Metabolic pathfinding based on genetic algorithms

Matias Gerard[1,2], Georgina Stegmayer[1] and Diego Milone[2]

[1] CIDISI-UTN-FRSF, CONICET, Lavaise 610 - Santa Fe (Argentina)
[2] SINC(I)-FICH-UNL, CONICET, Ciudad Universitaria - Santa Fe (Argentina)

Background

Metabolic pathway searching consists of finding a set of reactions allowing to transform a compound into another one. There are several search methods based on classical algorithms like breath-first search (BFS) [1] and depth-first search (DFS) [2] to perform this task. However, there are problems in which a very high number of solutions must be explored, making classical methods practically inapplicable. Genetic algorithms use stochastic search to explore multiple points of the search space at the same time.

Material and methods

We propose a genetic algorithm (EAMP) that perform a two-end metabolic pathway search and compare its performance with two classical search algorithms. To achieve this, the chromosomes were built by attaching a reaction to each gene, and the left-to-right sequence of genes encoded a metabolic pathway. A initialization strategy to build variable size chromosomes with a partially conserved sequentiality of reactions was proposed. The crossover and mutation operators were designed to promote the building of a reaction chain. Fitness function was built to consider the validity of the reactions sequence, the presence of the compounds to relate and the occurrence of repeated reactions.

Results

EAMP was studied for several mutation rates and different initialization strategies. Results indicate that minimum searching time was reached for a mutation rate of 0.04 and the initialization strategy with initial variable size for the chromosomes. Comparison of EAMP with BFS and DFS is shown on Figure 1. Boxplots correspond to searching time and number of reactions of 120 pathways founds with each algorithm. DFS perform the search with minimum search time but produce solutions with maximum number of reactions allowed. BFS found shortest pathways but employ greater time than EAMP. The genetic algorithm perform the search using an intermediate time to BFS and DFS, and not only found the shortest pathways but also solutions with greater number of reactions linking the two compounds.

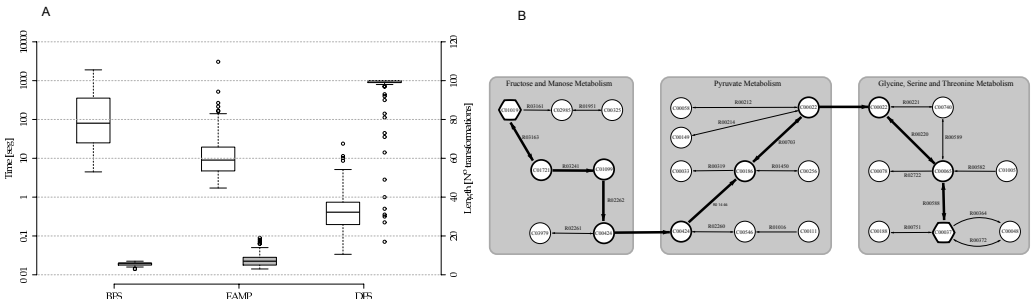

Figure 1. (A) Boxplots for searching time and number of reactions for EAMP, BFS and DFS algorithms. Searching times and the number of reactions are shown in in white and gray, respectively. Time is plotted in logarithmic scale. (B) Metabolic pathway linking C01019 and C00037. Pathway found is shown in bold line.

Conclusions

The proposed genetic algorithm found metabolic pathways using intermediate times to those required by BFS and DFS. Moreover, it builds metabolic pathways with variable size, including either shortest pathways and larger solutions. It is interesting from a biological viewpoint because pathways larger than shortest path could be provide relevant information about alternative metabolic pathways.

References

1. Ogata, H., Goto, S., Fujibuchi, W., Kanehisa, M.: Computation with the KEGG pathway database. BioSystems **47** (1998) 119–128
2. Faust, K., Croes, D., van Helden, J.: Metabolic Pathfinding Using RPAIR Annotation. Journal of Molecular Biology **388**(2) (2009) 390–414

sinc(*i*) Research Institute for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
M. Gerard, G. Stegmayer & D. H. Milone; "Metabolic pathfinding based on genetic algorithms"
3er. Congreso Argentino de Bioinformática y Biología Computacional (3CAB2C) - 2012, 2012.