



**Figure 1: Identity percentage of the 100 best hits per sequence after BLASTn against the ncbi database.** The results are provided for all sequences together, and for each major group independently. Means and medians are indicated as orange and black dots, respectively. Colors of the bars correspond to 100%, 95%, 80% and 50% confidence intervals, from light to solid. Lineage-specific vs. all sequence distribution differences are indicated as asterisks for the t-test and the Kolmogorov-Smirnov test.