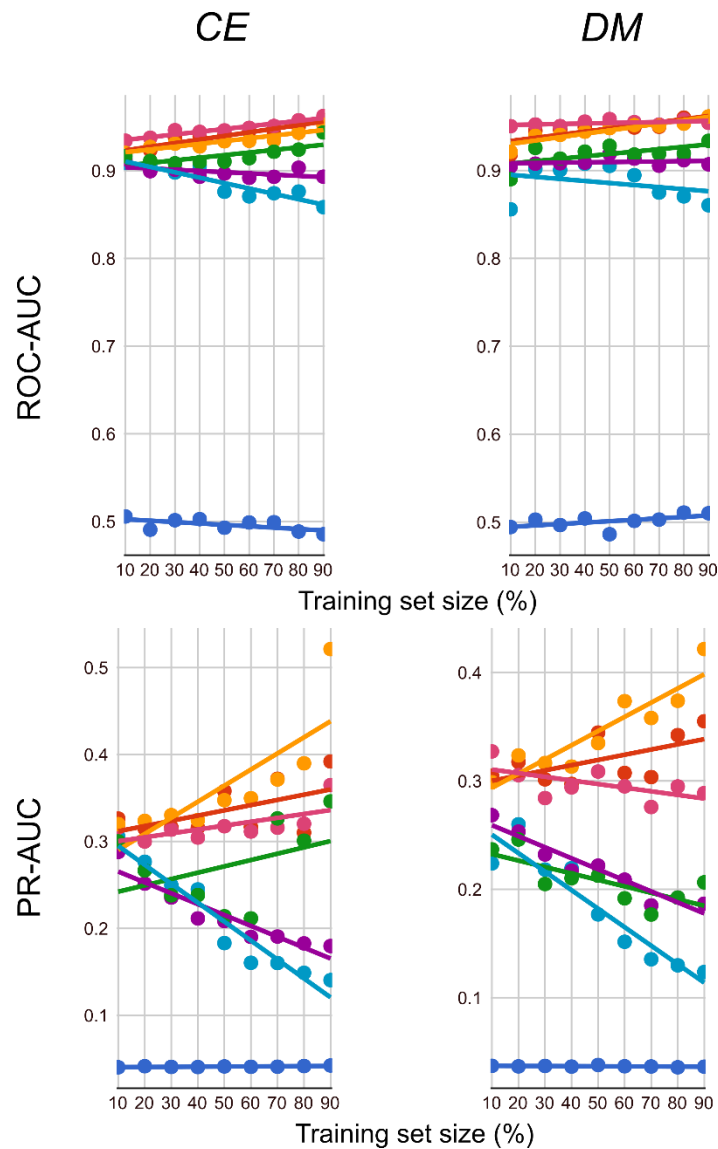**Figure S1.** Systematic evaluation of essential gene predictions within species using a non-redundant set. High-confidence essential and non-essential genes [22,23] (Campos et al., 2020a,b) were obtained, genes with protein sequence identity > 25% were clustered, and only the genes representing centroids were retained. Then, for each species, subsets (10-90%, 10% increments) of non-redundant essential and non-essential genes were randomly selected (x-axis) to train of six machine learning approaches/methods, following a feature selection strategy. Prediction performance (ROC-AUC and PR-AUC) was evaluated using the remaining data.