NES²RA: A TOOL FOR GRAPEVINE TRANSCRIPTOMIC DATA MINING

Stefania Pilati*, Giulia Malacarne, Samuel Valentini, Francesco Asnicar, Luca Masera, Marco Moretto, Paolo Sonego, Valter Caveczia, Enrico Blanzieri and Claudio Moser.

FONDATION E. MACH, Research and Innovation Center, via E. Mach, 1 - 38010 San Michele all'Adige (TN) - Italy
CNR, Institute of Materials for Electronics and Magnetism, via alla Cascata, 56/C - 38123 Trento - Italy
UNIVERSITY OF TRENTO, Dept of Information Engineering and Computer Science, via Sommarive, 9 - 38123 Trento -Italy
*stefania.pilati@fmach.it

Introduction

The development of “omics” technologies to study gene expression has revolutionized our perspective from the single gene to the gene network level. However, the complexity of the system biology approach requires appropriate mathematical, computational and statistical tools to analyze data and extract information. Grapevine transcriptomic data obtained with both microarrays and RNAseq technologies have been collected into the Vitis Expression Studies Platform Using COLOMBOS Compendia Instances (VESPUCCI, Moretto et al., 2016). Here we present the application of the algorithm of Network Expansion by Subsetting and Ranking Aggregation (NES²RA, Asnicar et al., 2016) to the VESPUCCI database in order to expand four Local Gene Networks (LGNs) related to the grapevine response to climate changes. NES²RA is based on the systematic and iterative application of the PC algorithm - aimed at identifying causal relationships from observational data - on subsets of the input data. To overcome the computational power requirement of NES²RA algorithm, it has been run as part of the gene@home project, a distributed computation project which relies on thousands of volunteers’ computers managed by the TN-Grid, an infrastructure based on BOINC system (Asnicar et al., 2015).

Conclusion

The present study proposes NES²RA algorithm as a suitable tool to mine grapevine transcriptomic data in order to highlight biologically relevant relationships among genes. The networks obtained integrating the information about the genes and their interactions found by NES²RA provide an in silico hint to identify new genes of partially known metabolic and/or regulatory networks, as shown here. Beside, we are testing new applications, such as for example, the use of NES²RA to discriminate among the isoforms of a gene family by exploiting the gene interaction information, in addition to the more common sequence structure analysis. Finally, we are developing a strategy to make NES²RA available as a web tool that the biologist can interrogate in real time. One possibility under investigation consists in separating the gene expansion step, by pre-computing it, and the post-processing ranking step, computed on demand.