

Notes on / comparison to Privman et al. alignment filtering paper

September 6, 2011

1 Comparison table

	Privman et al.	This study
Selection method	PAML M7/M8	SLR, PAML M7/M8
ω distr.	4-category discrete estimated from HIV alignment windows	30-category discrete estimated from mammalian alignments
% pos-sel sites	4.6% (gag), 2.9% (pol), 10.5% (env)	6%
Indel distribution	Variable rates estimated from HIV alignment windows (mean rate: 0.045 gag, 0.041 pol, 0.139 env)	Constant rate per simulation, range from 0 to 0.1
ω and indel covariation	Yes - separate sim parameters for each 20-aa window	No - constant indel rate and ω distribution
Sequence count	20	6, 17, 44
Tree shape	Fairly balanced, longest branches at root and tips	Balanced 6-species; mammalian / vertebrate shaped 17 and 44
Divergence level	(MPL values as input to INDELIBLE; divide by 1.5 to get dS) 0.497 gag, 0.492 pol, 0.57 env	Variable; 0.05 - 2.0 root-to-tip dS
Total branch length (at MPL=0.85)	8.5 gag, 9.1 pol, 9.1 env	3.9 17-taxon, 10.1 44-taxon, 3.74 6-taxon
Simulation replicates	50 replicates (x 3 genes) = 150 alignments	100 replicates per condition, 20k alignments
Aligners	ClustalW, MAFFT, MUSCLE, PRANK (amino acid)	ClustalW, MAFFT, PRANK (amino acid), PRANK (codon)
Filters	Column-based GUIDANCE, Gblocks, ALISCORE, optimal; residue-based GUIDANCE	Column-based Gblocks; residue-based GUIDANCE, T-Coffee, optimal

2 Major conclusions from Privman et al.

- Aligners have a strong impact: PRANK works best, then MAFFT/MUSCLE, then ClustalW
- No clear advantage for any of the column-based filtering methods. Residue-based filter showed small improvement (their explanation why: 'reduced loss of information')
- *env*-like simulations yielded worst performance, *pol*-like simulations yielded best performance. Obvious.
- Filtering should be used all the time. From the abstract:

Our study shows that the benefit of removing unreliable regions exceeds the loss of power due to the removal of some of the true positively-selected sites.

From the discussion:

Researchers cannot know the FP rate to expect for their data. It is not under control. Conversely, we show that in all three scenarios positive selection can still be detected after filtering. Therefore, filtering provides a more conservative approach...

3 Criticisms of Privman et al.

- Figure 1 - Why does MAFFT perform better than Prank in the pol simulations? Why does Gblocks often yield the best overall performance (e.g., MAFFT+Gblocks is best in the env simulations, slightly better than Prank+Gblocks and way better than GUIDANCE)

- Figure 2 - Why were MAFFT alignments used instead of Prank, if Prank was discovered to perform the best?
- Figure 3 - again, why use the middle-of-the-road aligner instead of the best one available? What is Prank's performance on the equivalent simulations?
- The 90% FPR number mentioned in their results section is misleading. It comes from Figure 3 env simulations, which used MAFFT instead of Prank and represents a gene-wide FPR (instead of sitewise FPR). We specifically show that Prank does better than other aligners because it keeps the FPR very low, even at high indel rates (see our Supp. Figure S1B for the best comparison).

4 How do their simulations fit into our results?

Although the simulations weren't performed in the exact same way, it seems that the ω distributions are roughly comparable (though pol has lower, and env has higher, % pos-sel sites), and gag/pol are around 0.05 mean indel rate, while env is around 0.15 mean indel rate. Divergence is around 0.3-0.5 dS (there is some uncertainty regarding the exact meaning of their branch lengths). So the divergence is at the low end of our range, and the indel rates are at the medium-to-high end.

Using Privman et al. Figure S3 (ClustalW/gag) we see TPR of 0.08 at a FPR of 1%. Our Figure 4c (ClustalW/17-taxon) shows a 0.3-0.4 TPR at FPR of 1%, MPL=0.5 and indel=0.05. Our numbers seem to trend higher than Privman et al. for the 'equivalent' simulation.

In terms of the benefit of filtering, we can try to compare our results to their gag simulations (which had a ω distribution most similar to ours). With ClustalW alignments and Gblocks filtering they find a 30% improvement in TPR at 5% FPR, while we saw no change (MPL=0.4, indel=0.08). For GUIDANCE filtering they find 100% improvement, while we saw 10% improvement. With Prank alignments, they saw 10-20% improvement for all methods. We saw no change for all methods.

Some of this difference could be that they used amino-acid Prank while we used codon Prank. Also, the difference between 1% and 5% FPR for the filtering measure could impact things.

More generally, it is plausible that the clustering of regions with high indel rates and positive selection is causing their alignments to be more 'difficult' overall, thus putting their equivalent position in our parameter space further towards the upper-right. Still, even with this explanation a number of inconsistencies remain (see Criticisms).

5 What to emphasize about our study versus Privman et al. / Fletcher and Yang

- We separated false positive and false negative rates using a constant SLR threshold – this allowed to separately analyze the sources of error, yielding the insight that all aligners suffer from false negatives, and Prank suffers from fewer false positives.
- We identified consistent trends across a wide range of indel rates / divergence levels
- We measured the impact of filtering when applied to the worst and best aligners, instead of using a middle-performing aligner that should not be used in practice by anyone worried about misalignment.

6 To-do

- Update introduction with some mention of Privman et al. and why our study differs from theirs
- Add some comparison to Results / Conclusion to Privman et al.'s results
- Subtle changes to manuscript (title / abstract / conclusion) emphasizing the unique / novel aspects of this study