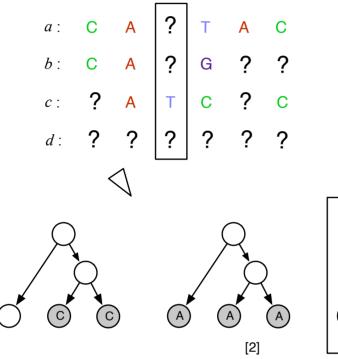
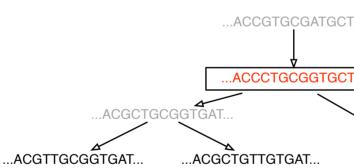


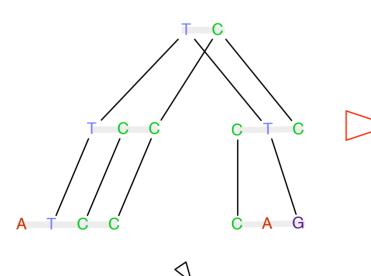
• Approach 1: heuristic^[8]

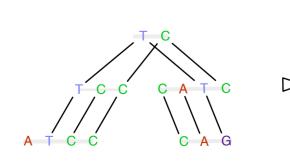


• Approach 2: naive Gibbs sampling ACCGTGCGATGCT. ...ACCCTGCGGTGC .ACGCTGC(ACGCTGTTGTGAT. ..ACGGTGCTGTGC (SSR)



Problem I: random walk behavior

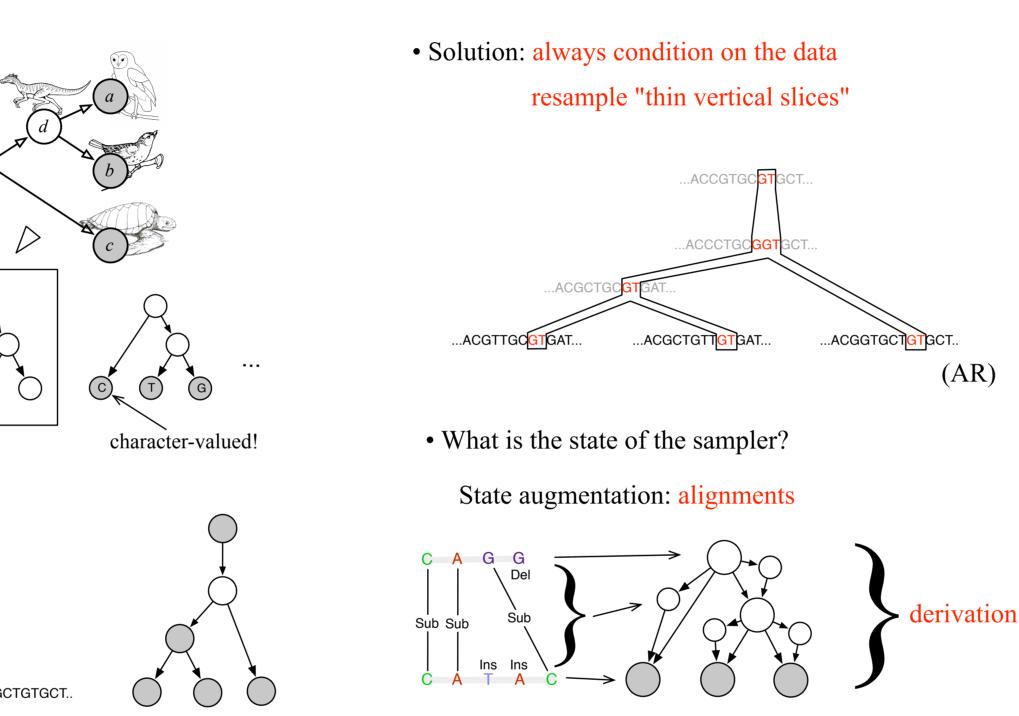




Problem II: each step is expensive (cubic in the sequence length)

Efficient Inference in Phylogenetic InDel Trees

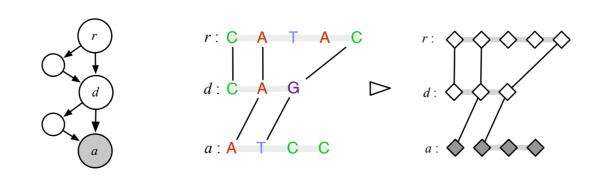
Alexandre Bouchard-Côté Michael I. Jordan Computer Science Division University of California at Berkeley



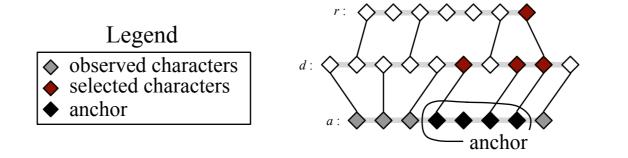
• How should "vertical slices" be defined?

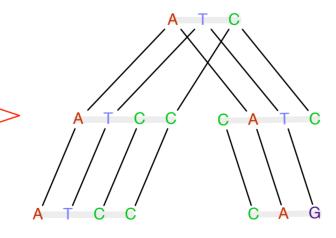
Deterministic functions of a random graphical model

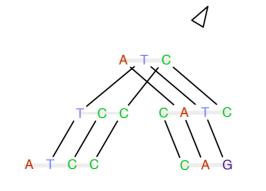
(AR)



Indexed by a span on an observed word (anchor)





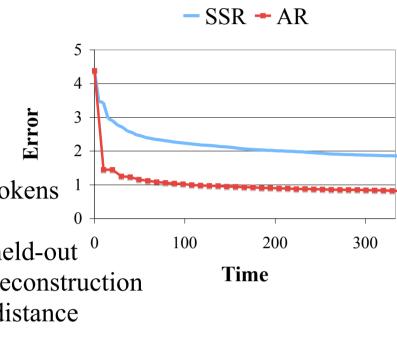


Dan Klein

• Experiments • Connected component of the anchor? 1. Reconstruction _____ $r : [\diamondsuit \diamondsuit \diamondsuit \diamondsuit \diamondsuit \diamondsuit \diamondsuit \diamondsuit \diamondsuit \diamondsuit$ - synthetic • Not irreducible - 4 characters types - 124010 character tokens • Non-contiguous lineages - tree has 7 nodes \diamond \diamond \diamond $(\diamond$ \diamond \diamond \diamond) \langle - all internal nodes held-out - evaluated on root reconstruction • Contiguous version? - loss: Levenshtein distance 2. Multi-alignment $\diamond \diamond \diamond \bullet \bullet \bullet \bullet \bullet \bullet$ - annotated protein d $(\diamond \diamond \diamond \diamond (\diamond \diamond \bullet \bullet))$ - source: BAliBASE \diamond - loss: Sum-of-Pairs Column-Score System CS SSR (Handel) 0.63 $> \diamond \diamond \diamond \diamond \diamond \bullet \bullet$ AR (this paper). 0.77 (b) F1g 1b $\diamond \diamond \diamond (\diamond \diamond \diamond \diamond) \diamond$ • Not reversible The taller the tree, the larger the gap: Figure 1: XXXXX • A correct definition: $r: \diamondsuit \diamondsuit$ System SP CS $r: \ \blacklozenge \ \blacklozenge \ \diamondsuit \ \diamondsuit \ \diamondsuit \ \diamondsuit$ SSR (Handel) 0.63 0.77 $: \diamond \diamond \diamond \dot{\diamond}$ SSR+AR $0.77 | \overline{0.86}$ $: \diamond \diamond \diamond (\diamond \diamond \diamond \diamond) \diamond$ $a: \diamondsuit \diamondsuit \checkmark$ Table 1: XXXXXX—clustalw • Next: how to sample? The multi-alignment system in the litter at that most resembles our lineage sampler is Handel. It models in a probabilistic fashion an InDel evolutionary derivation along a tree ad they a not be solutionary to the solutionary derivation along a tree ad they a not be solutionary to the protect of the solution of t

a multi-alignment as described above. The key difference with our reapproace had societat their inferences a markov is based on SSR rather than the lineage sampling moves that we achyocate "Punethus Star paper."

The BAliBASE multi-alignment dataset [13] is a standard bench product bench in the Built apply Sor purpie of the alignment alignment dataset for the order of the alignment dataset for the order of the alignment alignment alignment database for the order of the alignment alignment programs. Bioinformatics, 15:87-88, 1999. the dataset used to assess the performance of Handel in its original paper 18.3:3-16,19 While other comproved likelihoed of the approved likelihoed of the assess the performance of the and the assess of the performance of the assess of the performance of the and the assess of the performance of the and the assess of the performance of the performance of the assess of the performance of the performance of the assess of the performance of the performan Cylindric proposals: use a DP 1 own to perform better than Handel on this dataset [7 /] they leverage more conhisticated features such as



	gc2_cavpo	1	.ktisktkgaprm <u>PDVYTLPPSRD</u>
	alc_mouse	1	tiakvtvntfp <u>PQVHLLPPPSE</u>
	1yuh	1	.tkltvlgqpkss <u>PSVTLFPPSSE</u>
1 /	1mim	1	gtkleikr.tvaa <u>P</u> SVFIF <u>PPS</u> DE
lata	lnld	1	.ttvtvssastta <u>P</u> SVYPL <u>APVSG</u>
[16]			
/	gc2_cavpo	49	asnrvvsekpeykntppieda
	alc_mouse	49	lhgneelspesylveplkep
(SP)	1yuh	49	kvdgtpvtegmettqpskqs
	1mim	49	kvdnalqsgnsqesvteqdsk
e (CS)	lnld	49	<pre>nsgslssgvhtfpavlqs</pre>
	gc2_cavpo	92	<u>YTCSVMH</u> ealhnhvtqkaisrsp
SP	alc_mouse	95	<u>YSCMVGH</u> ealpmnftqktidrls
	1yuh	91	YSCQVTHeghtve.kslsra.
0.77	lmim	92	YACEVT <u>H</u> qglsspvt.ksfnrg.
0.96	lnld	87	<u>ITCNVAH</u> passtkvdkkieprg.
0.80			

