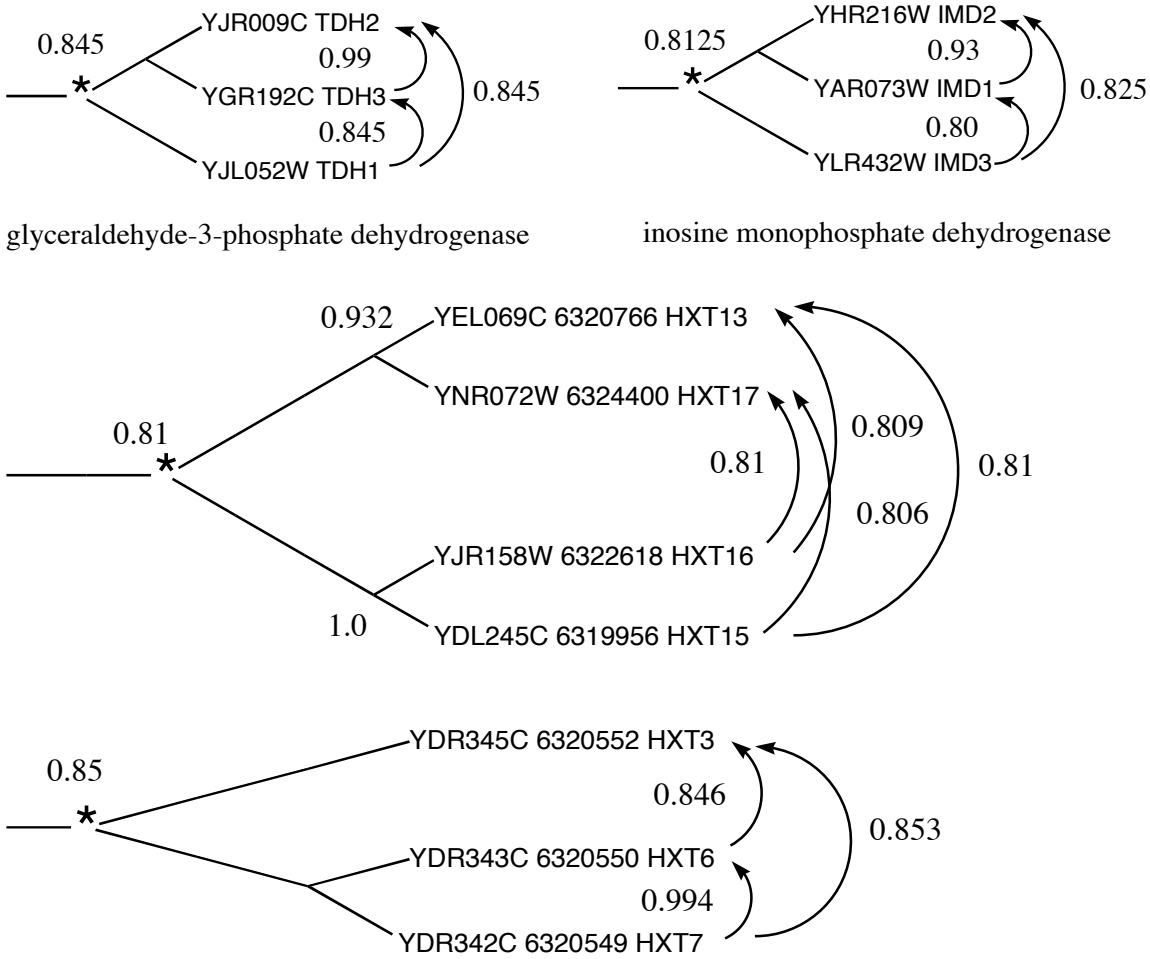
Supplementary Fig. 2



Trees showing f_2 values for families with more than two paralogs. Nodes marked with an * are those representing duplications where $0.80 < f_2 < 0.86$.

Because some families have suffered more than one duplication, it is possible to have more paralog pairs then there are duplications. Specifically, when a family contains duplications following the duplication within the window, this will lead to more paralogs pairs than duplications having $0.80 < f_2 < 0.86$. That is, some of the duplications in the window will be counted more than once.

Specifically, the f_2 values for TDH1-TDH3 and TDH1-TDH2 are two estimates of the same f_2 value for duplication between TDH1 and the ancestor of TDH2/3. The TDH2-TDH3 pair, arising from a later duplication has a higher value of f_2 (0.99). This is illustrated in the figure below.

Similarly, the f_2 values for IMD1-IMD3 and IMD2-IMD3 are two estimates of the same f_2

value for duplication between IMD3 (YLR432W) and the common ancestor of IMD1 (YAL073W) and IMD2 (YHR216W). The IMD1- IMD2 pair, arising from a later duplication, has a higher value of f_2 (0.93). This is illustrated in the figure below.

The same analysis applies to the sugar transporters genes HXT13 (YEL069C), HXT15 (YDL245C), HXT16 (YJR158W), and HXT17 (YNR072W), generating two pairs in the table with $0.80 < f_2 < 0.86$ (but only one duplication in the $0.80 < f_2 < 0.86$ window) and the HXT3 (YDR345C), HXT6(YDR343C) and HXT7 (YDR342C), generating two pairs in the table (but only one duplication in the $0.80 < f_2 < 0.86$ window). This is illustrated in the figure below.