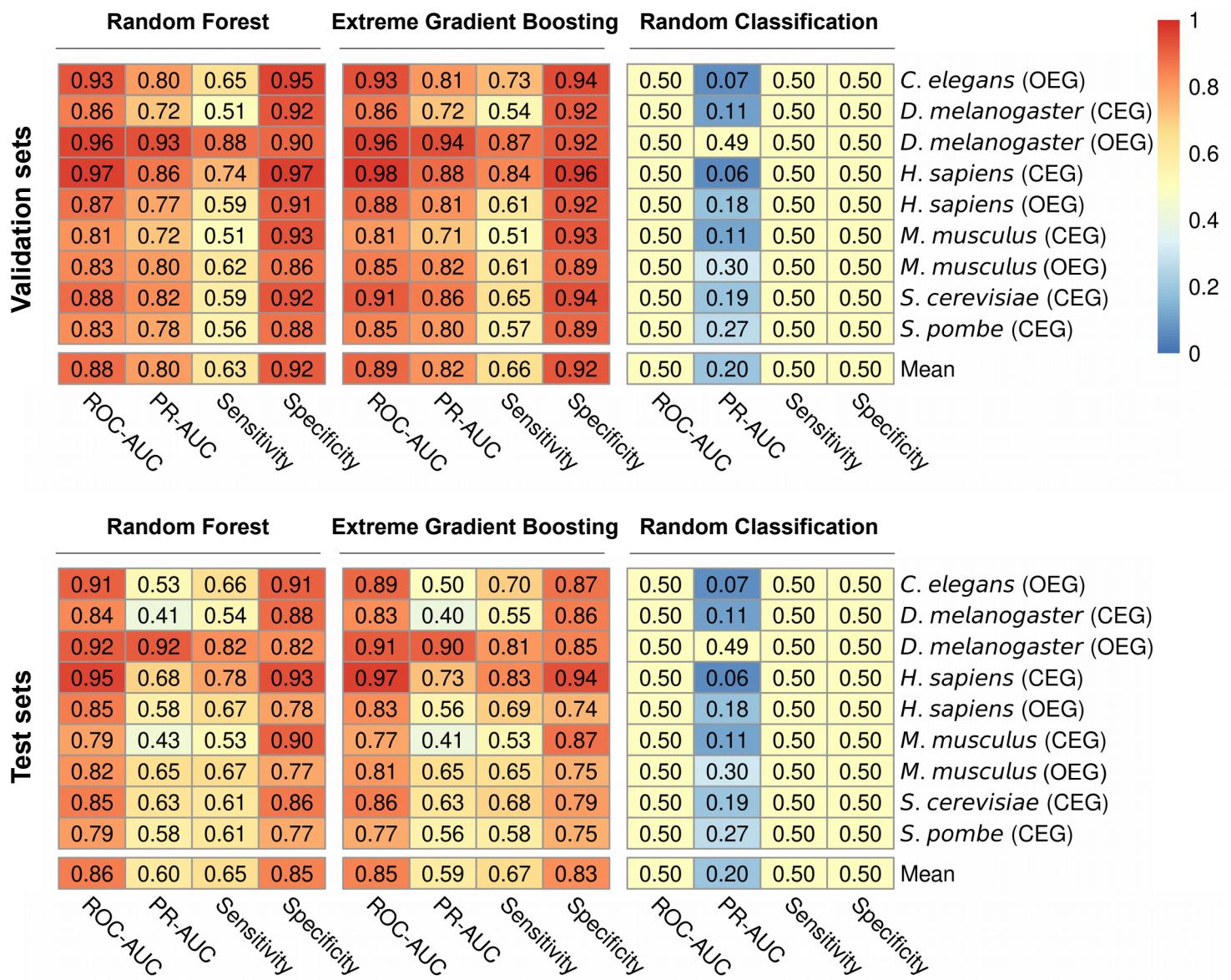
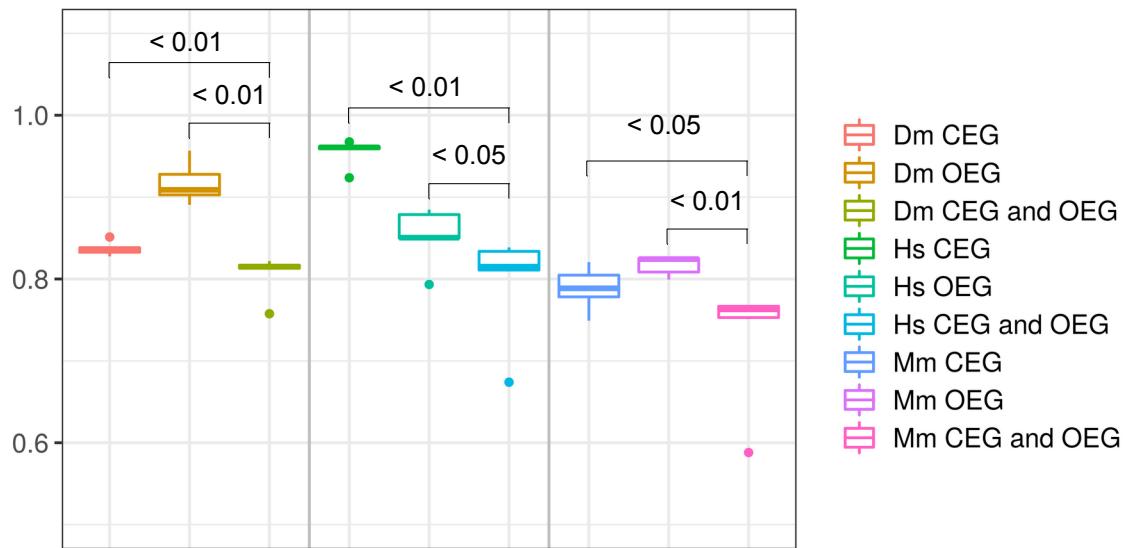


Supplementary Figure 1. Schematic overview of the machine learning workflow. A two step feature reduction was performed using Least Absolute Shrinkage and Selection Operator (LASSO) and highly correlating features were removed. Performance of the classifier was evaluated on the validation and testing sets. The validation set is a fraction of the training set and was used for hyperparameter tuning and performance evaluation in each iteration. The testing data was not used for training the algorithm, but used to evaluate the performance of the fully specified classifier. To improve generalizability, we performed a stratified randomized 5-fold cross validation, resulting in five trained machines for each data set.

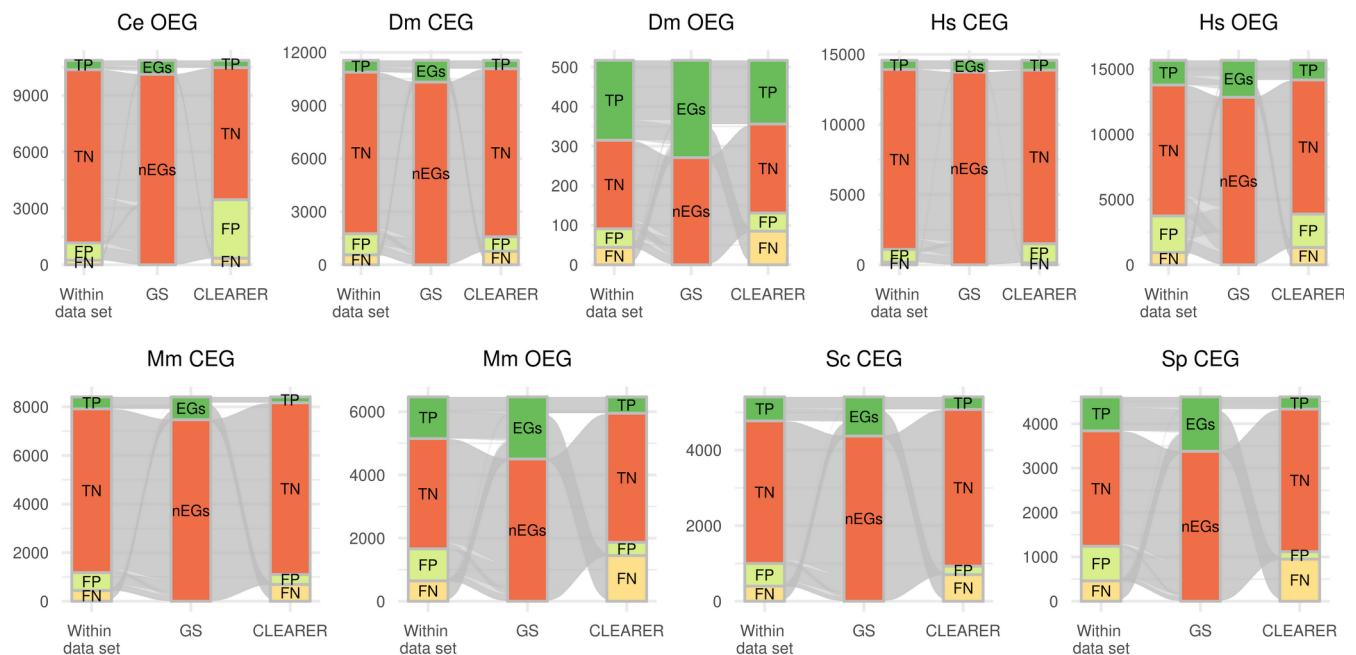


Supplementary Figure 2. Machine learning performance of classifiers applied to validation and test sets. Heatmaps showing the performance of the two used machine learning algorithms i.e. Random Forest and Extreme Gradient Boosting on the nine essential gene data sets. The performance metrics ROC-AUC, PR-AUC, sensitivity and specificity are shown. The ROC-AUC is the area under the curve (AUC) of the true positive against false positive rate. Consequently, it represents the measure of separability of essential and non-essential genes. The PR-AUC is the AUC of the precision and recall (= sensitivity) curve, which shows the fraction of true positives out of all positively predicted *versus* the fraction of true positives out of all essential genes. Sensitivity measures the proportion of essential genes that are correctly identified (when at least half of all trees predict a gene to be essential) and specificity the proportion of correctly identified non-essential genes. A random classifier would have a sensitivity and specificity of 0.5. Performance and standard deviation didn't decrease substantially when comparing the results from training and testing, reflecting that the classifiers didn't overfit.

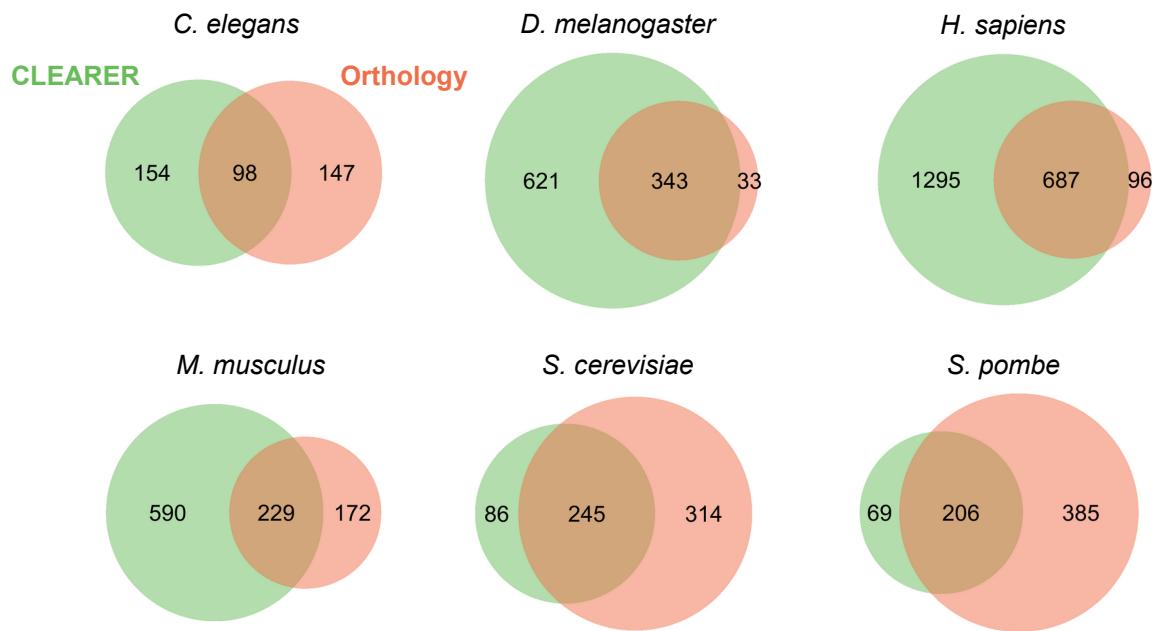
ROC-AUC



Supplementary Figure 3. Box plots showing the ROC-AUC values for the Random Forest testing sets. A significant improvement (Wilcoxon tests) can be seen when cellular essential genes (CEG) and organismal essential genes (OEG) are handled separately. Organisms for which CEG and OEG data were available are *D. melanogaster* (Dm), *H. sapiens* (Hs) and *M. musculus* (Mm).



Supplementary Figure 4. Sankey diagrams show the correspondence between the gold standard (GS) and predictions of classifiers trained and tested on the same data set (within data set) and the ensemble classifier (CLEARER). Essential (EGs) and non-essential genes (nEGs) from the gold standard are overlaid with the predictions. The sum of true positive (TP) and true negative (TN) predictions i.e. correctly identified EGs and nEGs are similar, demonstrating the similar accuracy of both approaches. However, the proportions of false positives (FP) and false negatives (FN) changed.



Supplementary Figure 5. Comparison of CLEARER with the orthology-based approach. Venn diagrams show the true positives i.e. essential genes correctly classified to be essential by CLEARER (green) and the orthology-based approach (orange).