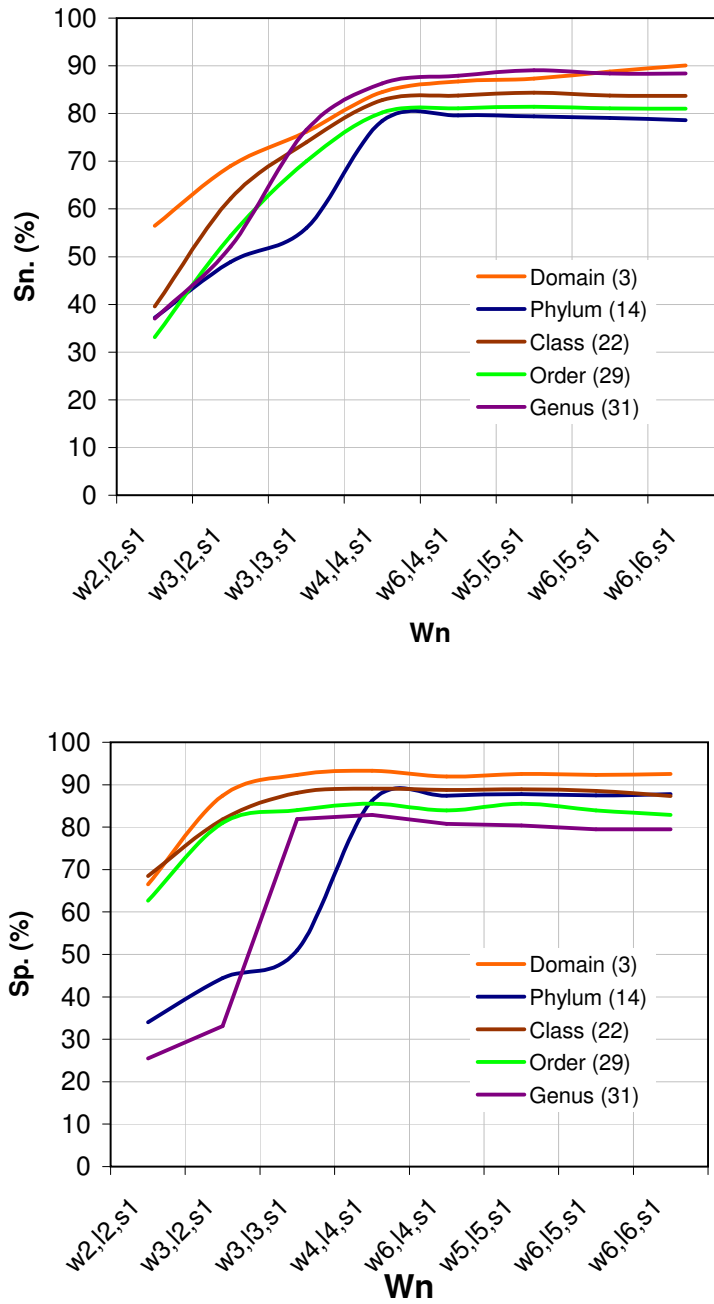


Supplementary Figure 2



**Supplementary Figure 2.**  $W_n$  parameter search for the sequence composition space with the highest classification accuracy for 15 kb fragments of unknown organisms at different phylogenetic levels. The legend gives in brackets the number of clades for each of the analyzed levels. The composition space is defined by the word length  $w$ , the number of literal characters  $l$ , and the step size  $s$ . The sensitivity (upper panel) and specificity (lower panel) is shown, that is attainable with 15 kb genomic fragments from unknown organisms.