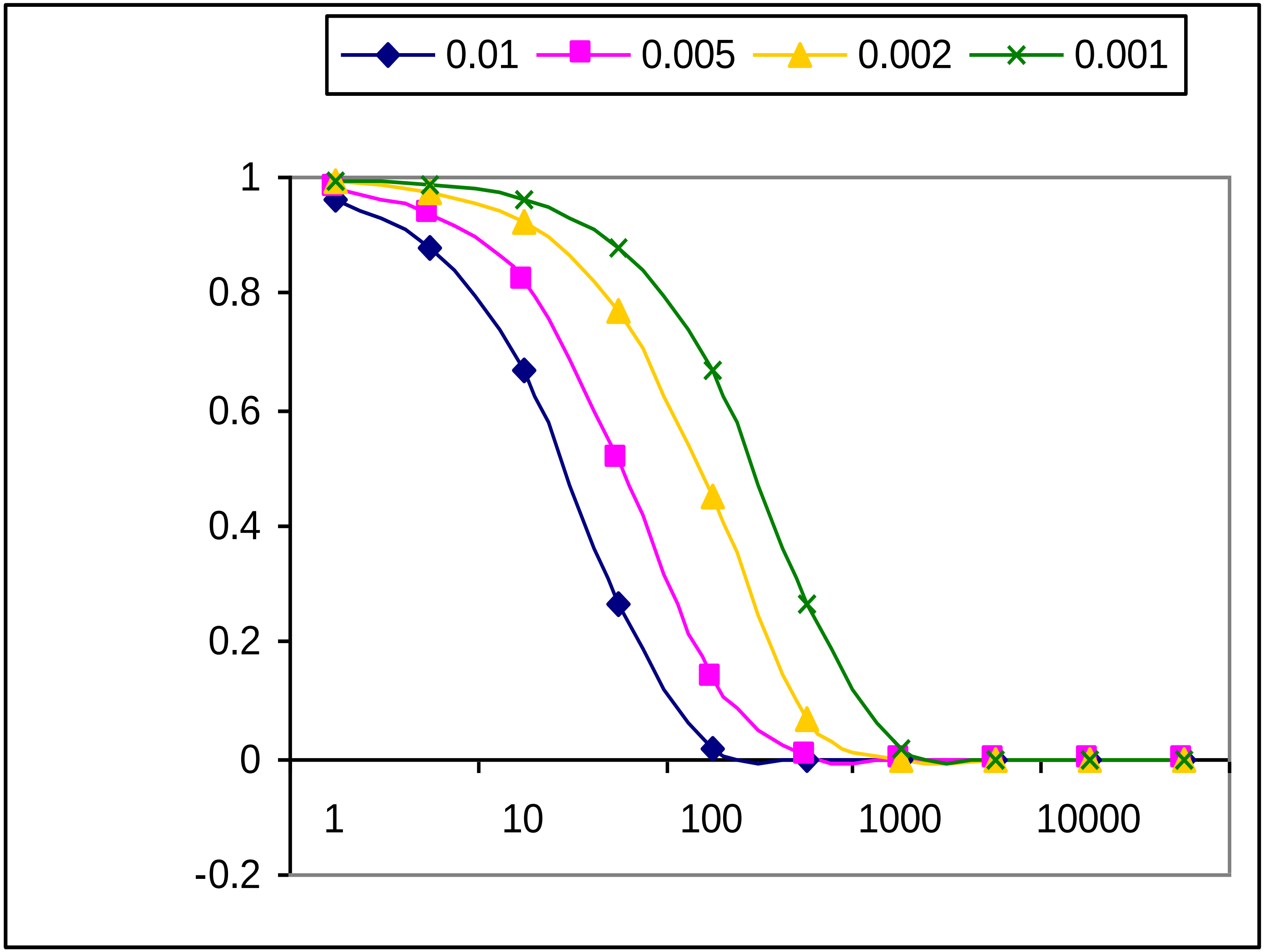


Key:
Ma Major-splicing
Mi Minor-splicing
T Trans-splicing

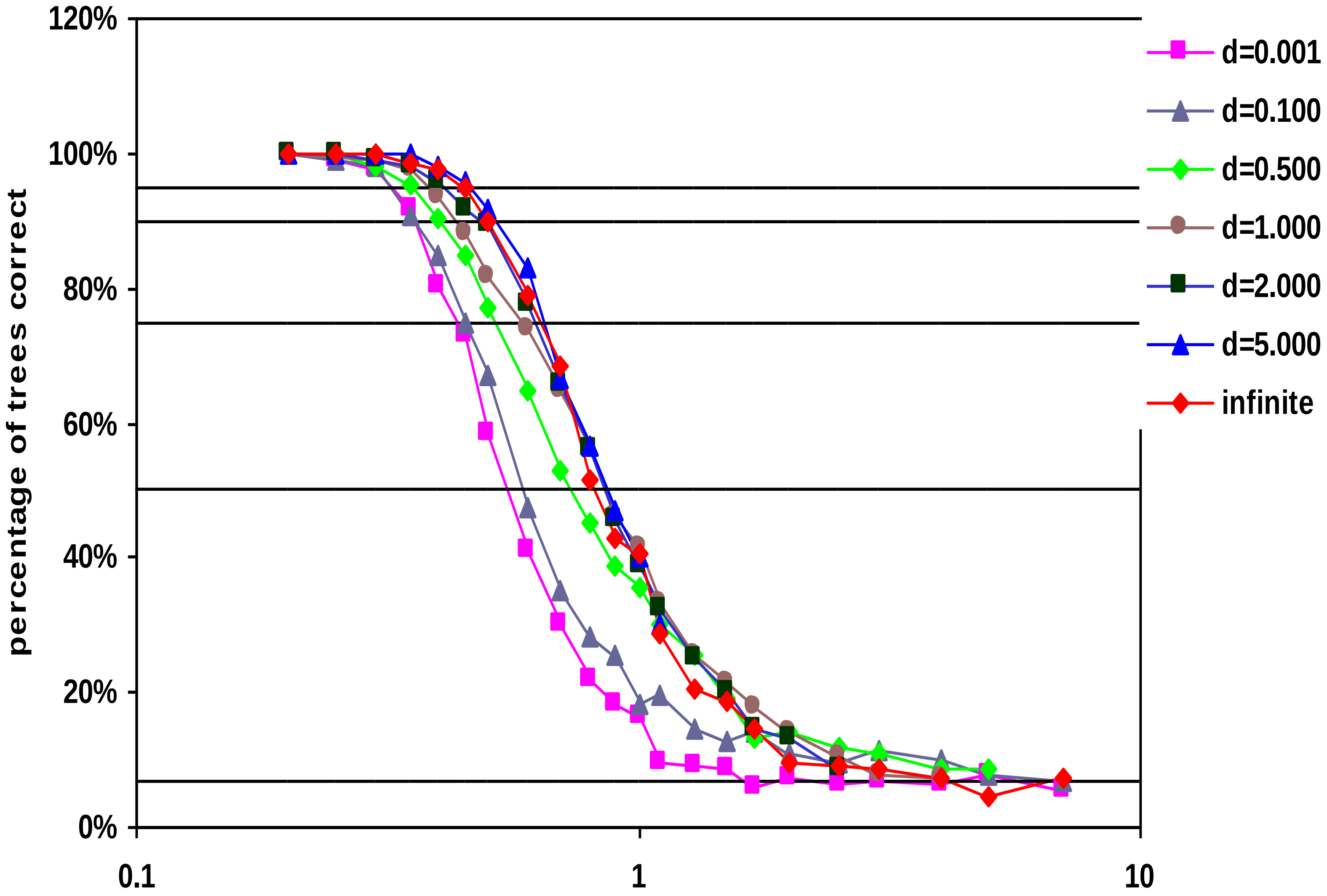
damned eukaryotes!



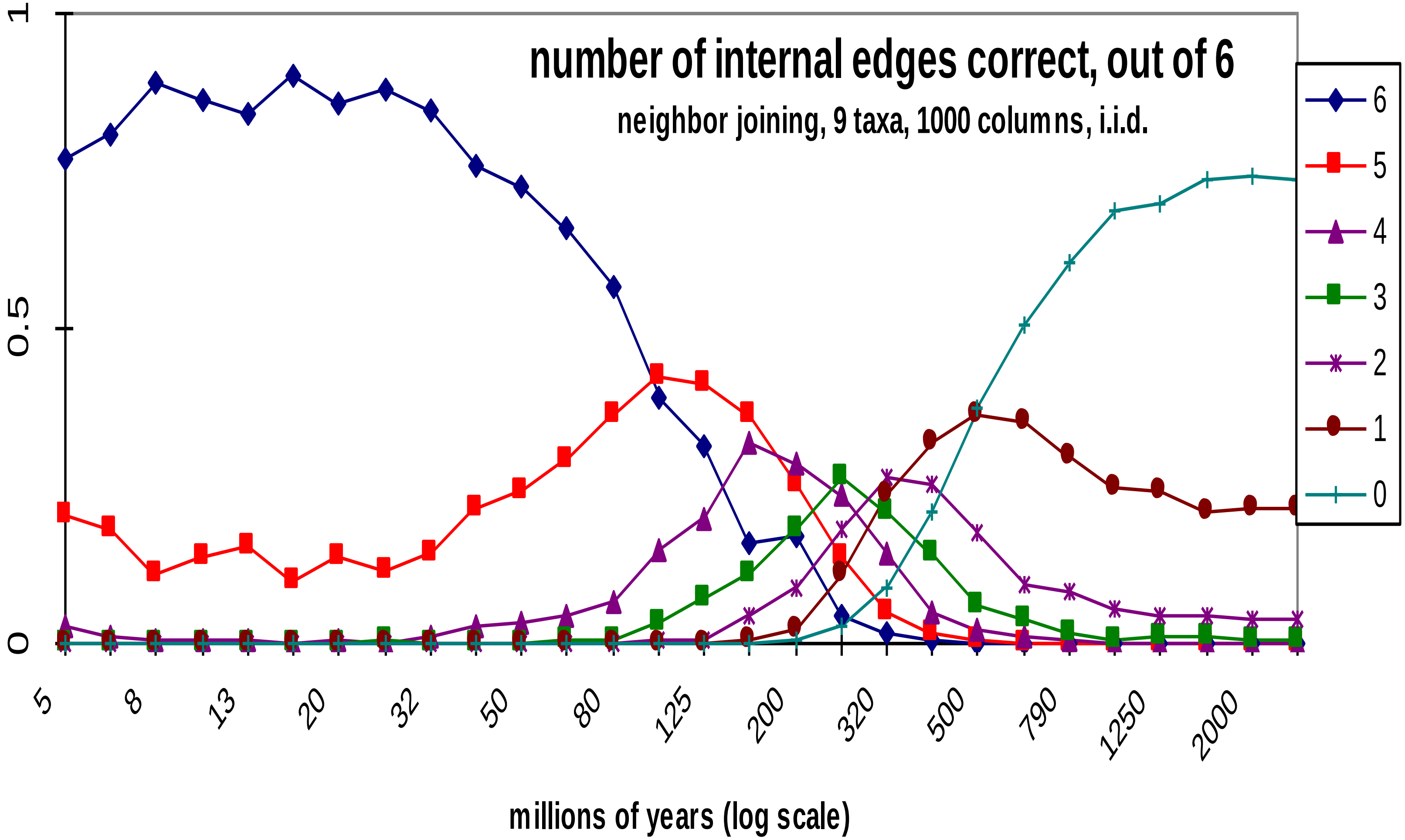
Calculated results, Δ ≤ ¼ + ne^{-qt}



1. simulation results with covarion model



simulation results with standard model



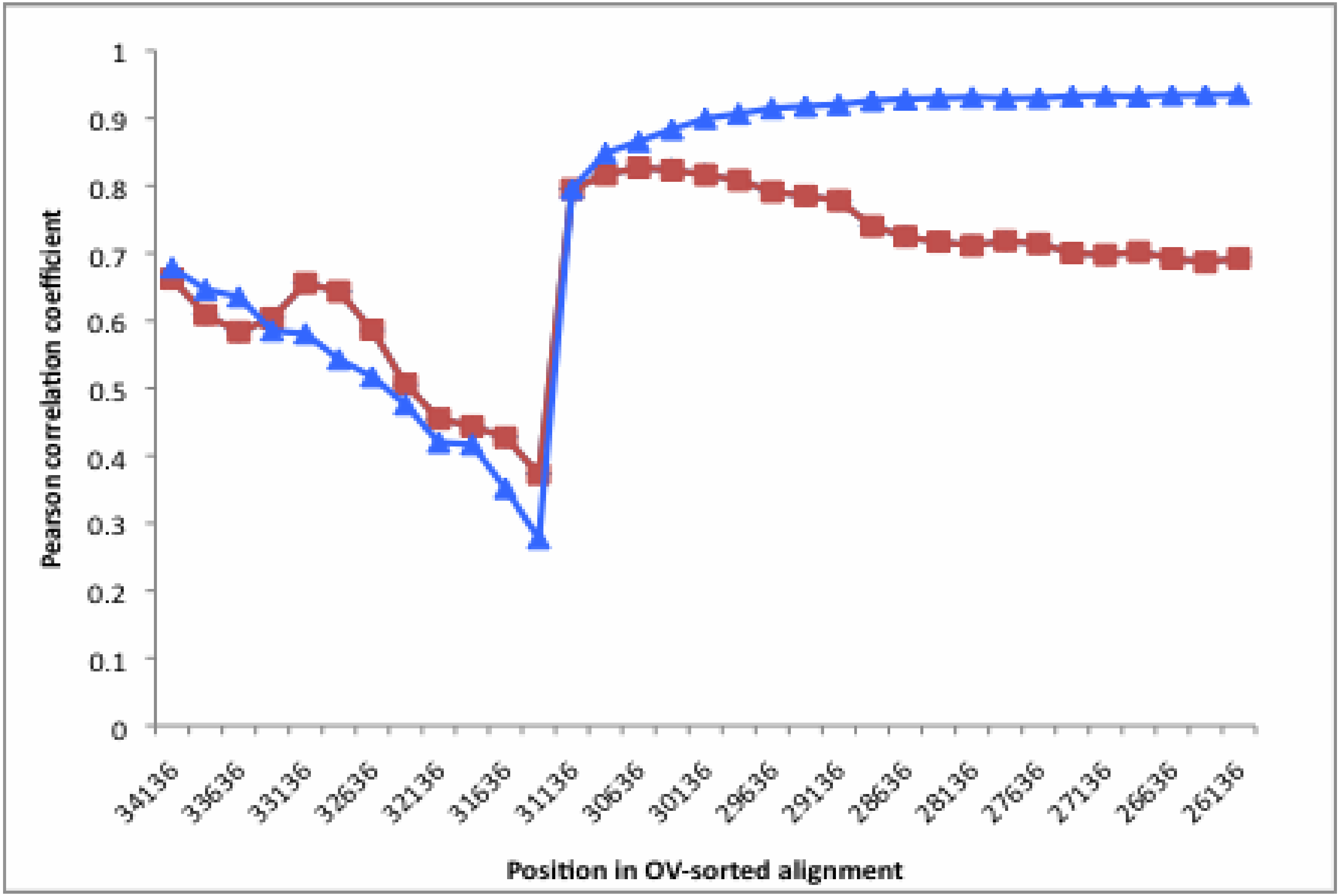
1 idea, delete fast sites

If there were a mixture of

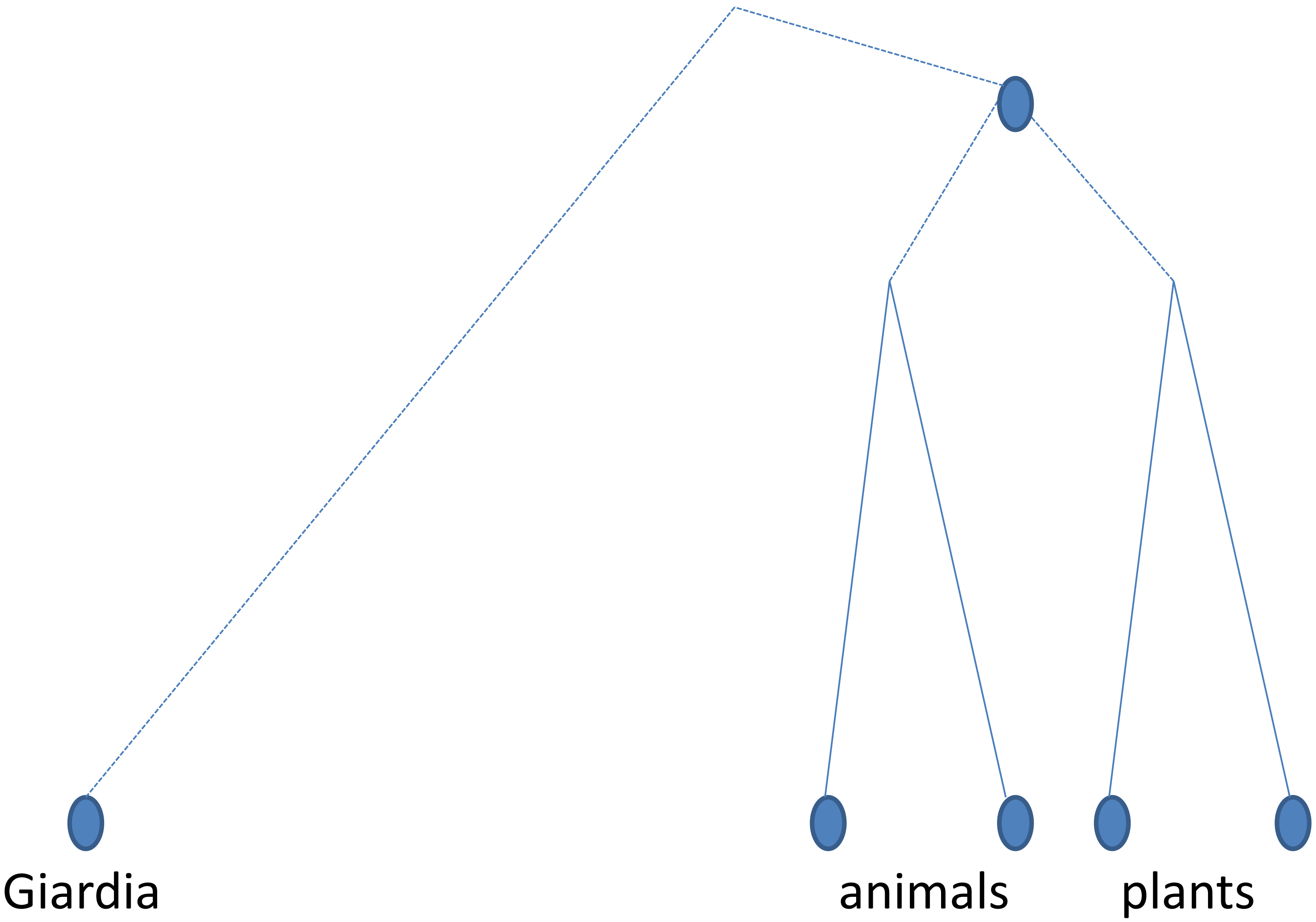
a) faster evolving sites, and
b) and we could identify them
c) and remove them

would that help go further back in time?

deleting faster sites

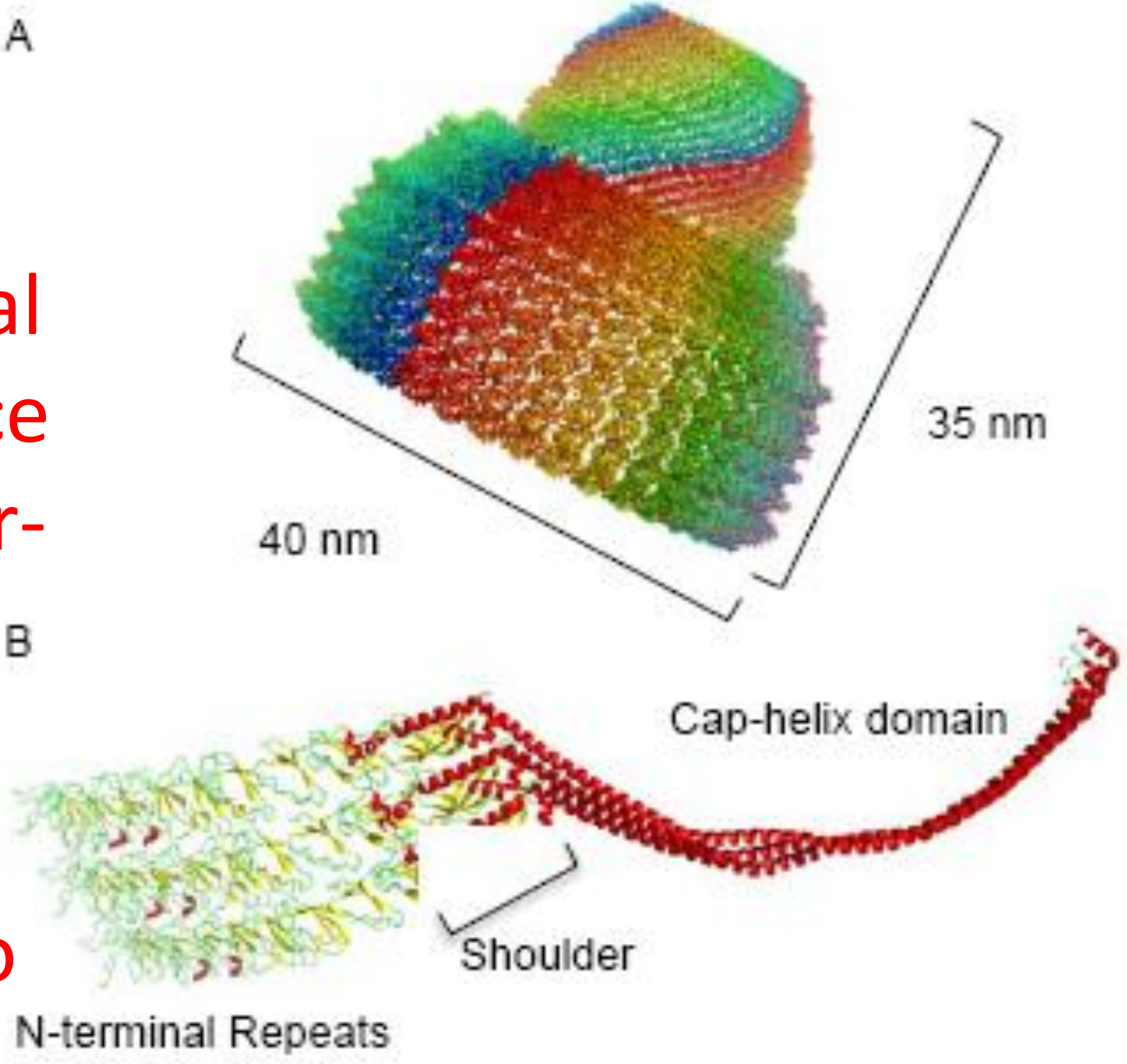


Ancestral Sequence Reconstruction

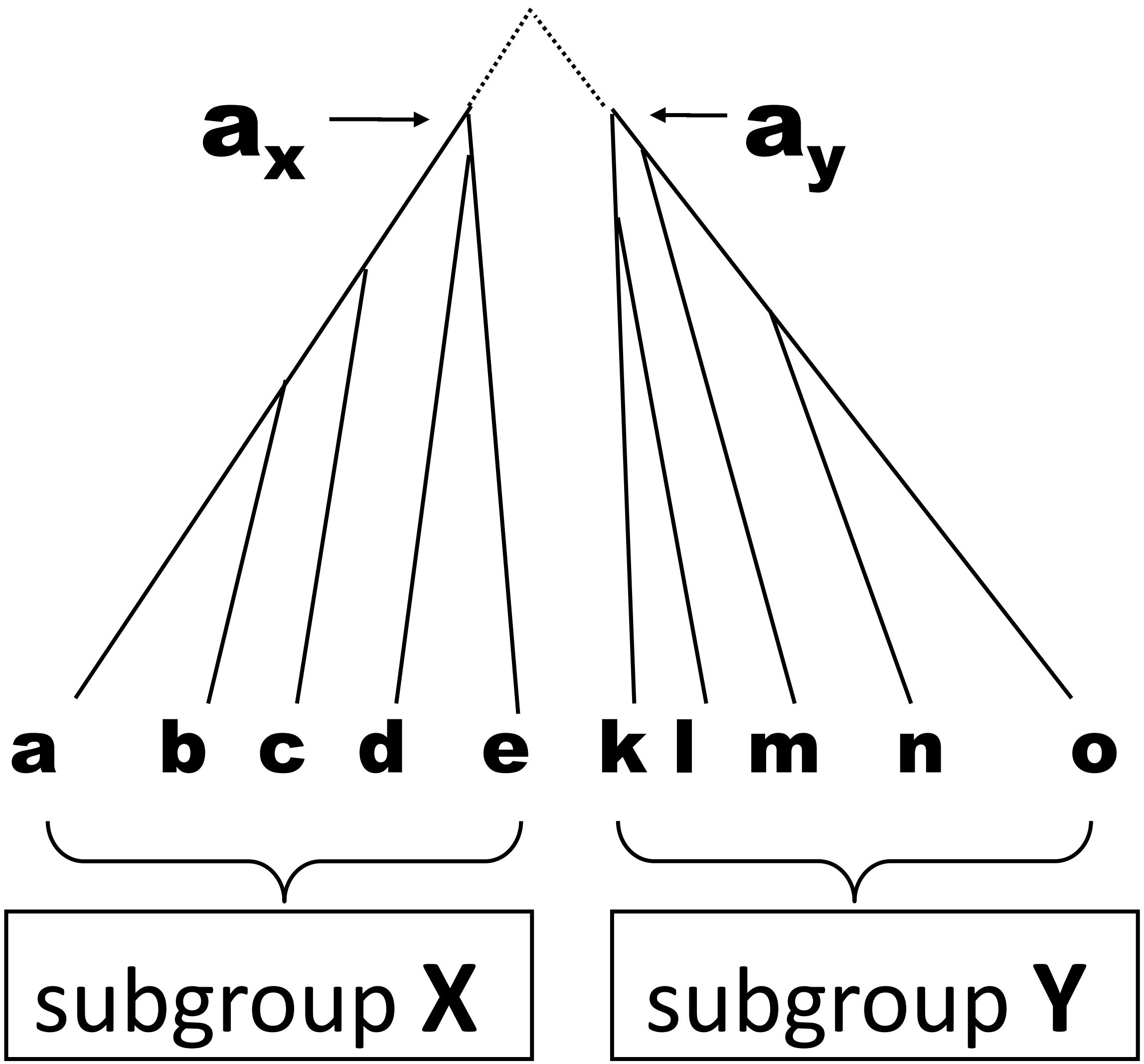


2, 3, testing Ancestral Sequence Reconstruction

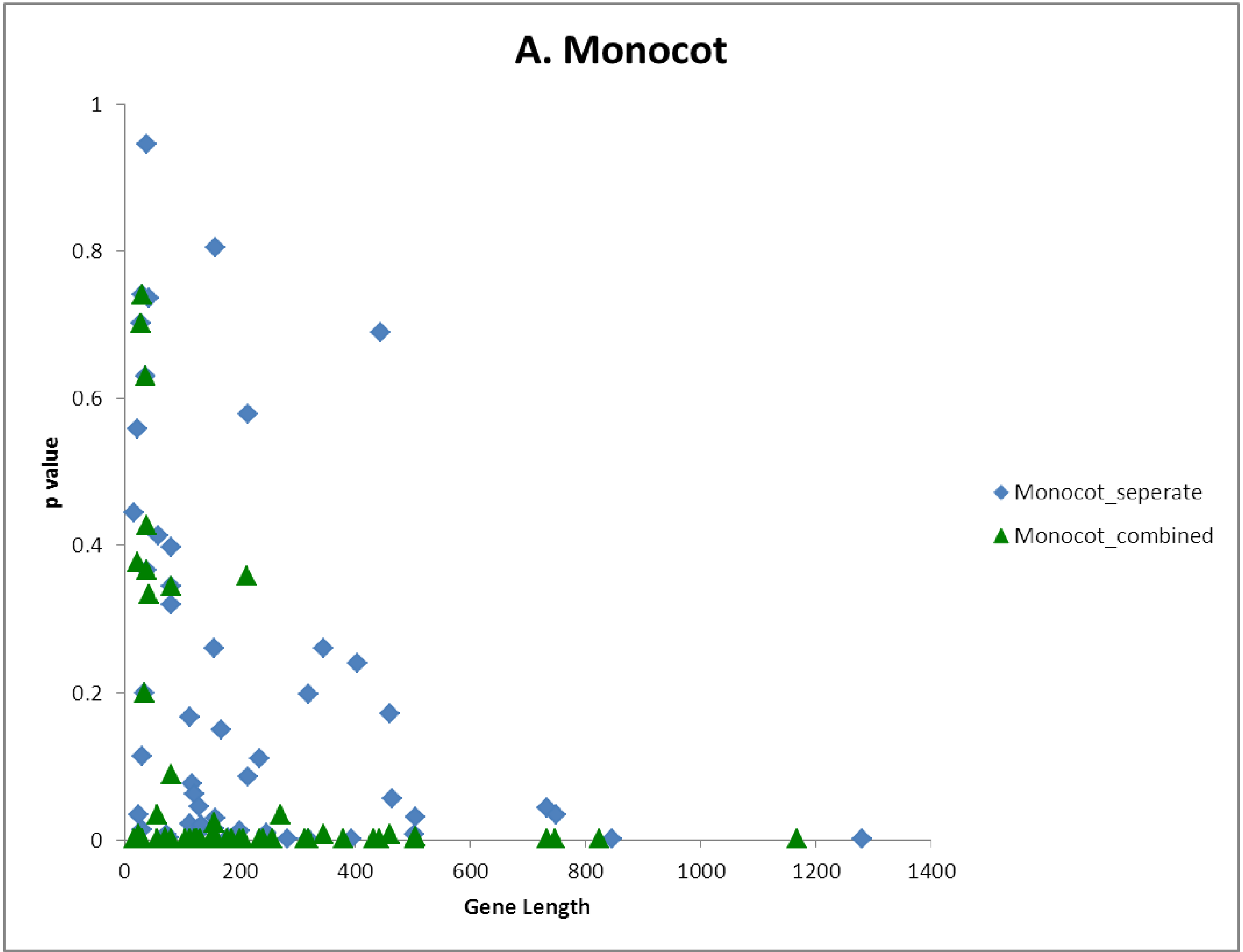
vaults
3-D info



subgroups X and Y

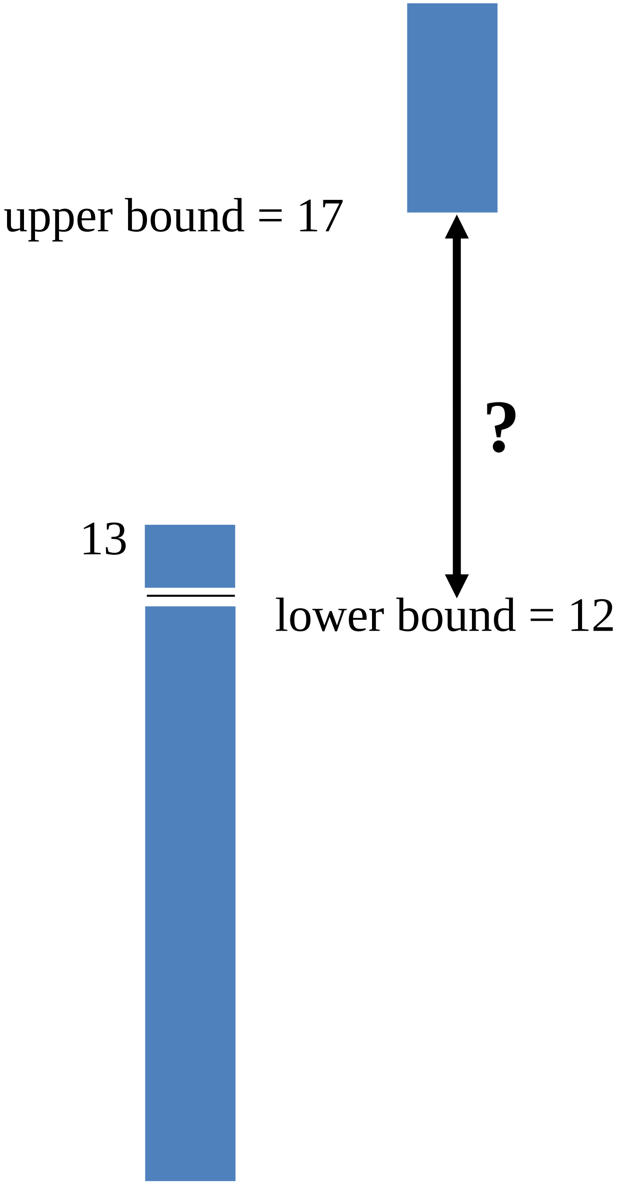


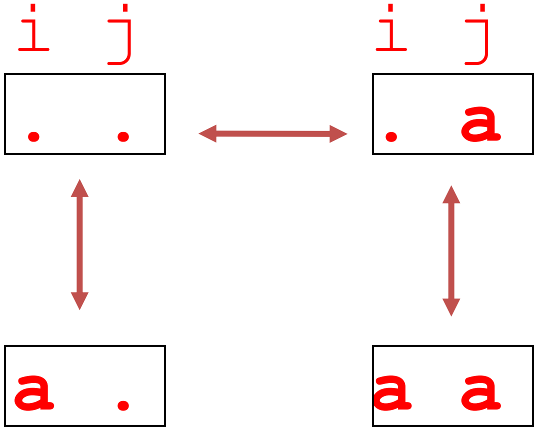
4. gene length vs similarity



5, Weighting

	i	*j*	3	4	5	6	7	8	9	0	1	2
f_1	a	.	.	.	.	a	.	.	.	.	.	.
f_2	.	.	.	.	.	a	.	.	.	.	a	.
f_3	.	.	.	.	.	a	.	.	.	.	.	a
f_4	.	.	.	.	.	a	.	.	.	.	.	.
g_1	.	a	a	a	a	.	a	a	.	.	.	.
g_2	a	a	a	a	a	.	.	a	.	.	a	.
h_1	.	.	a	a	a	.	.	.	a	a	.	.
h_2	.	.	a	.	a	.	.	.	a	a	.	.
h_3	a	.	.	a	a	.	.	.	a	a	.	.





Would weighting by incompatibilities help?

information from sequence order not used

Alignment original sequence order	Reordered Alignment shuffled/reordered
AIIFLNSALGPSPELFPIILATKVL	ASAGPSPPATPLLIIIIILLFFNEKV
AIMFLNSALGPPTELFPVILATKVL	ASAGPPTPATPLLIMVILLFFNEKV
SIMFLNHTLNPTPELFPIILATETL	SHTNPTPPATPLLIMIILLFFNEET
TILFLNSSLGLQPEVTPTVLATKTL	TSSGLQPPATPLLILTVLVTFNEKT
TLLFLNSMLKPPSELFPIILATKTL	TSMKPPSPATPLLLLIILLFFNEKT
ALLFLNSTLNPPTELFPLILATKTL	ASTNPPTPATPLLLLLILLFFNEKT
AILFLNSFLNPPKEFFPIILATKIL	ASFNPPKPATPLLILIILFFFNEKI

c! ways to reorder alignment

shuffle by columns & by taxa

6. So, could we use ‘words’ of 2, 3, 4, 5, … letters

7. Alphabet reduction