NG86 limitations: ignoring phylogenies

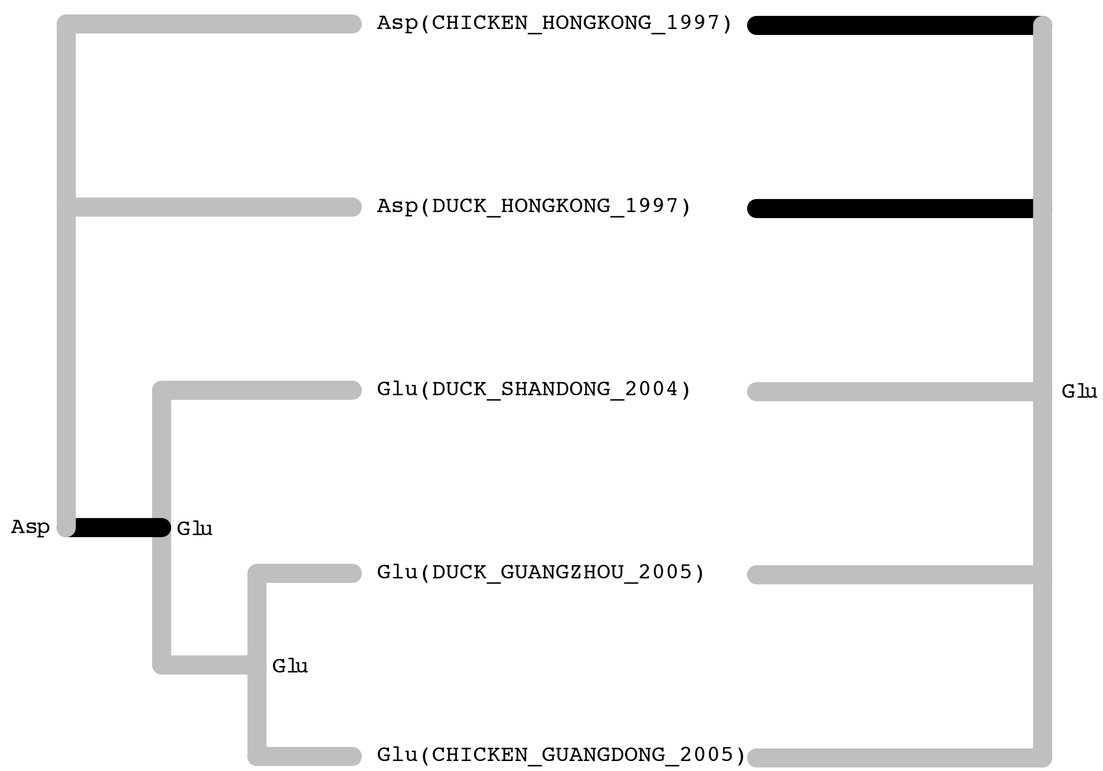

Fig. 1.1. Effect of phylogeny on estimating synonymous and nonsynonymous substitution counts in a dataset of Influenza A/H5N1 haemagglutinin sequences. Using the maximum likelihood tree on the left, the observed variation can be parsimoniously explained with one nonsynonymous substitution along the darker branch, whereas the star tree on the right involves at least two.

http://www.hyphy.org/resources/hyphybook2007.pdf

NG86 limitations: averaging across all sites in a gene

- Different sites in a gene will be subject to different selective forces.
- A *gene-wide* measure of selection is going to average these effects.
- **Most** sites in **most** genes will be maintained by purifying selection.
- Positively selected sites are of great biological interest, because they point to how a particular gene can respond to selective pressures.
- Negatively selected sites are also of interest, because they point to functional constraint, and could be used to guide drug or vaccine design.
- Must develop methods that are able to disentangle the contributions of individual sites.