

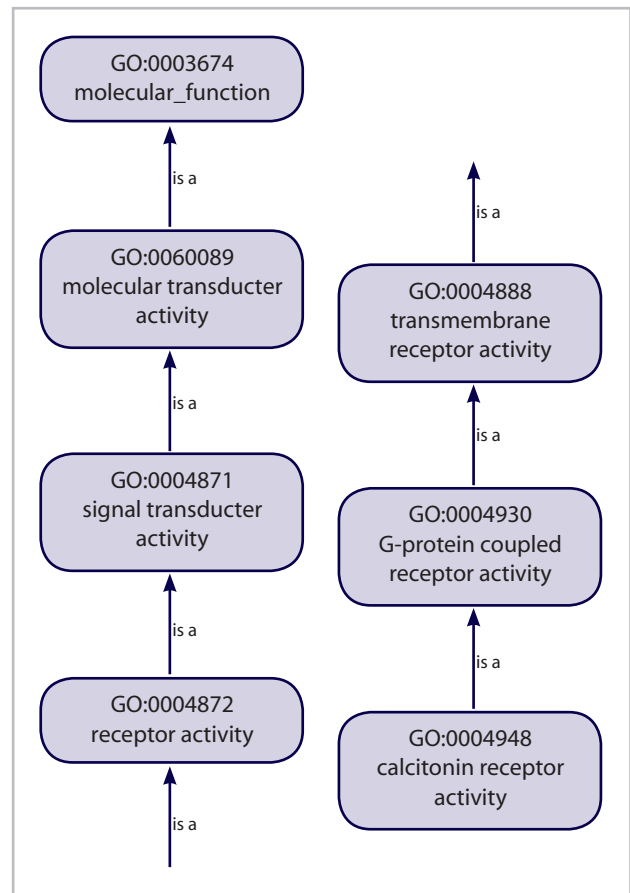
## INFORMATION TECHNOLOGY APPLIED TO BIOENERGY GENOMICS: PROBABILISTIC ANNOTATION USING ARTIFICIAL INTELLIGENCE

Ricardo Zorzetto Nicolielo Vêncio

Ribeirão Preto School of Philosophy, Sciences and Literature / University of São Paulo (USP)

FAPESP Process 2009/53161-6 | Term: Feb 2010 to Jan 2012

An alternative to the problem of fossil fuels depletion is the use of renewable energy. In Brazil, sugarcane (*Saccharum officinarum*) is used for years as alternative energy source and, therefore, Brazil has become a key player in alternative energy development. Our main aim is to develop methods and tools to attack some of the bioinformatics issues raised in sugarcane genomics research. In order to achieve this, we opt for Sifter (Engelhardt *et al.*, 2006), a powerful method based on Bayesian Networks. Our major aim was to establish a local implementation of the Sifter methodology for application in bioenergy related problems and the following is to improve the original source code performance, potentially allowing it to be used in a genome-wide scale.

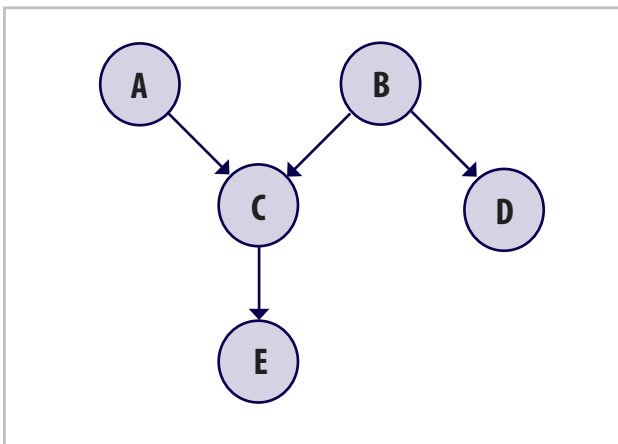


*A given gene's molecular function  
in the Gene Ontology representation*



## MAIN RESULTS

We currently have completed scripts that allow full automation of the pipeline of Sifter methodology, with average performance gain of about 72.5% (quad core machine) and 67.7% (dual core) in relation to original scripts supplied with the software. To achieve this goal, we changed the originally proposed pipeline, and, beyond that, we added new functions to the scripts aiming user friendly software and a better detection performance, under evaluation. This new pipeline is designed to enable the analysis proposed by Sifter methodology in high-throughput analysis.



Example of Bayesian Network (BN). The nodes represent random variables and the directed edges represent statistical dependence relationships. This BN compactly represents the following probability distribution:  $\Pr(A,B,C,D,E) = \Pr(E|C) \Pr(C|A,B) \Pr(D|B) \Pr(B) \Pr(A)$

## PRODUCTS/PUBLICATIONS

### POSTERS

Silva DCDA, Waldemarin RC, Vêncio RZN. Information technology applied to bioenergy genomics: probabilistic annotation using artificial intelligence. *X-Meeting 2010 – 6<sup>th</sup> International Conference of The Brazilian Association for Bioinformatics and Computational Biology*. November 15-18, 2010. Centro de Convenções e Artes da UFOP, Ouro Preto, Minas Gerais, Brazil.

---

**Ricardo Zorzetto Nicolliello Vêncio**

Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto  
Universidade de São Paulo (USP)  
Departamento de Computação e Matemática

Av. Bandeirantes, 3.900 – Monte Alegre  
14040-901 – Ribeirão Preto, SP – Brazil

rvencio@usp.br  
55.16.3602-3718