BIOINFORMATICS

Vol. 00 no. 00 2014
Pages 1–2

Automated HPO-Annotations for newly sequenced proteins by homology inference

K. Nagaraj[1], M. Hanumanthappa[1], O. Tarabai[1], S. Seitz[1]

[1]Fakulät für Informatik, Boltzmannstr. 3, 85748 Garching

Received on 28.02.2014

Associate Editor: I12 - Department for Bioinformatics and Computational Biology - Fakultät für Informatik, Boltzmannstr. 3, 85748 Garching

ABSTRACT

Motivation: Rapid genome sequencing and high-throughput technology, automatic function prediction for a novel sequence is an essential matter in bioinformatics. Automatic annotations based on local alignments suffer from several drawbacks (2). With our de novo method we try to improve the precision and recall of automatic annotations.

Results: Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text

Availability: The webinterface for our created prediction-method is available at https://dataminer.informatik.tu-muenchen.de/ omar.tarabai/.

Contact: name@bio.com

1 INTRODUCTION

In the databases many proteins are found for which the sequence is known, but the function is still not determined. With the increasing number of sequences, caused for example by genomic scale projects, traditional experimental approaches have become outpaced. This leads to the need for rapid and reliable functional annotation methods (1).

Many different approaches have been taken annotate protein function by computational methods, including methods based on sequence, expression, interaction and tertiary structure. Despite this taken effort and the following increase of number and variety of prediction methods, automated annotation remains difficult. Reasons for these difficulties can for example be found in the inherent limitations of current tools and databases or the ambiguity of the definition of function itself (1).

To overcome this problems and to be able to annotate protein function without relying on tertiary data, this method is created to reliably predict protein function by sequence alone.

2 MATERIAL & METHODS

Our method mainly relies on using protein sequence similarity as an indicative of functional similarity in order to transfer Human Phenotype Ontology (HPO) annotations from known to unknown protein sequences. Therefore, in order to achieve a reasonable prediction, we needed a set of properly annotated and reviewed protein sequences to use as a reference database. For this we used...

We used BLAST 2.2.26 (BLAST) which is a widely-used tool for protein sequence alignment, it takes as input a pre-generated database of reference protein sequences and a target sequence. It employs a heuristic algorithm to search the reference database for sequences that are most similar to the target sequence. Its main output is the top N hits sorted from the most similar to the least similar protein sequence. Additionally, it outputs a number of other values defining statistics about the degree of similarity discovered, most relevant to our method is the "bit score" value which is a statistical measure of how good the alignment is (BLAST score).

3 RESULTS

4 DISCUSSION

Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text. Figure **??** shows that the above method Text Text Text Text Text Text Text Text Text Text Text Text. **?** might want to know about text text text text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text. Figure **??** shows that the above method Text Text Text Text Text Text Text Text Text Text Text Text. **?** might want to know about text text text text Text Text Text Text Text Text Text Text Text.

Table **??** shows that Text Text Text Text Text Text Text Text Text Text Text. Figure **??** shows that the above method Text Text. Text Text Text Text Text Text Text Text Text. Figure **??** shows that the above method Text Text. Text Text Text Text Text Text Text Text Text. Figure **??** shows that the above method Text Text.

ACKNOWLEDGEMENT

Without the great help and guidance by the Rostlab, and every groupmember there, we wouldn't have been able to succeed in creating our method. Also we'd like to thank the Rostlab for letting us access their computers and equipment.

REFERENCES

[BLAST]BLAST: Basic Local Alignment Search Tool. http://blast.ncbi.nlm.nih.gov/Blast.cgi

[BLAST score]BLAST Scores and Statistics. http://www.ncbi.nlm.nih.gov/books/NBK21097/#_A614_

[1]Rodrigues, Ana PC and Grand, Barry J and Godzik, Adam and Friedberg, Iddo (2007) The 2006 Automated Function Prediction Meeting, *BMC Bioinformatics*, **8**, S1.

[2]Sasson, Ori and Kaplan, Noam and Linial, Michal (2006) Functional annotation prediction: All for one and one for all, *Protein Science*, **15**, 1557-1562.