Protein Prediction 2

Exercise
Team 6
10.12.2013

Automated Testing

1. Picking 1000 random sequences from annotated sequences

2. Query blast for top 5 hits

3. Create tree for each hit

4. Merge trees using one of the defined merging algorithm

5. Apply thresholds from 0 to 1.0 with a step of 0.1

6. Calculate precision, recall and f-max with each threshold

Final Merging Method

- Top 5 blast hits

- Using blast "score"

- Score is added for a term seen more than once

- Score normalized

- Threshold of 0.3 applied

- Testing on 1000 test sequences, an f-max of 0.36