

FIG. W2. The amino acid hierarchical scoring schema adapted from pattern-induced multi-sequence alignment (PIMA). In PIMA, each leaf represents an amino acid, each ancestral node represents an amino acid class, and the number in the cardinality column denotes the level of each node. The similarity score of amino acids x and y is calculated as six minus the level of the amino acids' most recent ancestral node.

