



Universidade de São Paulo

Biblioteca Digital da Produção Intelectual - BDPI

Departamento de Computação e Matemática - FFCLRP/595

Artigos e Materiais de Revistas Científicas - FFCLRP/595

2015

Sifter-T: A scalable framework for phylogenomic probabilistic protein domain functional annotation

BMC Bioinformatics. 2015 Apr 30;16(Suppl 8):A4

<http://www.producao.usp.br/handle/BDPI/49964>

Downloaded from: Biblioteca Digital da Produção Intelectual - BDPI, Universidade de São Paulo

MEETING ABSTRACT

Open Access

Sifter-T: A scalable framework for phylogenomic probabilistic protein domain functional annotation

Danillo C Almeida-E-Silva, Ricardo Z N Vêncio*

From Latin American Student Council Symposium 2014 (LA-SCS 2014)
Belo Horizonte, Brazil. 27 October 2014

Background

In the functional annotation field, Sifter v2.0 is regarded as one of the best when it comes to annotation quality. Recently, it has been considered one of the best tools for functional annotation according to the initiative “Critical Assessment of Protein Function Annotation” (CAFA), an open collaborative experiment designed for large-scale assessment of protein function prediction tools. Sifter combines two powerful ideas: phylogenomics and bayesian graphical models. Nevertheless, it is still not widely used. This contradictory observation is probably due to issues with usability and suitability of the framework to a high throughput scale.

Although powerful in terms of approach, it can be considered prototype level in terms of software. The current Sifter version does not allow nucleotide or amino acid sequences input directly, nor accepts current standards in gene annotation formats. Moreover, several parameters are still hardcoded and difficult to be tuned by the end user. Finally, its relationship to third party dependence software is cumbersome, along with its output.

Description

In this study, we had two goals: (i) enhance the tool’s usability, through local implementations or a web-based front end; and (ii) optimize the original source-code for better performance, allowing it to be used in genome-wide scale.

Among the implemented strategies we have: parallel threads; CPU load balancing; best use of disk access, memory usage and runtime; adaptation to the currently used biological databases formats; improved user accessibility; expansion of accepted input types; automation of the reconciliation process; new output format; detailed documentation; and other minor implementations.

The increased performance allowed, for example, the reannotation of 419,029 *Saccharum officinarum* (sugar-cane) ESTs to be performed by Sifter-T in 5 days, while BLAST took 49 days in a standard bioinformatics laboratory machine.

Conclusions

This implementation result is presented as Sifter-T (Sifter Throughput-optimized), an open source tool with better usability and performance when compared to the original Sifter workflow implementation. The new Sifter-T features allow researchers to have easy and quick access to the Sifter’s powerful annotation mathematical method, now with enhanced experiment customization and keeping the inference engine intact. Sifter-T, and its online interface, is freely available at <http://labpib.fmrp.usp.br/methods/sifter-t/>.

Published: 30 April 2015

doi:10.1186/1471-2105-16-S8-A4

Cite this article as: Almeida-E-Silva and Vêncio: Sifter-T: A scalable framework for phylogenomic probabilistic protein domain functional annotation. *BMC Bioinformatics* 2015 **16**(Suppl 8):A4.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



* Correspondence: rvencio@usp.br
Universidade de São Paulo, São Paulo, Brazil