



Figure 5.6: Sequence diversity tree of cross-validation dataset D215DSC and the external validation set D302TSA. The first colourful band around the trees indicates different families. The second band shows which sequence belongs to which dataset. Finally, the last band is a color bar regarding the T_m . Low-temperature, stable enzymes are represented with green and high temperature ones with red. It is expected to see sequences from the same families close together regardless from which dataset they belong to. Figure is generated using xTol, an in-house sequence diversity tree illustrator.