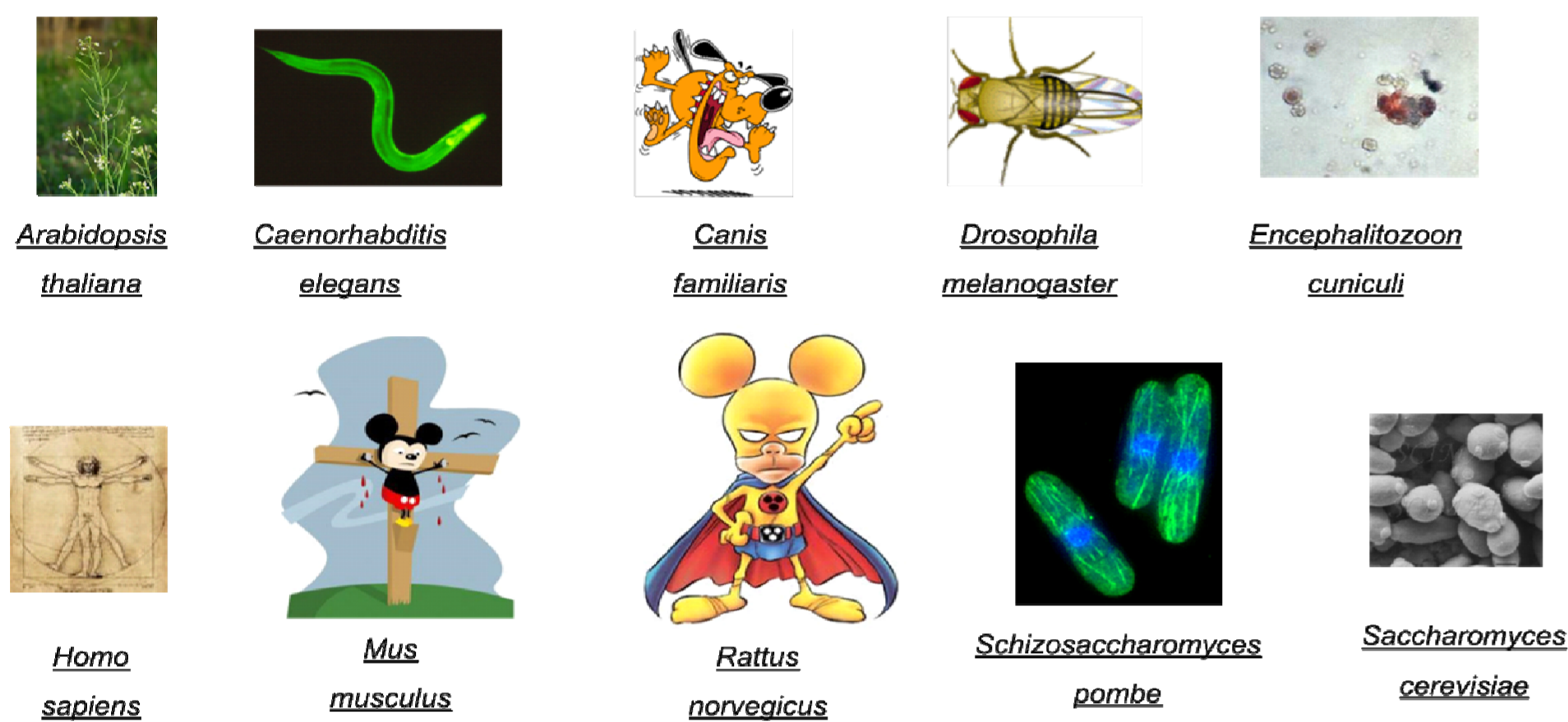


INTRODUCTION

ORTHOLOGS ARE PROTEINS WHICH COME FROM SPECIATION EVENTS, AS CONTRASTED WITH PARALOGS, WHICH DERIVE FROM DUPLICATION EVENTS. DETECTION OF ORTHOLOGS IS BECOMING MUCH MORE IMPORTANT AS THE RAPID PROGRESS IN GENOME SEQUENCING. AN ORTHOLOGS DATABASE IS VERY HELPFUL TO IDENTIFY NEW SEQUENCES AND PREDICT THEIR FUNCTION. IN THIS WORK WE EVALUATED THE ORTHOMCL-DB AS SOURCE FOR

METHODS

WE CONSTRUCTED A PROTEIN DATABASE USING ALL ORTHOMCL-DB ENTRIES FOR TEN PROMINENT ORGANISMS, TOTALIZING 270327 PROTEINS CLUSTERED UNDER 50785 DISTINCT ENTRIES.



THE ESTS WERE DOWNLOADED FROM THE DBEST DATABASE, AND TOTALIZED 360883, 302080, 375360 AND 365619 ESTS FROM ATH, CEL, DME AND HSA, RESPECTIVELY. UNIQUES TO BE ANALYZED WERE OBTAINED BY RANDOMLY SELECTING 150 THOUSAND ESTS THAT WERE CLUSTERED AND ASSEMBLED WITH TGICL SOFTWARE.

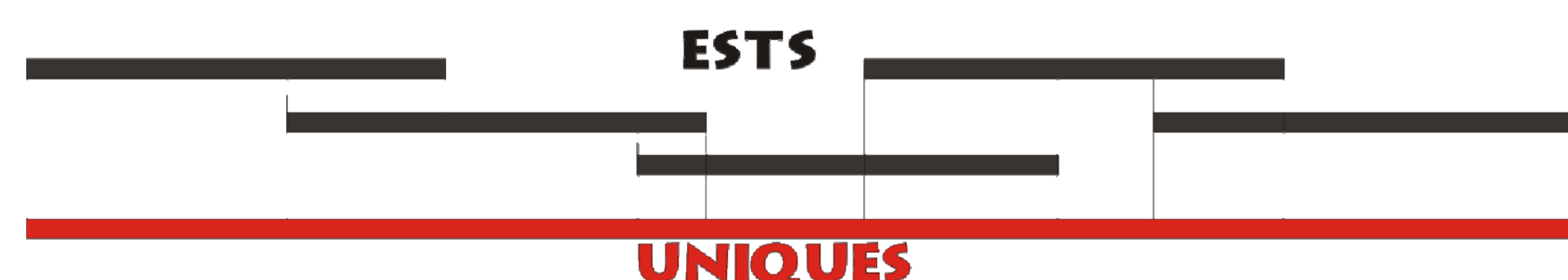


FIGURE 1 REPRESENTATION OF THE TGICL MECHANISM WHICH CREATES UNIQUES BASED ON THE ALIGNMENT AND OVERLAP REGIONS OF THE ESTS.

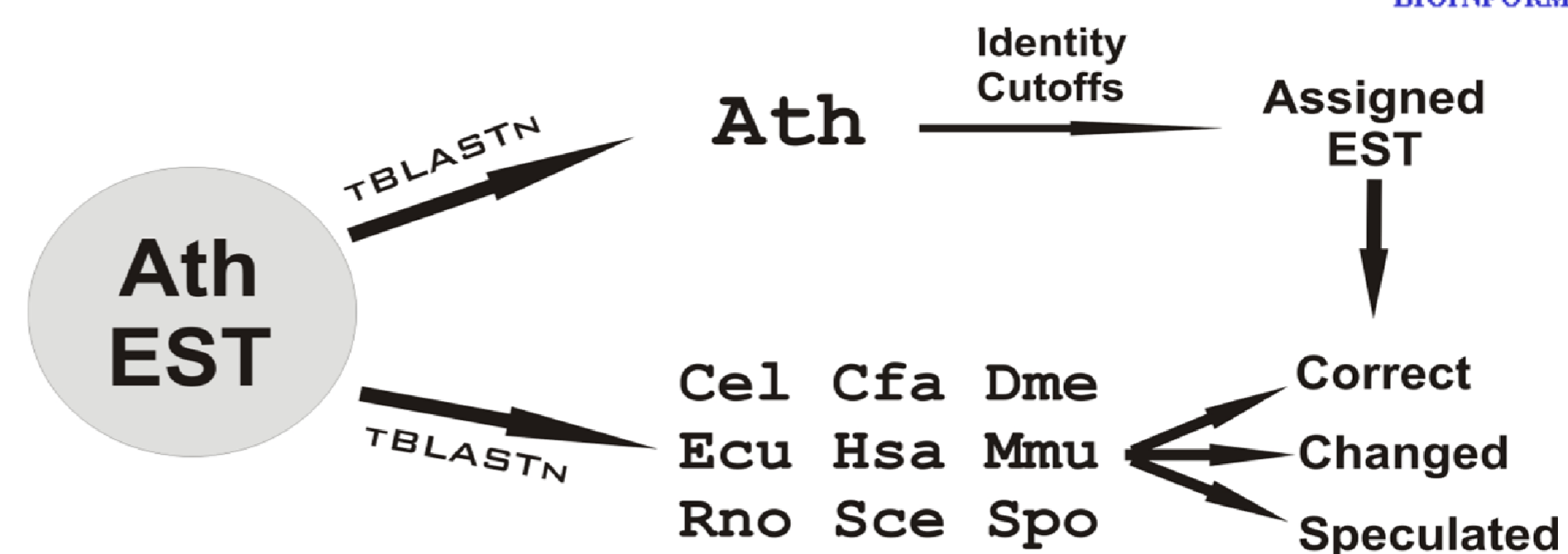
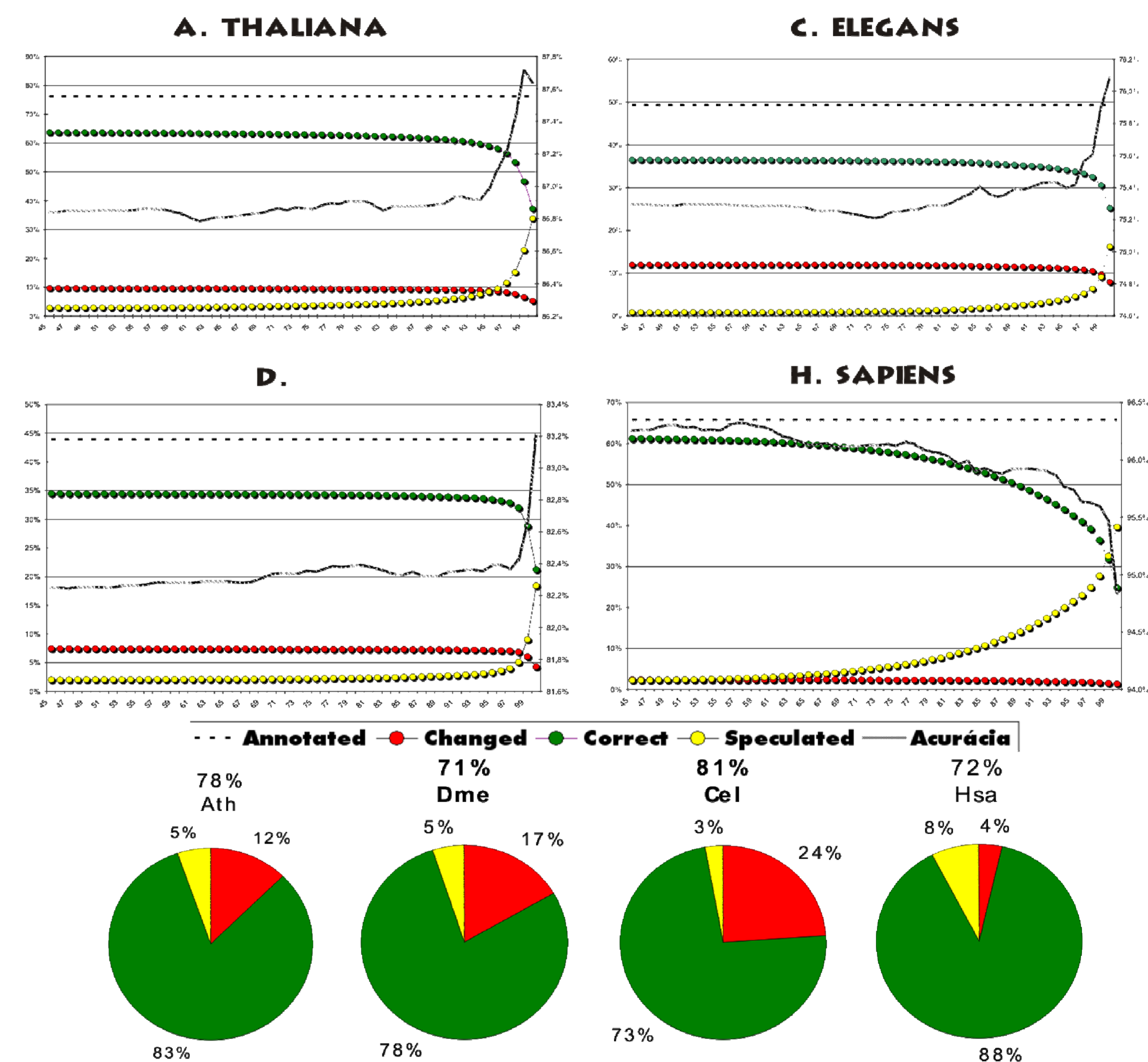


FIGURE 2 REPRESENTATION OF THE BLAST EXPERIMENT USING AS EXAMPLE THE A. THALIANA'S EST. THE FIRST STEP WORKS AS A POSITIVE CONTROL, AND THE SECOND STEP SIMULATES A NEW TRANSCRIPTOME. COMPARING THE RESULTS WE ESTABLISHED LABELS FOR THE ANNOTATION: CORRECT, CHANGED AND SPECULATED.

RESULTS



DISCUSSION

THE ORTHOMCL-DB IS A GOOD SOURCE FOR AUTOMATED ANNOTATION. IN SOME CASES, THE CHANGED ANNOTATION PROPORTION IS HIGH, IT HAPPENS BECAUSE THE CHANGES OCCURS IN CLUSTERS WITH THE SAME DEFINITION, AND DOES NOT PREJUDICE THE FUNCTION ANNOTATION.