

# Gap (indel) models

# Using gap characters

GARLI implements two models appropriate for use on gap characters in *fixed alignments*

DIMM model based on Rivas and Eddy (2008)

- Real model of insertion-deletion process
- Non-reversible, dollo (one insert, multiple delete)

Variant of  $Mk_v$  (Lewis, 2001) model (no full gap columns)

- Allows many gap-base transitions

# Fixed alignment analysis options

## gapped alignment

Taxon \ Character	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
boylliMVZ148929	C	G	C	-	-	T	A	T	A	G	T	A	T	A	-	-	-	C	C	A	C	A	T	T	A	C	C	T	T	-	-	G	A																																																																			
luteiventris_MT	C	G	C	-	-	T	A	T	-	T	A	-	T	C	-	T	C	-	C	C	A	C	A	T	T	A	C	C	T	T	-	-	C	A																																																																		
muscosaMVZ149006	C	G	C	-	-	T	A	T	-	T	A	-	T	C	-	C	C	A	C	A	T	T	A	C	C	T	T	-	-	C	A																																																																					
cascadæMVZ148946	C	G	C	-	-	T	A	T	-	T	A	-	T	C	-	T	C	-	C	C	A	C	A	T	T	A	C	C	T	T	-	-	C	A																																																																		
temporariaDMH84R1	C	G	T	C	-	T	A	T	-	T	A	-	T	C	-	T	C	-	C	C	A	C	A	T	T	A	C	C	T	T	-	-	G	A																																																																		
svylvaticaMVZ13742	C	G	C	-	-	T	A	T	-	T	A	-	T	C	-	T	C	-	C	C	A	C	A	T	T	A	C	C	T	T	-	-	C	A																																																																		
svylvaticaMVZ13964	C	G	C	-	-	T	A	T	-	T	A	-	T	C	-	T	C	-	C	C	A	C	A	T	T	A	C	C	T	T	-	-	C	A																																																																		
septentrionalis	C	G	C	-	-	T	A	T	-	T	A	-	T	C	-	T	C	-	C	C	A	C	A	T	T	A	C	C	T	T	-	-	T	A																																																																		
catesbeianaDH84R2	C	G	C	-	-	T	A	T	-	T	A	-	T	C	-	T	C	-	C	C	A	C	A	T	T	A	C	C	T	T	-	-	C	A																																																																		
virgatipesMVZ1759	C	G	C	-	-	T	A	T	-	T	A	-	T	C	-	T	C	-	C	C	A	C	A	T	T	A	C	C	T	T	-	-	C	A																																																																		
palmipesVenAMNH	C	G	C	-	-	T	A	T	-	T	A	-	T	C	-	T	C	-	C	C	A	C	A	T	T	A	C	C	T	T	-	-	G	A																																																																		
palmipesEcukU20442	C	G	C	-	-	T	A	T	-	T	A	-	T	C	-	T	C	-	C	C	A	C	A	T	T	A	C	C	T	T	-	-	C	A																																																																		
Sp_1_ecuadorQCA	C	G	C	-	-	T	A	T	-	T	A	-	T	C	-	T	C	-	C	C	A	C	A	T	T	A	C	C	T	T	-	-	C	A																																																																		
bwanaQCAZ13964	C	G	C	-	-	T	A	T	-	T	A	-	T	C	-	T	C	-	C	C	G	A	C	A	T	T	A	C	C	T	T	-	-	C	A																																																																	
julianiTNHG0324	C	G	C	-	-	T	A	T	-	T	A	-	T	C	-	T	C	-	C	C	A	C	A	T	T	A	C	C	T	T	-	-	C	A																																																																		
vallantiKU195299	C	G	C	-	-	T	A	T	-	T	A	-	T	C	-	T	C	-	C	C	A	C	A	T	T	A	C	C	T	T	-	-	C	A																																																																		
sierramadrensisKU	C	G	T	-	-	T	A	T	-	T	A	-	T	C	-	T	C	-	C	C	A	C	A	T	T	A	C	C	T	T	-	-	G	A																																																																		

↓

ignore gaps, e.g. GTR  
(the usual)

# Fixed alignment analysis options

## gapped alignment

Taxon \ Character	C	G	C	-	-	T	A	T	A	G	T	A	-	-	T	C	C	A	C	A	T	T	A	C	C	T	T	T	-	-	G	A		
boylliMVZ148929	C	G	C	-	-	T	A	T	A	G	T	A	-	-	T	C	C	A	C	A	T	T	A	C	C	T	T	T	-	-	G	A		
luteiventris_MT	C	G	C	-	-	T	A	T	-	T	A	-	T	C	-	T	C	-	C	C	A	C	A	T	T	A	C	T	T	-	-	G	A	
muscosaMVZ149006	C	G	C	-	-	T	A	T	-	T	A	-	T	C	-	T	C	C	C	A	C	A	T	T	A	C	T	T	-	-	G	A		
cascadæMVZ148946	C	G	C	-	-	T	A	T	-	T	A	-	T	C	-	T	C	C	C	A	C	A	T	T	A	C	T	T	-	-	G	A		
temporariaDMH84R1	C	G	T	C	-	-	T	A	T	-	T	A	-	T	C	-	T	C	C	C	A	C	A	T	T	A	C	T	-	-	G	A		
sylvaticaMVZ13742	T	G	C	-	-	-	-	-	T	A	-	T	A	-	T	C	-	T	C	C	C	A	C	A	T	T	T	T	C	A	T	A		
sylvaticaDMH84R3	T	G	C	-	-	-	-	-	T	A	-	T	A	-	T	C	-	T	C	C	C	A	C	A	T	A	C	T	T	T	C	A		
septentrionales	T	G	C	-	-	-	-	-	T	A	-	T	A	-	T	C	-	T	C	C	C	A	C	A	T	A	C	T	T	T	C	A		
catesbeianaDMH84R2	T	G	C	-	-	-	-	-	T	A	-	T	A	-	T	C	-	T	C	C	C	A	C	A	T	A	C	T	T	T	C	A		
virgatipesMVZ1759	T	G	C	-	-	-	-	-	T	A	-	T	A	-	T	C	-	T	C	C	C	A	C	A	T	A	C	T	T	T	C	A		
palmipesVenAMNH	C	G	C	-	-	T	A	T	-	T	A	C	T	-	T	C	-	T	C	C	C	A	C	A	T	T	A	C	T	T	-	-	G	A
palmipesEcukU20442	C	G	C	-	-	T	A	T	-	T	A	C	T	-	T	C	-	T	C	C	C	A	C	A	T	T	A	C	T	T	-	-	G	A
Sp_1_ecuadorQCA	C	G	C	-	-	T	A	T	-	T	A	C	T	-	T	C	-	T	C	C	C	A	C	A	T	T	A	C	T	T	-	-	G	A
bwanaQCAZ13964	C	G	C	-	-	T	A	T	-	T	A	C	T	-	T	C	-	T	C	C	C	A	C	A	T	T	A	C	T	T	-	-	G	A
julianiTNHC60324	C	G	C	C	G	A	T	A	-	T	A	A	T	-	T	C	-	T	C	C	C	A	C	A	T	T	A	C	T	T	-	-	G	A
vallantiKU195299	C	G	C	-	-	T	A	T	-	T	A	A	T	-	T	C	-	T	C	C	C	A	C	A	T	T	A	C	T	T	-	-	G	A
sierramadrensisKU	C	G	T	-	-	T	A	T	-	T	A	C	A	T	T	C	-	T	C	A	C	A	T	T	A	C	T	T	-	-	G	A		



Taxon \ Character	C	G	C	-	-	T	A	T	A	G	T	A	-	-	T	C	C	A	C	A	T	T	A	C	C	T	T	T	-	-	G	A		
boylliMVZ148929	C	G	C	-	-	T	A	T	A	G	T	A	-	-	T	C	C	A	C	A	T	T	A	C	C	T	T	T	-	-	G	A		
luteiventris_MT	C	G	C	-	-	T	A	T	-	T	A	-	T	C	-	T	C	C	A	C	A	T	C	C	T	T	T	-	-	G	A			
muscosaMVZ149006	C	G	C	-	-	T	A	T	-	T	A	-	T	C	-	T	C	C	A	C	A	T	C	C	T	T	T	-	-	G	A			
cascadæMVZ148946	C	G	C	-	-	T	A	T	-	T	A	-	T	C	-	T	C	C	C	A	C	A	T	C	C	T	T	-	-	G	A			
temporariaDMH84R1	C	G	T	C	-	-	T	A	T	-	T	A	-	T	C	-	T	C	C	C	A	C	A	T	C	T	T	-	-	G	A			
sylvaticaMVZ13742	T	G	C	-	-	-	-	-	T	A	-	T	A	-	T	C	-	T	C	C	C	A	C	A	T	T	T	C	A	T	A			
sylvaticaDMH84R3	T	G	C	-	-	-	-	-	T	A	-	T	A	-	T	C	-	T	C	C	C	A	C	A	T	A	C	T	T	T	C	A		
septentrionales	T	G	C	-	-	-	-	-	T	A	-	T	A	-	T	C	-	T	C	C	C	A	C	A	T	A	C	T	T	T	C	A		
catesbeianaDMH84R2	T	G	C	-	-	-	-	-	T	A	-	T	A	-	T	C	-	T	C	C	C	A	C	A	T	A	C	T	T	T	C	A		
virgatipesMVZ1759	T	G	C	-	-	-	-	-	T	A	-	T	A	-	T	C	-	T	C	C	C	A	C	A	T	A	C	T	T	T	C	A		
palmipesVenAMNH	C	G	C	-	-	T	A	T	-	T	A	C	T	-	T	C	-	T	C	C	C	A	C	A	T	T	A	C	T	T	-	-	G	A
palmipesEcukU20442	C	G	C	-	-	T	A	T	-	T	A	C	T	-	T	C	-	T	C	C	C	A	C	A	T	T	A	C	T	T	-	-	G	A
Sp_1_ecuadorQCA	C	G	C	-	-	T	A	T	-	T	A	C	T	-	T	C	-	T	C	C	C	A	C	A	T	T	A	C	T	T	-	-	G	A
bwanaQCAZ13964	C	G	C	-	-	T	A	T	-	T	A	C	T	-	T	C	-	T	C	C	C	A	C	A	T	T	A	C	T	T	-	-	G	A
julianiTNHC60324	C	G	C	G	A	T	A	-	T	A	A	T	-	T	C	-	T	C	C	C	A	C	A	T	C	T	T	-	-	G	A			
vallantiKU195299	C	G	C	-	-	T	A	T	-	T	A	A	T	-	T	C	-	T	C	C	C	A	C	A	T	C	T	T	-	-	G	A		
sierramadrensisKU	C	G	T	-	-	T	A	T	-	T	A	C	A	T	T	C	-	T	C	A	C	A	T	T	A	C	T	T	-	-	G	A		

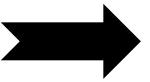
5-state approach  
(e.g., Rivas and Eddy)

ignore gaps, e.g. GTR  
(the usual)

Taxon \ Character	C	G	C	-	-	T	A	T	A	G	T	A	-	-	T	C	C	A	C	A	T	T	A	C	C	T	T	-	-	G	A			
boylliMVZ148929	C	G	C	-	-	T	A	T	A	G	T	A	-	-	T	C	C	A	C	A	T	T	A	C	C	T	T	-	-	G	A			
luteiventris_MT	C	G	C	-	-	T	A	T	-	T	A	-	T	C	-	T	C	C	A	C	A	T	C	C	T	T	-	-	G	A				
muscosaMVZ149006	C	G	C	-	-	T	A	T	-	T	A	-	T	C	-	T	C	C	C	A	C	A	T	C	C	T	-	-	G	A				
cascadæMVZ148946	C	G	C	-	-	T	A	T	-	T	A	-	T	C	-	T	C	C	C	A	C	A	T	C	C	T	-	-	G	A				
temporariaDMH84R1	C	G	T	C	-	-	T	A	T	-	T	A	-	T	C	-	T	C	C	C	A	C	A	T	C	T	-	-	G	A				
sylvaticaMVZ13742	T	G	C	-	-	-	-	-	T	A	-	T	A	-	T	C	-	T	C	C	C	A	C	A	T	T	C	A	T	A				
sylvaticaDMH84R3	T	G	C	-	-	-	-	-	T	A	-	T	A	-	T	C	-	T	C	C	C	A	C	A	T	A	C	T	T	C	A			
septentrionales	T	G	C	-	-	-	-	-	T	A	-	T	A	-	T	C	-	T	C	C	C	A	C	A	T	A	C	T	T	T	C	A		
catesbeianaDMH84R2	T	G	C	-	-	-	-	-	T	A	-	T	A	-	T	C	-	T	C	C	C	A	C	A	T	A	C	T	T	C	A			
virgatipesMVZ1759	T	G	C	-	-	-	-	-	T	A	-	T	A	-	T	C	-	T	C	C	C	A	C	A	T	A	C	T	T	T	C	A		
palmipesVenAMNH	C	G	C	-	-	T	A	T	-	T	A	C	T	-	T	C	-	T	C	C	C	A	C	A	T	T	A	C	T	T	-	-	G	A
palmipesEcukU20442	C	G	C	-	-	T	A	T	-	T	A	C	T	-	T	C	-	T	C	C	C	A	C	A	T	T	A	C	T	T	-	-	G	A
Sp_1_ecuadorQCA	C	G	C	-	-	T	A	T	-	T	A	C	T	-	T	C	-	T	C	C	C	A	C	A	T	T	A	C	T	T	-	-	G	A
bwanaQCAZ13964	C	G	C	-	-	T	A	T	-	T	A	C	T	-	T	C	-	T	C	C	C	A	C	A	T	T	A	C	T	T	-	-	G	A
julianiTNHC60324	C	G	C	G	A	T	A	-	T	A	A	T	-	T	C	-	T	C	C	C	A	C	A	T	C	T	T	-	-	G	A			
vallantiKU195299	C	G	C	-	-	T	A	T	-	T	A	A	T	-	T	C	-	T	C	C	C	A	C	A	T	C	T	T	-	-	G	A		
sierramadrensisKU	C	G	T	-	-	T	A	T	-	T	A	C	A	T	T	C	-	T	C	A	C	A	T	T	A	C	T	T	-	-	G	A		

# Fixed alignment analysis options

# gapped alignment



## **5-state approach** (e.g., Rivas and Eddy 2008)

**“gapcode”  
matrix  
&  
use partitioned  
model**

**ignore gaps, e.g. GTR  
(the usual)**

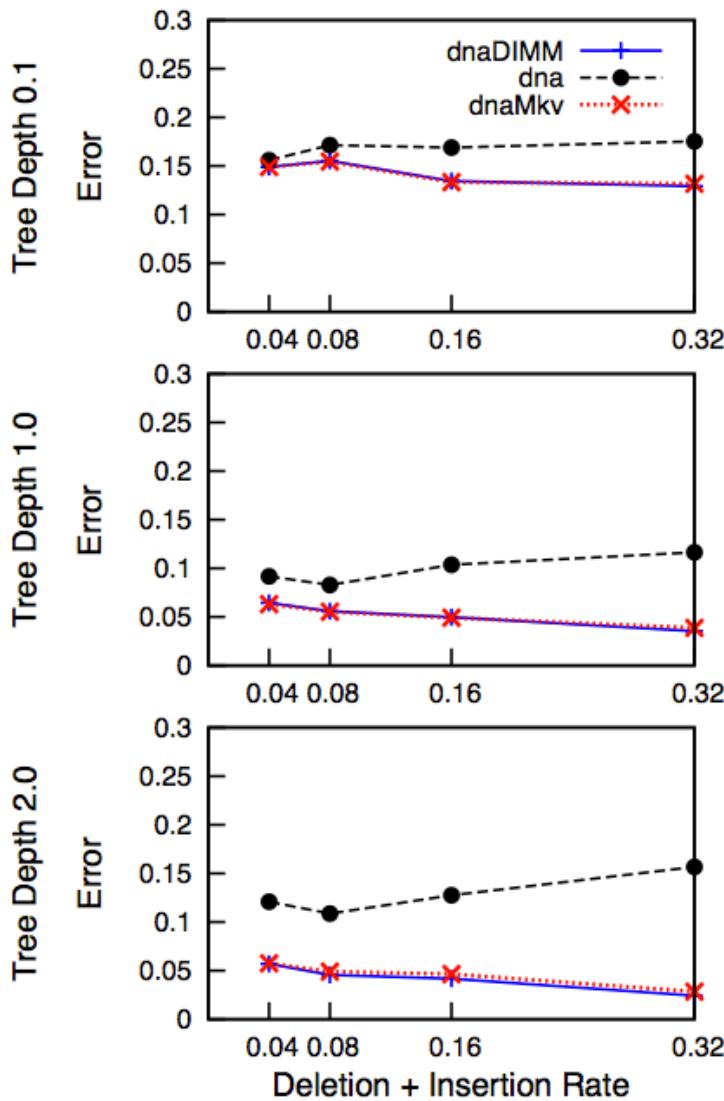
GTR

DIMM  
(or  $Mk_v$ )

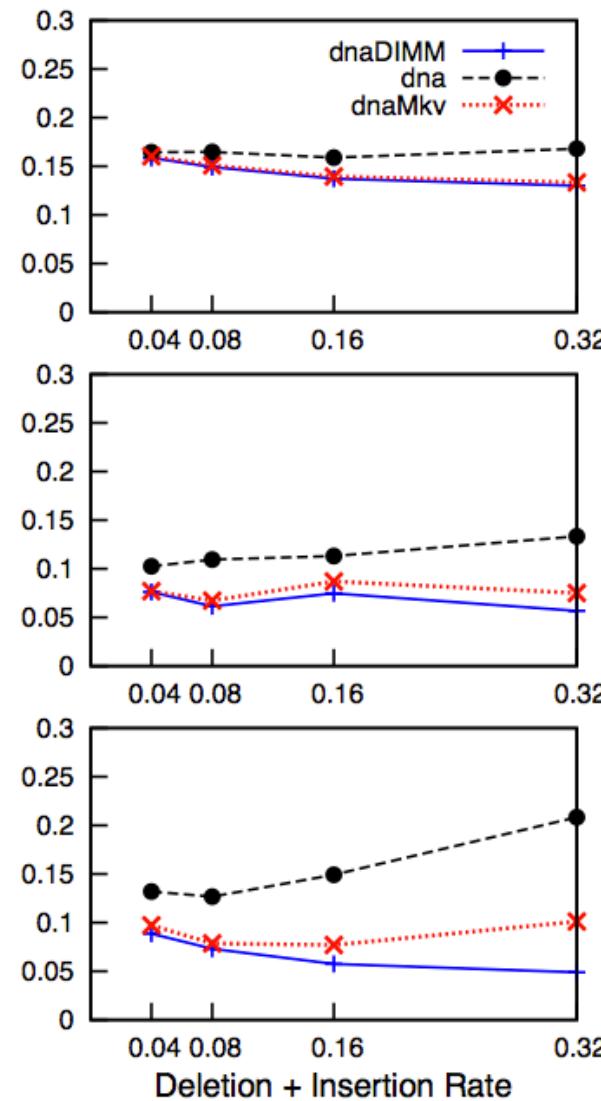
# Gap model accuracy – 64 taxa, true alignments

deeper trees

## single site indels

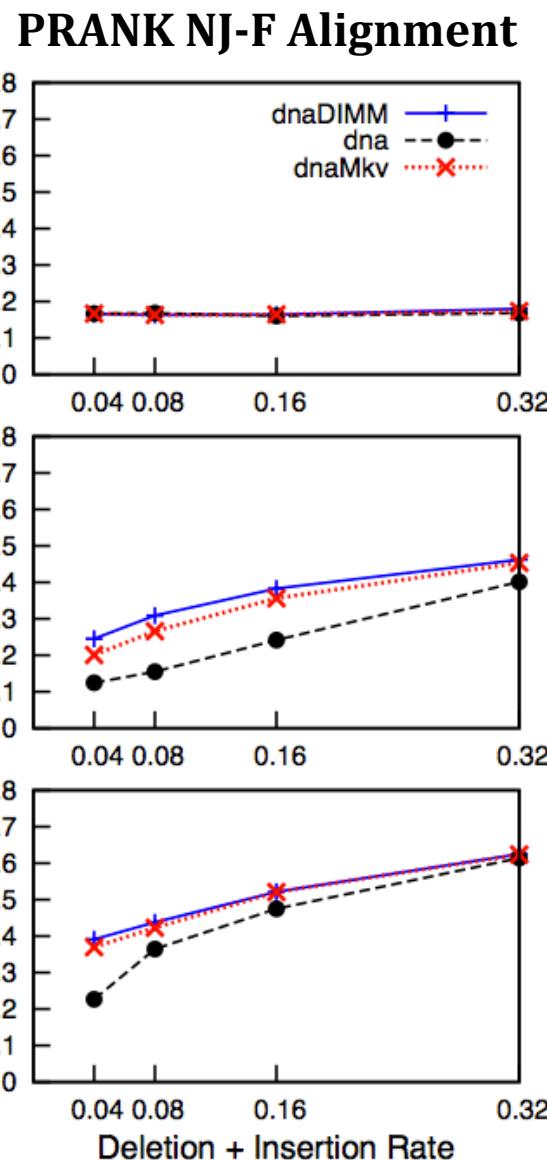
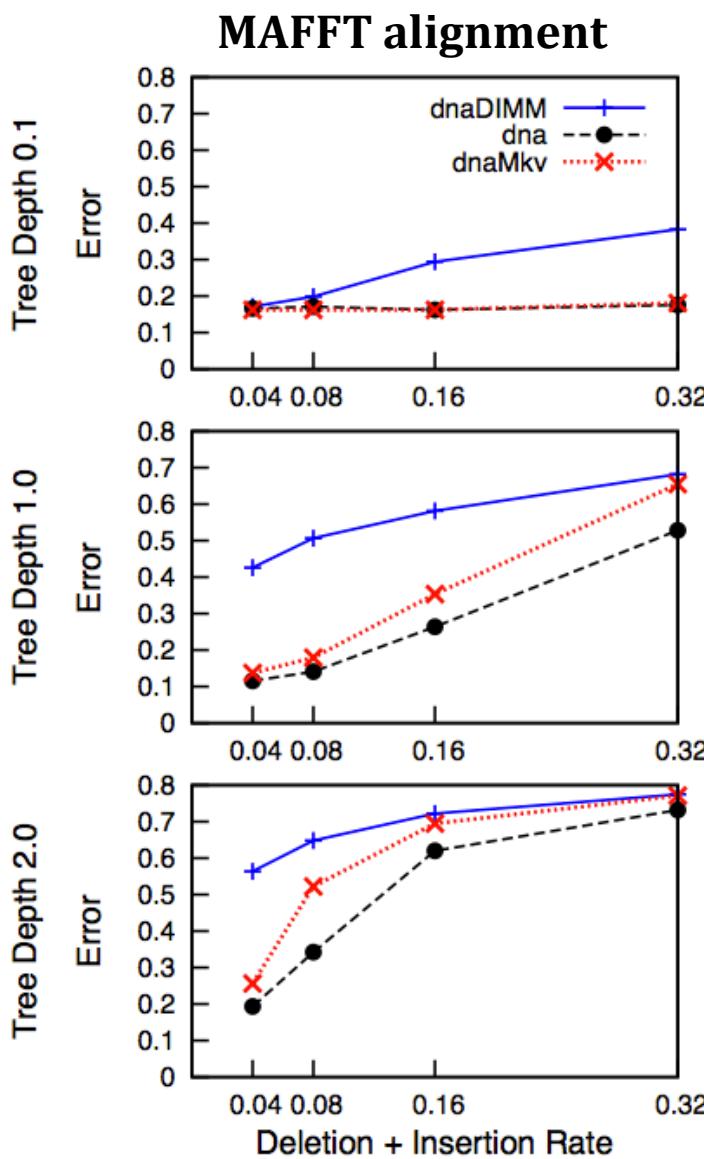


## multi-site indels



higher indel rates

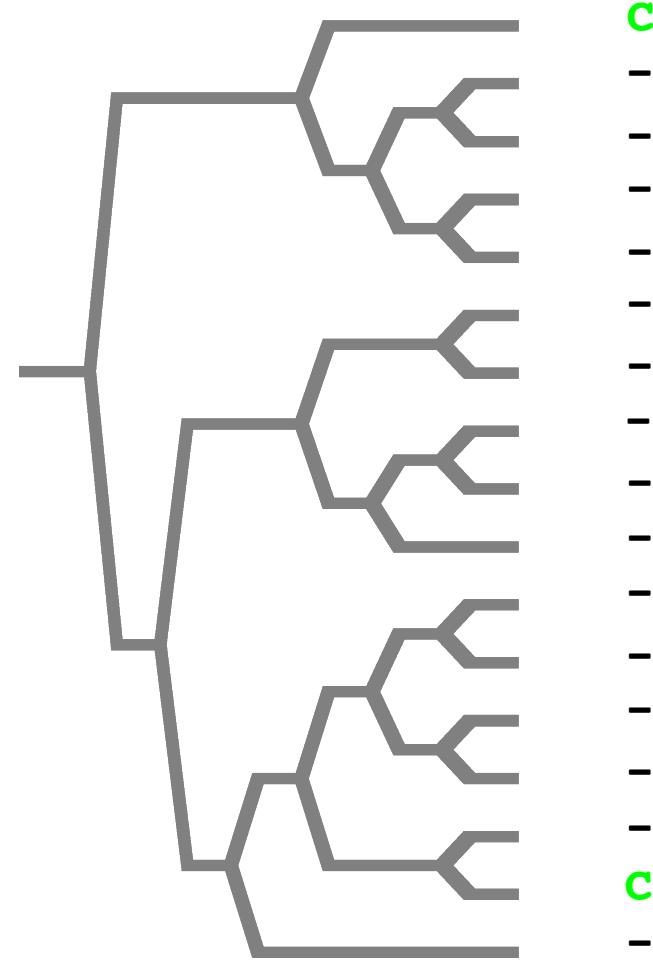
# Gap model accuracy – 64 taxa, estimated alignments



# Part of the problem: innocuous alignment errors

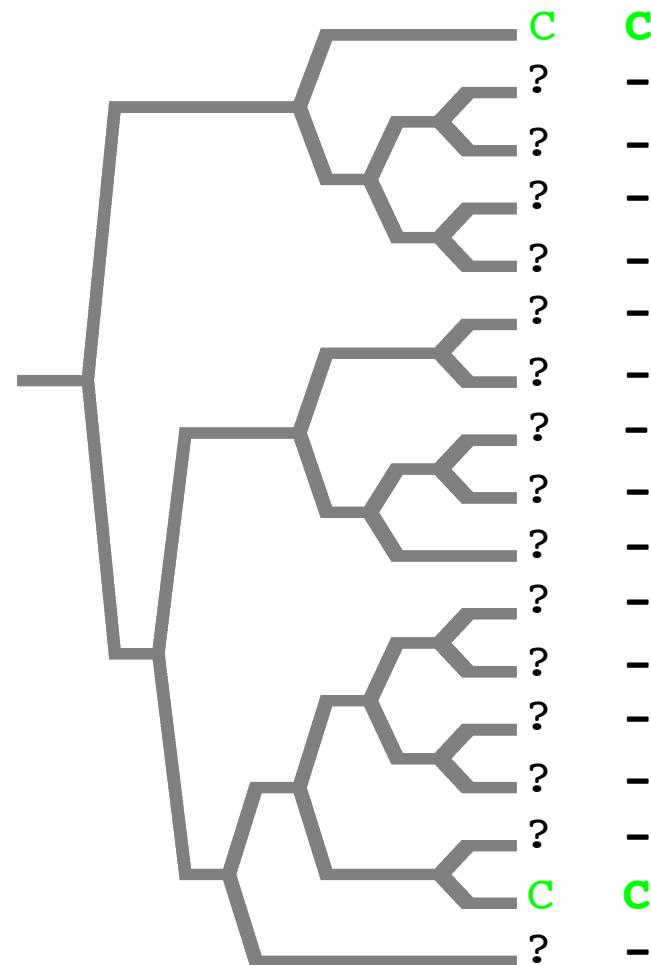
Taxon \ Character																																				
boyliiMVZ148929	C	G	C	-	-	-	T	A	T	A	G	T	A	T	A	-	-	-	C	C	A	C	A	T	T	A	C	C	T	T	-	-	C	A		
luteiventris_MT	C	G	C	-	-	-	T	A	T	-	-	T	A	T	-	T	C	-	T	C	C	A	C	A	T	T	A	C	C	T	T	-	-	C	A	
muscosaMVZ149006	C	G	C	-	-	-	T	A	T	-	-	T	A	T	-	T	C	-	C	C	A	C	A	T	T	A	C	C	T	T	-	-	C	A		
cascadaeMVZ148946	C	G	C	-	-	-	T	A	T	-	-	T	A	T	-	T	C	-	T	C	C	A	C	A	T	T	A	T	C	T	T	-	-	C	A	
temporariaDMH84R1	C	G	T	C	-	-	T	A	T	-	-	T	A	T	-	T	C	-	T	C	C	A	C	A	T	T	A	C	C	T	T	-	-	C	A	
sylvaticaMVZ13742	T	G	C	-	-	-	-	-	-	-	-	T	A	T	A	-	T	C	-	C	C	A	C	A	T	T	A	C	T	T	C	A	T	A		
sylvaticaDMH84R43	T	G	C	-	-	-	-	-	-	-	-	T	A	T	A	-	T	C	-	C	C	A	C	A	T	T	A	C	T	T	C	A	T	A		
septentrionales	T	G	C	-	-	-	-	-	-	-	-	T	A	T	A	-	T	C	-	C	C	A	A	T	T	A	C	C	T	T	-	-	T	A		
catesbeianaDMH84R2	T	G	C	-	-	-	-	-	-	-	-	T	A	T	A	-	T	C	-	C	C	A	A	T	T	A	C	C	T	T	-	-	T	A		
virgatipesMVZ1759	T	G	C	-	-	-	-	-	-	-	-	T	A	T	A	-	T	C	-	C	C	A	A	A	T	A	C	C	T	T	-	-	T	A		
palmipesVenAMNH	C	G	C	-	-	-	T	A	T	-	-	T	A	C	T	-	T	C	-	C	C	A	C	A	T	A	C	C	T	T	-	-	C	A		
palmipesEcuKU20442	C	G	C	-	-	-	T	A	T	-	-	T	A	C	T	-	T	C	-	C	C	A	C	A	T	A	C	C	T	T	-	-	C	A		
Sp_1_ecuadorQCA	C	G	C	-	-	-	T	A	T	-	-	T	A	T	-	T	C	-	C	C	A	C	A	T	T	A	C	C	T	T	-	-	C	A		
bwanaQCAZ13964	C	G	C	-	-	-	T	A	T	-	-	T	A	T	-	T	C	A	C	C	G	C	A	T	T	A	C	C	T	T	-	-	C	A		
julianiTNHC60324	C	G	C	G	C	A	T	A	T	-	-	T	A	T	-	T	C	-	C	C	A	C	A	T	T	A	C	C	T	T	-	-	C	A		
vailantiKU195299	C	G	C	-	-	-	T	A	T	-	-	T	A	T	-	T	C	-	C	C	A	C	A	T	T	A	C	C	T	T	-	-	C	A		
sierramadrensisKU	C	G	T	-	-	-	T	A	T	-	-	T	A	C	A	C	T	C	-	A	C	C	A	C	A	T	T	A	C	C	T	T	-	-	C	A

# Innocuous alignment errors?



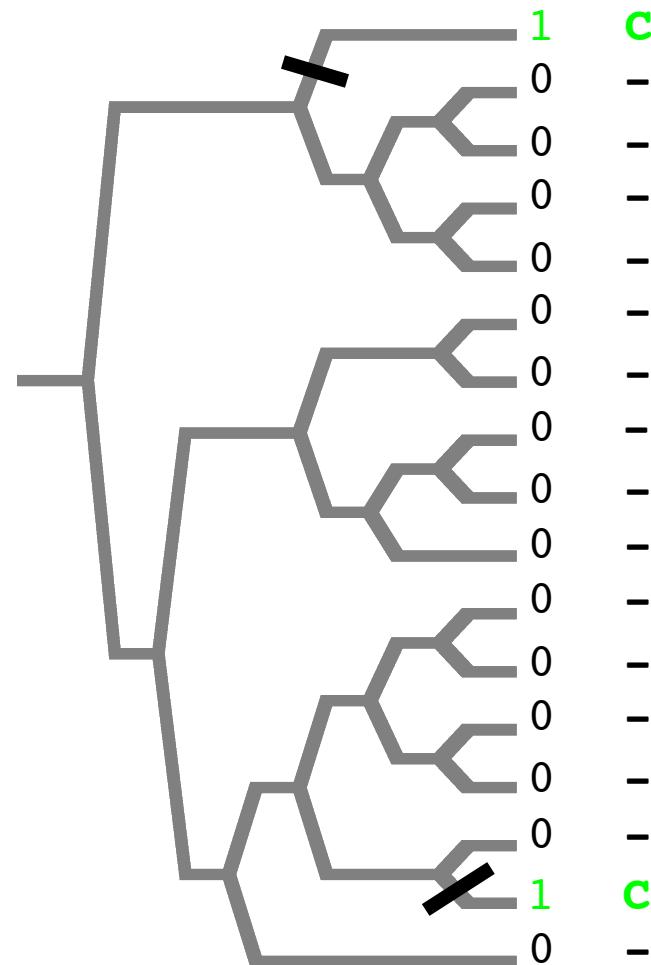
# Innocuous alignment errors?

- GTR – no events, little effect



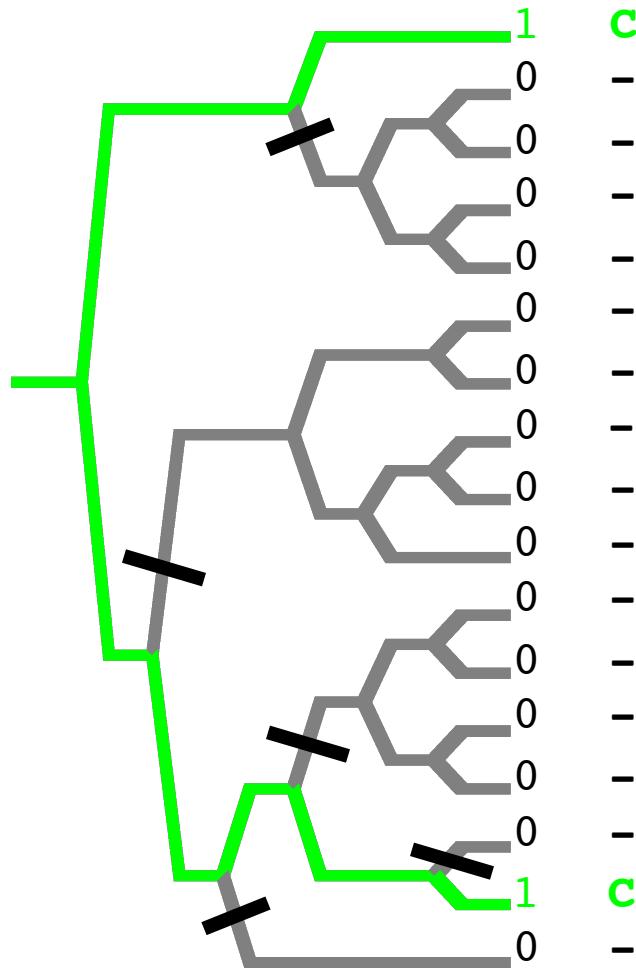
# Innocuous alignment errors?

- GTR – no events, little effect
  - $Mk_v$  – 2 events, moderate effect



# Innocuous alignment errors?

- GTR – no events, little effect
- $Mk_v$  – 2 events, moderate effect
- DIMM – 5 deletions! strong effect  
(Dependent on tree size!!)



# Conclusions: gap models

Indel events can provide useful signal *if* properly modeled

BUT, without correct column homology they can do more evil than good