SUPPLEMENTARY FIGURES:

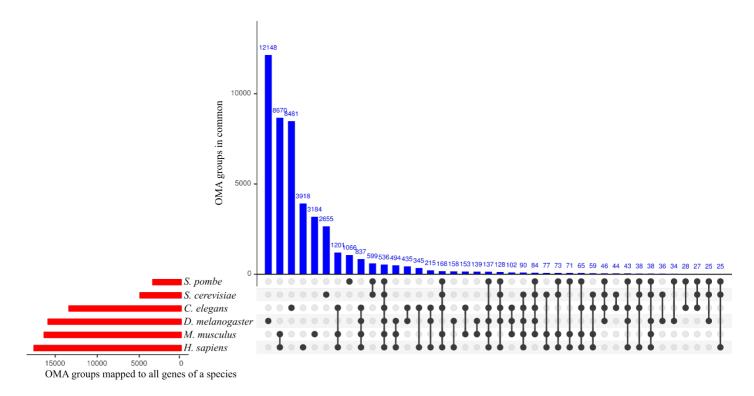


Fig. S1. Diagram exhibiting the total (red) and shared (blue) ortholog identifiers (complete predicted proteomes) for the species used in the present study, obtained from the OMA database.

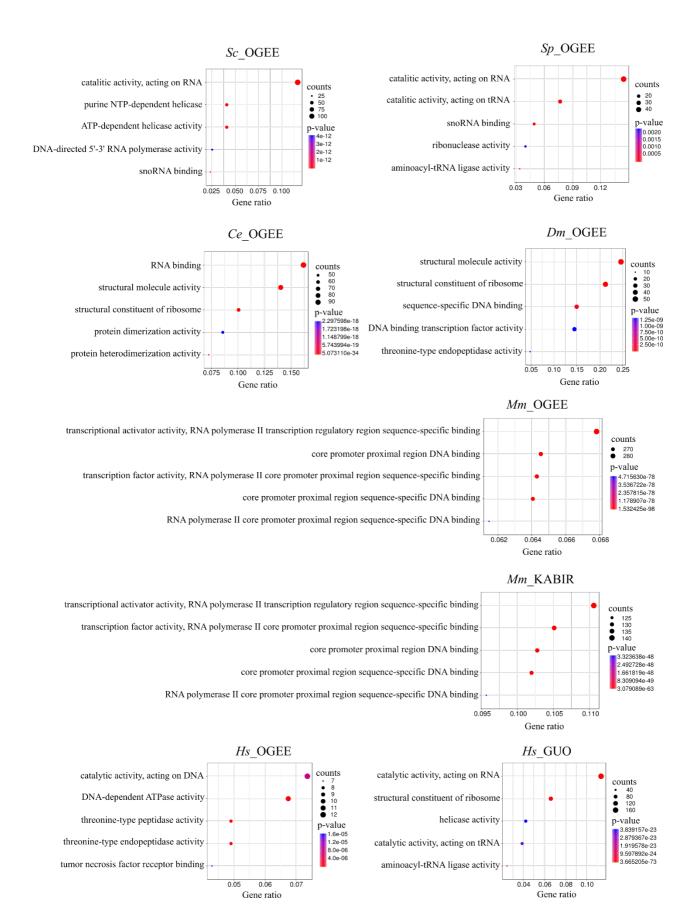


Fig. S2. Molecular function enrichment analysis ("enrichGO" function, "clusterProfiler" package in R) of gene ontology terms of essential genes per data set.

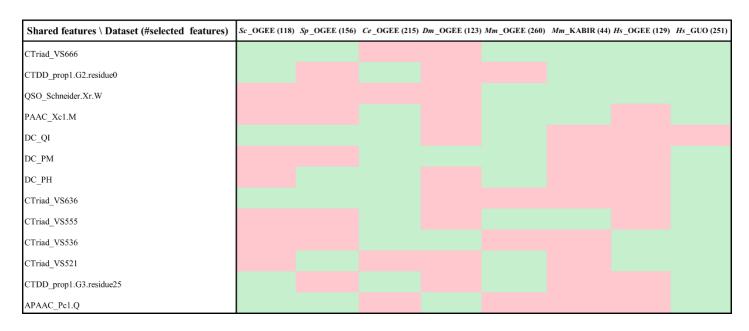


Fig. S3. List of shared features selected by both ElasticNet and Ensemble SPLS methods among at least 4 data sets. The total number of features selected for each species/data set is reported.

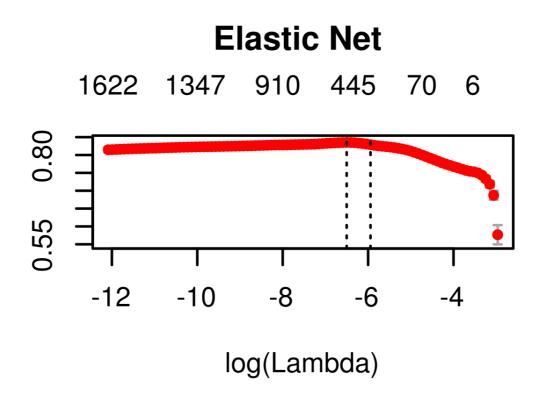


Fig. S4. An example of the application of ElasticNet as a feature-selection procedure using *C. elegans* data. The x-axis represents the lambda parameter tested (bottom) and the corresponding retained features (top). The y-axis represents the ROC-AUC using 10-fold cross-validation. The final set of selected features maximises this metric.

OTHER SUPPLEMENTARY FILES:

- -Table S1. Selected features of each data set and importance of feature
- R "Sessioninfo" file
- Fasta files (data sets) and R scripts (online https://bitbucket.org/tuliocampos/essential and a static version referring to this publication https://doi.org/10.6084/m9.figshare.8063069)