



**Supplementary Figure 1**

**Genes Retained Per Strain at Incrementing PID Thresholds.**

Genes Retained Per Strain at Incrementing PID Thresholds. The number of genes retained in each strain-specific model is dependent on the threshold utilized for binarization of the homology matrix. The effect of the threshold will also be dependent on how closely related the target strains are to the reference strain. For example, within the strains in the Supplementary Tutorial notebooks we see that CU651637.1 and CP002167.1 are more dissimilar to reference model iML1515 as the drop off in retained genes occurs in a steeper fashion. We suggest using a threshold of 80 when comparing strains of the same species to ensure a sufficient similarity metric to include a gene in the draft models.