



**Figure S1:** Heat map indicating information content (IC) inferred with MARE when analysing the amino acid dataset (sm-aa). Species are shown in rows and genes are shown in columns. The IC is colour-coded as follows: Genes which are not present are white coloured and genes with a tree-likeness = 0 are red coloured. The darker the blue, the higher is the tree-likeness. The information content of the amino acid dataset calculated with MARE is 0.65 and matrix saturation is 0.83.