

Article

An integrative bioinformatic approach to aid in the gene ontology characterization of data from OMICs technologies: application to the study of a proteomics data set

Eduardo Zúñiga-León ¹, Ulises Carrasco-Navarro ¹ and Francisco Fierro-Fierro ^{1*}

¹ Departamento de Biotecnología, Universidad Autónoma Metropolitana-Unidad Iztapalapa, México D.F., México; pgen10@hotmail.com (E.Z.L.); ulises.c.n@gmail.com (U.C.N)

* Correspondence: mailto:degfff@yahoo.com (F.F.F)

Received: date; Accepted: date; Published: date

Abstract: Omics technologies generate a huge volume of data. In the last decade, overrepresentation or enrichment tools have played a successful role in the functional analysis of large genes/proteins list, which is clearly evidenced by thousands of publications citing these tools. Here, we present GOmics, a functional enrichment analysis tool, that identifies statistically overrepresented biological terms within given gene/protein set. This tool provides a hypergeometric distribution test to calculate biological terms enriched significantly. GOmics is adapted to use updated information from the two main annotations databases such as Gene Ontology and KEGG. GOmics is a tool with greater coverage to identify biological terms compared to others similar tools. GOmics also is able of builds networks-based different graphical representations from enriched results, this feature makes it an integral complete. GOmics is freely accessible at <https://github.com/bioinfproject/bioinfo/>

Keywords: Enrichment analysis, Gene Ontology, KEGG Pathways, Proteome, plot visualization

1. Introduction

Omics technologies are revolutionizing biological research by enabling comprehensive monitoring of a complex biological system. Functional annotation of data from these approaches is appealing for several reasons because reducing the complexity from hundreds or thousands of genes/proteins to few processes or pathways in which they are involved, and have more explanatory power than a simple list of identifiers. Due to increasing of these approaches and its importance on biological systems, several bioinformatics tools have been developed to perform functional annotations. Over-representation analysis (ORA) is the most popular bioinformatic methodology to obtain significant functional information (enrichment) from sets of related genes/proteins [1]. ORA method search in biological databases (e.g., Gene Ontology or KEGG) and use statistical testing to find biological terms and functional annotations that are significantly enriched in a list of genes/proteins. The aim of enrichment analysis is testing whether any biological annotations are over-represented in the query genes/proteins list compared to what would be expected in the whole population. In others words, if a set of proteins are significantly enriched in certain biological processes or pathways, they are likely to play a similar role *in vivo*. However, in most cases, the results of these analyses are very long lists of biological terms or pathways associated to genes/proteins that are difficult to digest and interpret, and sometimes it's necessary an appropriate visualization method. Although there are several enrichment tools, they do not all included graphical visualizations. Commonly the most of data related with annotations

are represented as Bar or Pie chart plots, however, not provide more details about additional relationship between the genes/proteins and GO terms/pathways. For instance, one protein could be involved in three or more relevant biological process or pathways, and a Bar or Pie chart plot no provide such information. The networks analysis has become an increasingly popular tool to deal with the complexity of a large dataset of all sorts. The importance of using networks lies in observation of relationships between factors, rather than seeing as isolated entities [2]. The intersection network is a bipartite network and applying this concept to biological systems, these connections allow to detect multifunctional proteins, that is to say, genes/proteins with more than two functions and involved in more of two processes or pathways.

Here we present Gomics, a Functional Annotation Tool developed in programming language Python and R that integrate ORA methodology and networks-based visualization. Gomics is an analysis tool command-line user friendly, which applies appropriate statistical methods to identify significantly enriched GO terms or pathways among a given list of genes/proteins. Gomics supports all organisms deposited in UniprotKB and KEGG databases. It also provides four types of graphical visualizations for show enriched results.

2. Materials and Methods

2.1. Implementation and usage

Gomics uses PERL commands and Python modules to enrichment analysis of a gene or protein list. It can provide plots using R packages to enrichment results visualization. The code is configured to use updated information from UniProt-Gene Ontology Annotation (UniProt-GOA) [3], UniProt Knowledgebase [4] and KEGG [5] databases. Gomics is a bioinformatic free tool, which run in Linux terminal. All the options of this tool have been extensively tested on Linux distribution within the bioinformatics community, named Ubuntu.

2.2. General usage

The flow chart of the data processing is depicted in figure 1. Gomics is available for download from GitHub: <https://github.com/bioinfproject/bioinfo/> as a Python script and can be simply executed without any need for special procedure. There are two forms of execute Gomics on command-line.

- a) `$ python Gomcis`
- b) `$ chmod +x Gomics`
- `$./Gomics`

Gomics can perform three enrichment analysis differents using updated databases (Figure S1A). The first analysis (1) is gene ontology using all information stored on UniProt-GOA (Complete GO Annotation) and UniprotKB (GO annotations). The second analysis (2) using all annotations stored on KEGG database for find pathways. Finally, third analysis (3) is more flexible because identify KEGG PATHWAYS from sequences proteins and can to use for both annotated and non-annotated organisms. After executing one of the previous analyzes, automatically generate graphics that are deposited in specific directories.

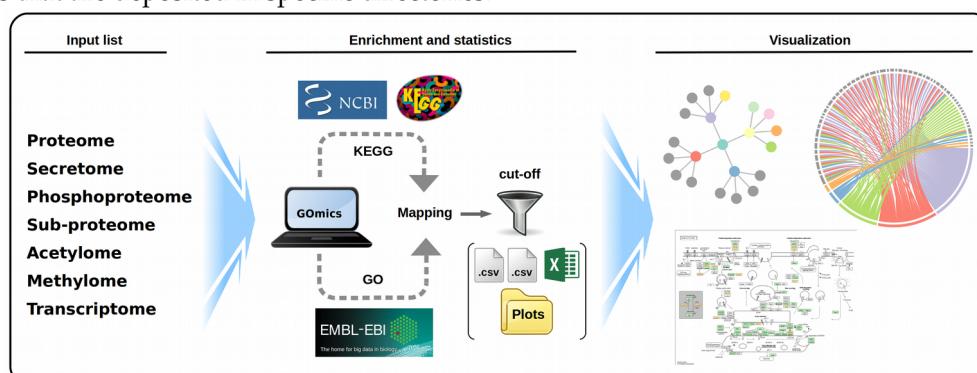


Figure 1. Schematic overview of Gomics. Gomics is composed of three main sections: input list, enrichment/statistics and visualization. A connection is made to the UniprotKB, Uniprot-GOA or

KEGG databases and after the content of input list is analyzed. The enrichment/statistics section organize and analyze data of input list, retrieves information for all genes or proteins and results are stored in three output files, two in csv format and one in Microsoft Excel formart. The visualization section provides four types of graphical representations in png format and high definition.

2.3. Input file

GOmics use a plain text containing a list of genes (KEGG gene ID) or proteins (Uniprot Entry ID). This list contains gene or proteins that are affected in any particular condition in study. The file can contain three columns (in Tabular format) depending on the results obtained from "Omics" approaches. First column correspond to gene or protein list with changes of expression or abundance, respectively. Second column correspond to numerical values of expression or abundance, or any value related with study. The third column in input file correspond to a background list, this can be all genes identified in a transcriptome, proteins identified in a proteome, or complete proteome. GOmics tool comes with two additional protein list for *Homo Sapiens* and *Beauveria bassiana*, for test this tool. For facilitates the use of identifiers, GOmics allows only identifiers compatibles with UniprotKB and KEGG database.

2.4. Annotations sources

The Gene Ontology Annotation (GOA) database aims to provide highquality electronic and manual annotations to the UniProt Knowledgebase (Swiss-Prot, TrEMBL and PIR-PSD) using the standardized vocabulary of the Gene Ontology (GO) . The GOA database contained all annotated proteomes. This database displays both manually and electronically assigned annotations. The GOA database has nearly 60000 species, more than 160000 taxa, with more than 32 million annotations. The GOA association files are download from FTP in tab-delimited format (<ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/>). GO annotation in UniProtKB (Swiss-Prot and TrEMBL, PIR-PSD) are flat files and can be downloaded from: (<http://www.ebi.ac.uk/uniprot/>). In Swiss-Prot entries, only the manually annotated information is displayed. To view the complete GO annotation for a Swiss-Prot entry, you should download or browse the master copy of the data in the GOA association files. GOA updates its annotation display weekly, while UniprotKB occurs every four weeks. KEGG cover information to different molecular levels. The KEGG PATHWAYS (<https://www.genome.jp/kegg/pathway.html>) database is a collection of manually curated with molecular interaction, reactions and relation network information. The KEGG database has more than 24 million of genes annotated and more than 6 million pathway linked genes.

2.5. Background

If the third column is absent in input file, the program automatically uses the entire proteome in the case of Uniprot-GOA or the entire KEGG database. For gene ontology analysis, GOmics build backgrounds organized by category (e.g., Biological Process, Molecular Function and Cellular Component) of a specific organism, which are used for mapping the proteins list. GOmics download a GOA association file (<ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/proteomes/>), of which extract all GO terms. For KEGG PATHWAYS analysis GOmics build a background file with all genes of a specific organism for mapping the gene list.

2.6. Enrichment Analysis

GOmics analyze the input list against user's preferred background and retrieves the genes or proteins ranked by functions with statistical significance. This tool adopts GeneMerge1.4 statistical algorithm for obtain over-representation of particular functions or categories of input list [6]. An hypergeometric distribution test is perform to calculate the discrete probability of x (term or pathway) of one kind in a sample (protein list) of size k drawn from a background of size n (proteome), where m is equal to proteins annotated of x in the background (Figure S1B). An FDR or modified Bonferroni correction of P-value is applied to identify the statistically more represented functional annotations.

2.7. Output file

GOmics generate three final output files stored in specific directories with particular functions or categories over-represented with statistical significance (Figure 1). I) An edges file with two columns as Source and Target, which contains the values of individual nodes that are linked together. This file can be used with other network-based tools like Cytoscape [7] and Gephi [8]. II) A nodes file with information as nodes ID, expression values, P-value and additional information for build plots. III) Finally, an excel file with enrichment analysis results for plotting in others programs. GOmics also provides automatically four types of graphical representations in high definition to facilitate the analysis and interpretation of results, these are: Circular and Random Network, Chord diagram [9] and UpSet plot [10]. These graphics are built with R packages as tidyverse, tidygraph, ggraph, circlize, igraph, networkD3 and UpSetR, which are used to build networks. The graphics are configured according to the amount and type of data over-represented in enrichment analysis.

3. Results

GOmics is a integral tool with two major features: it allows enrichment analysis a given list with data from different “Omics” approaches and builds different graphical representations in network form from enrichment results. GOmics is an enrichment analysis tool command-line, and is able to analyze from five to thousands of identifiers with any of three analysis mentioned above. This tools is easy of use and fast, although the time depends on internet connection and amount of data entered. The databases used by GOmics are automatically updated, which allows that the users will always receive the most updated analysis results. For to use GOmics it's not necessary to download any database, neither a special edition of files is required.

3.1. Testing data set

To demonstrate the versatility and functionality of GOmics we selected two subset of proteins from different experimental data proteomics from two recent publications. The first data set come from a platelet proteome of patients in early-stage cancer [11], and second set come from a proteome in *Beauveria bassiana*, a fungal insect pathogen [12].

3.2. Case study 1: Enrichment analysis in early-stage lung or head of pancreas cancer

Platelets play an important role in tumor angiogenesis, growth and metastasis [13]. The study from Sabrkhana et al, 2018 [11] identified several differentially expressed proteins associated with early-stage cancer. This experiment was performed in 12 cancer patients (8 with Lung cancer and 4 with Pancreatic cancer) and 11 controls. To illustrate how to work GOmics, we analyze an input list with used 31/18 upregulated/downregulated proteins, respectively, of platelets from patients with early-stage cancer compared to platelets from healthy controls, and a background of 4496 proteins (identified in proteome). In this analysis, we perform an enrichment analysis for gene ontology (FDR 0.02) and KEGG pathways (FDR 0.05). In our study we identified seventeen processes associated to inflammatory response, immune response and cancer using UniprotKB annotations. On the other hand, four proteins (P06702, P05109, Q9UKW4, P28907) are involved in more than two processes, which may be multifunctional proteins, as is the case of P28907 protein, which it's classified as moonlight protein. Also were identified four proteins are related directly with angiogenesis process. The enrichment with KEGG database shows five pathways related to immune response. From input list, just three differentially expressed proteins (Q95365/3106, P30499/3107 and Q9UKW4/10451) share several pathways.

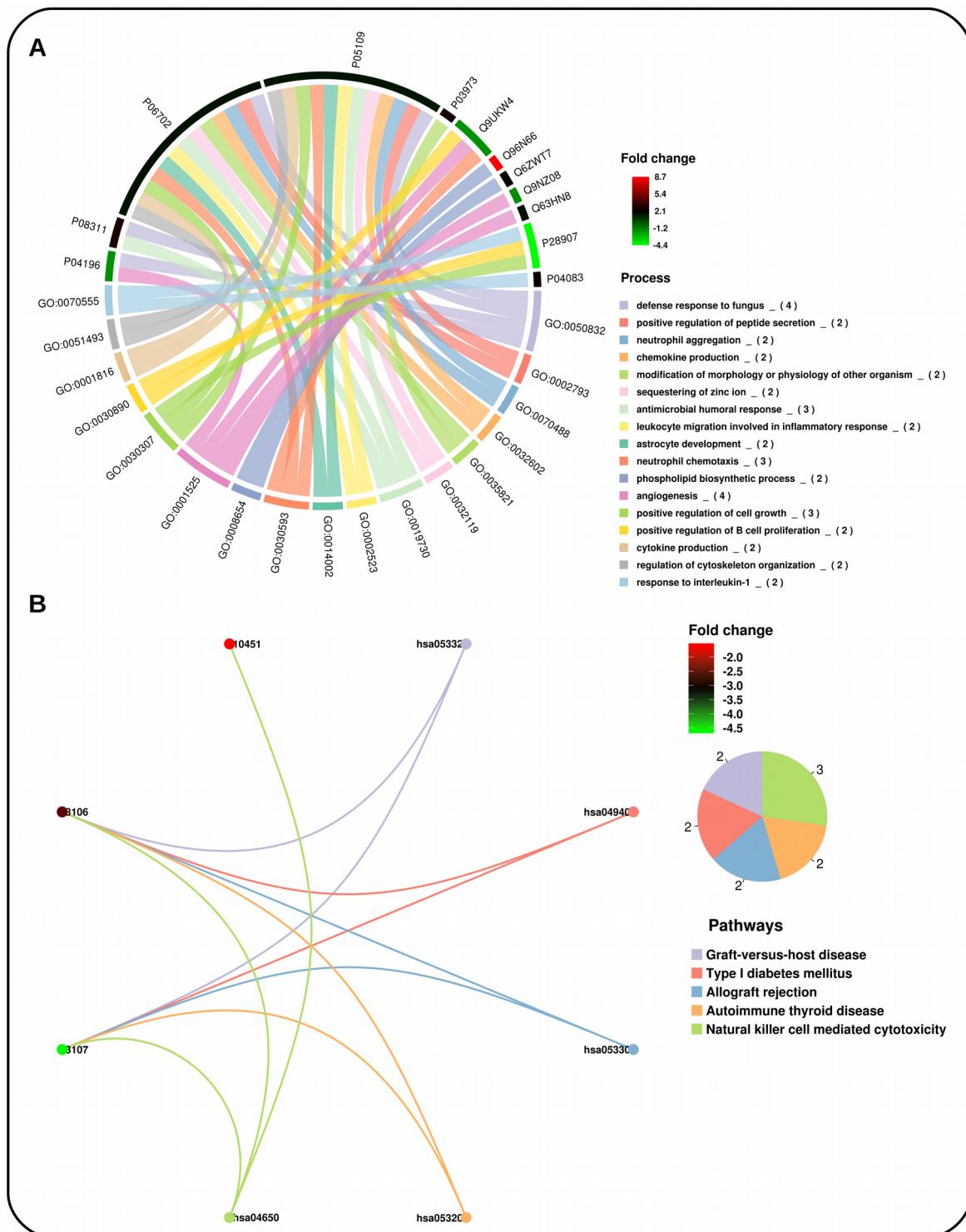


Figure 2. Enrichment analysis in early-stage cancer. A) Chord diagram clustered by colors with a color scale for fold change values and a description of GO terms enriched. Chord diagram allows shows up-regulated (fold change ≥ 1.5)/down-regulated (fold change ≥ 1.5) proteins. B) Circular network clustered by colors were the nodes color of proteins correspond to up- or downregulated, with one mini Pie Chart with the number of genes or proteins by pathway. The identifier of nodes differentially expressed correspond to NCBI-GeneID.

3.3. Case study 2: Identification of GO terms from Proteome and Phosphoproteome dataset in *Beauveria bassiana*

A proteome indicates the quantitative protein expression profile of a cell, an organism, or a tissue under exactly defined conditions [14]. A recent study by Wang et al, 2016 [15] with Cdc14, a

dual-specificity phosphatase that regulates nuclear behavior by dephosphorylating phosphotyrosine and phosphoserine/phosphothreonine in fungi and found that Cdc14 to act as a positive regulator of cytokinesis, asexual development and multiple stress responses in *Beauveria bassiana*, a fungal insect pathogen. This experiment was performed in *B. bassiana* $\Delta cdc14$ mutant versus WT under different stress conditions. In our study we used 86/239 and 80/99 proteins/phosphoproteins significantly regulated under NaCl (1 M) and H₂O₂ (3 Mm) stress, respectively, and performed the enrichment analysis using GOmics (Figure 3). This analysis was done using a FDR of 0.05 (FDR 0.01 for 239 phosphoproteins under NaCl was used) and complete proteome as background. Fatty acid elongation, glyoxylate cycle, and fatty acid biosynthesis process were ranked as the most significantly affected processes in $\Delta cdc14$ mutant under NaCl (1 M) stress, showing all proteins are downregulated. On the other hand, the most of phosphoproteins are involved in several translational related processes (Figure 3). Proteolysis, amine metabolic process and response to oxidative stress, and glycogen biosynthetic process, proton export across plasma membrane and regulation of protein catabolic process, in proteome and phosphoproteome, respectively, were ranked as the most significantly affected processes in $\Delta cdc14$ mutant under H₂O₂ (3 Mm) stress (Figure 4). On the other hand, in the proteomes under NaCl (1 M) and H₂O₂ (3 Mm) stress, both showed oxidation-reduction process, which is related to stress conditions.

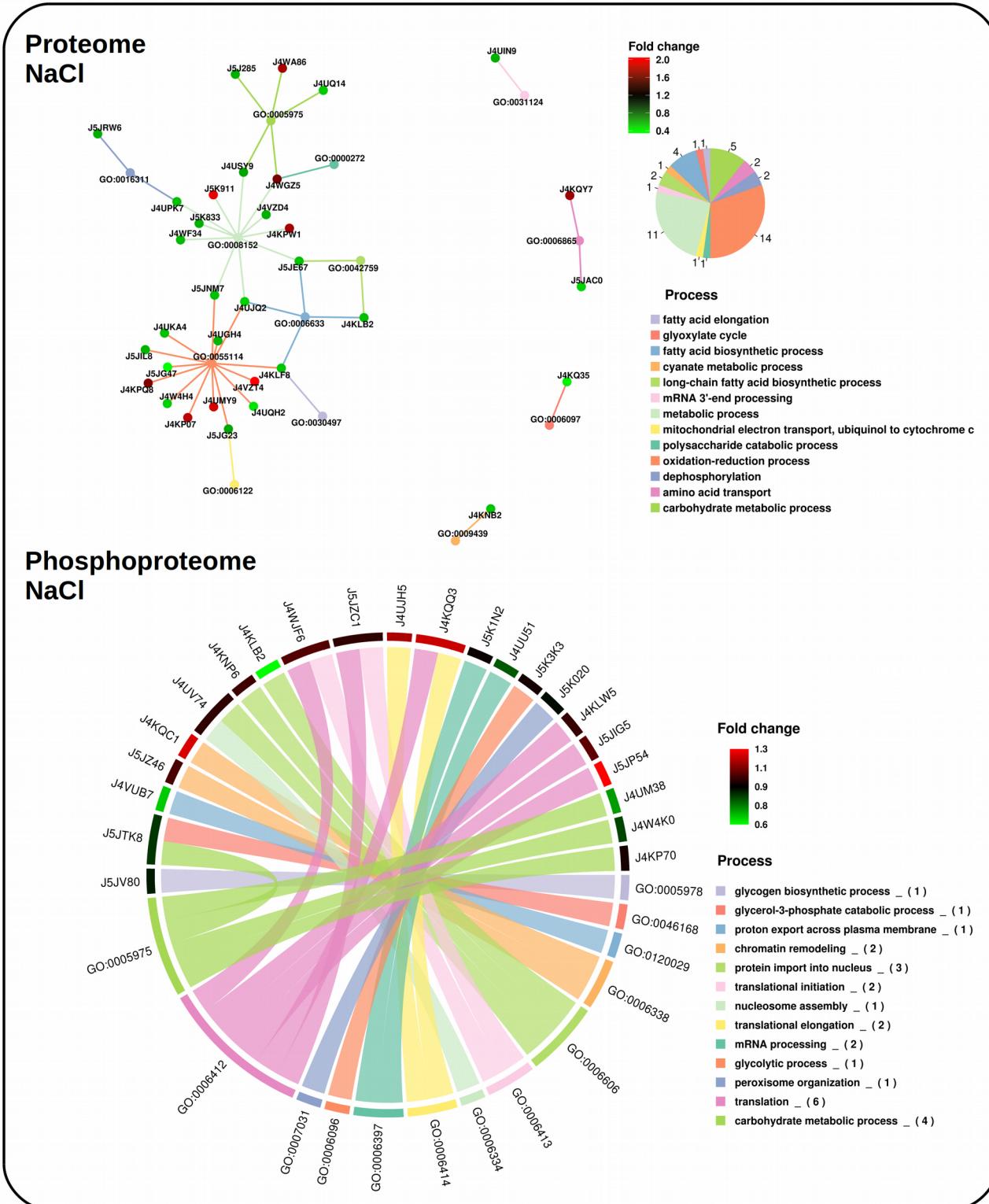
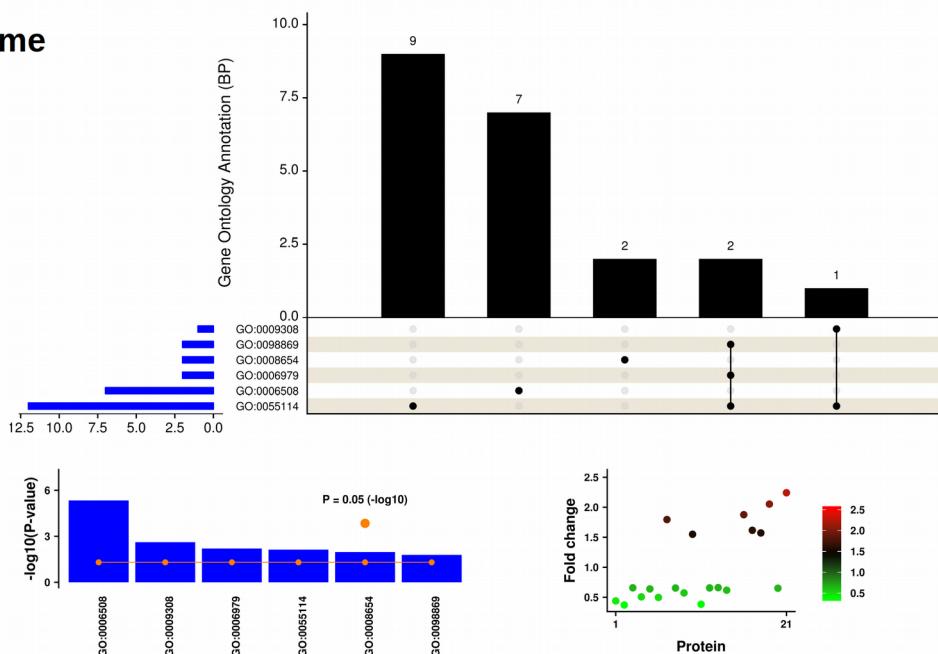


Figure 3. Enrichment analysis of proteomic and phosphoproteomic data under NaCl (1 M) stress. GOMics build different graphical representations using results from enrichment analysis showed GO terms ranked by statistical significance. The graphics included the ratio of each protein abundance of *Δcdc14* mutants over WT (up- or downregulated if the ratio is >1.5 or <0.67 , respectively). The upper figure presents a network clustered by colors were the nodes color of proteins correspond to up- or downregulated, with one mini Pie Chart with the number of proteins by process. The lower figure shows a plot called Chord diagram clustered by colors with a color scale for abundance values and a description of processes enriched.

Proteome H_2O_2



Phosphoproteome H_2O_2

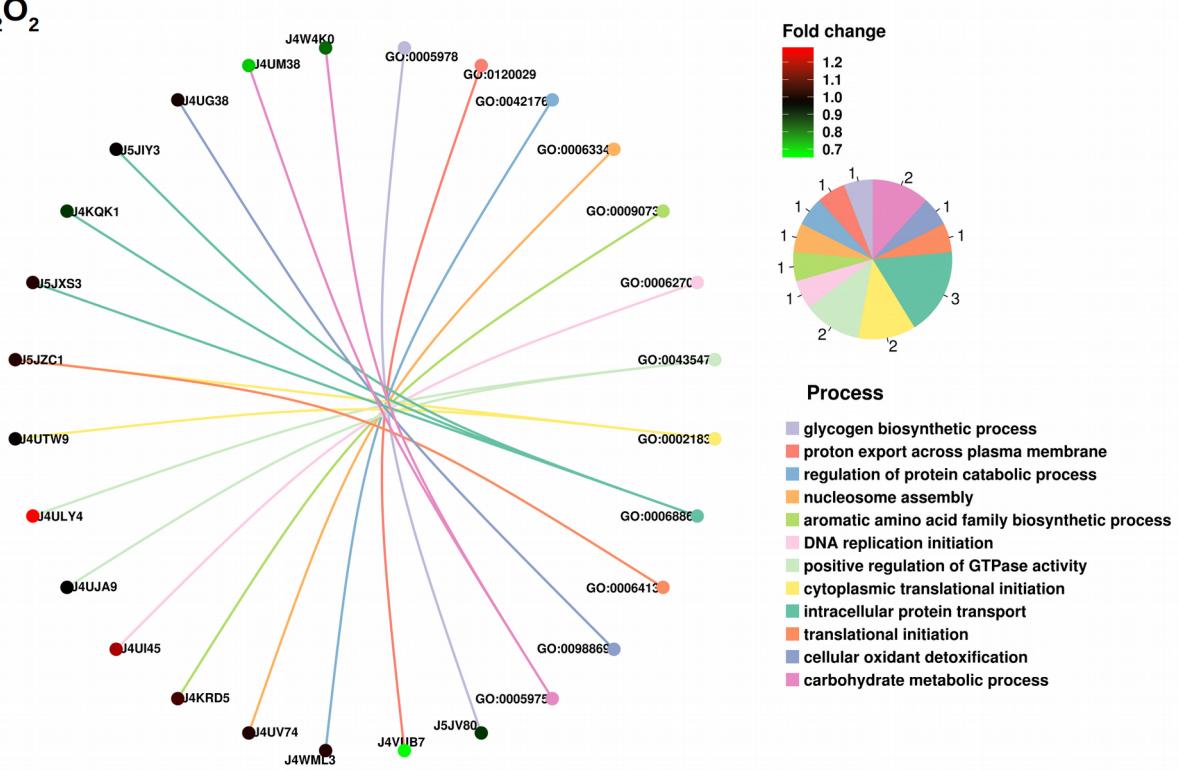


Figure 4. Enrichment analysis under H_2O_2 (3 mM) stress. GOmics build different graphical representations using results from enrichment analysis showed GO terms ranked by statistical significance. The graphics included the ratio of each protein abundance of $\Delta cdc14$ mutants over WT (up- or downregulated if the ratio is >1.5 or <0.67 , respectively). The upper figure shows a Upset plot shows intersections between GO terms and additional plots based on elements in the intersections as processes with $P\text{-value} \leq$ or ≥ 0.05 (left) and expression values (right). The lower figure presents a Circular network clustered by colors were the nodes color of proteins correspond to up- or downregulated, with one mini Pie Chart with the number of proteins by process.

3.4. Comparison and advantages of GOMics over others functional enrichment analysis tools

GOMics is designed to perform enrichment analysis using updated information of the two main annotation databases (Gene Ontology and KEGG) of genes and proteins. GOMics has the versatility to use Uniprot-GOA and UniprotKB as databases separated, for obtaining complete and manually cured information, respectively, in order to obtain as much information as possible. There are other available tools that can perform enrichment analysis on a set of gene or protein lists. However, these tools update their functional information less frequently and thus become less reliable for performing functional enrichment analysis. On the other hand, some of these tools not included more than two databases and not offer graphical representations to aid to visualize results (Table S1). GOMics supports all organisms deposited in UniprotKB (<https://www.uniprot.org/proteomes/>) and KEGG (https://www.kegg.jp/kegg/catalog/org_list.html) databases, a number much higher than others similar tools as g:Profiler [16], GOrilla [17], GOEAST [18] and WebGestalt [19], which limits the enrichment analysis excluding species of non-model organisms (Table S1). To test whether GOMics offers a difference on results of enrichment analysis, we perform a comparison between some publicly available enrichment tools such as g:Profiler, GOrilla, GOEAST, and WebGestalt of GO terms related to Biological Processes using the same data of Case study 1 previously described, with a FDR 0.05 in all cases. Although some of these tools have restrictions in organisms number, the *Homo sapiens* species is found in all these tools. As shown in the figure 5, GOMics showed greater coverage of GO terms identified compared to others tools. With GOMics were detected GO terms as GO:0002523 (leukocyte migration involved in inflammatory response), GO:0050727 (regulation of inflammatory response), GO:0030890 (positive regulation of B cell proliferation), GO:0030307 (positive regulation of cell growth), GO:0001525 (angiogenesis), GO:0070555 (response to interleukin-1), GO:0060333 (interferon-gamma-mediated signaling pathway), GO:0050729 (positive regulation of inflammatory response) and GO:0002224 (toll-like receptor signaling pathway) absent in the others tools, and which are related to processes immune response and cancer. Unlike GOMics, the others tools do not use update functional information for enrichment analysis, therefore there were fewer proteins detected compared to GOMics. In this analysis GOrilla covered only 54% (12 proteins) of proteins, while g:Profiler and WebGestalt covered 36 (8 proteins) and 40% (9 proteins), respectively, indicating that the GOMics provides better functional enrichment information. This comparative example allowed to show the coverage and versatility of GOMics about others similar tools. Since GOMics performs analysis by directly using the information from updated databases, users need not additional tools to update the functional information. On the other hand, GOMics, in addition to performing enrichment analysis also generates a set of graphs to facilitate the visualization of results. GOMics is capable of generating four different graphics by analysis depending of type and amount data, while some of the analyzed tools generate simple visualizations.

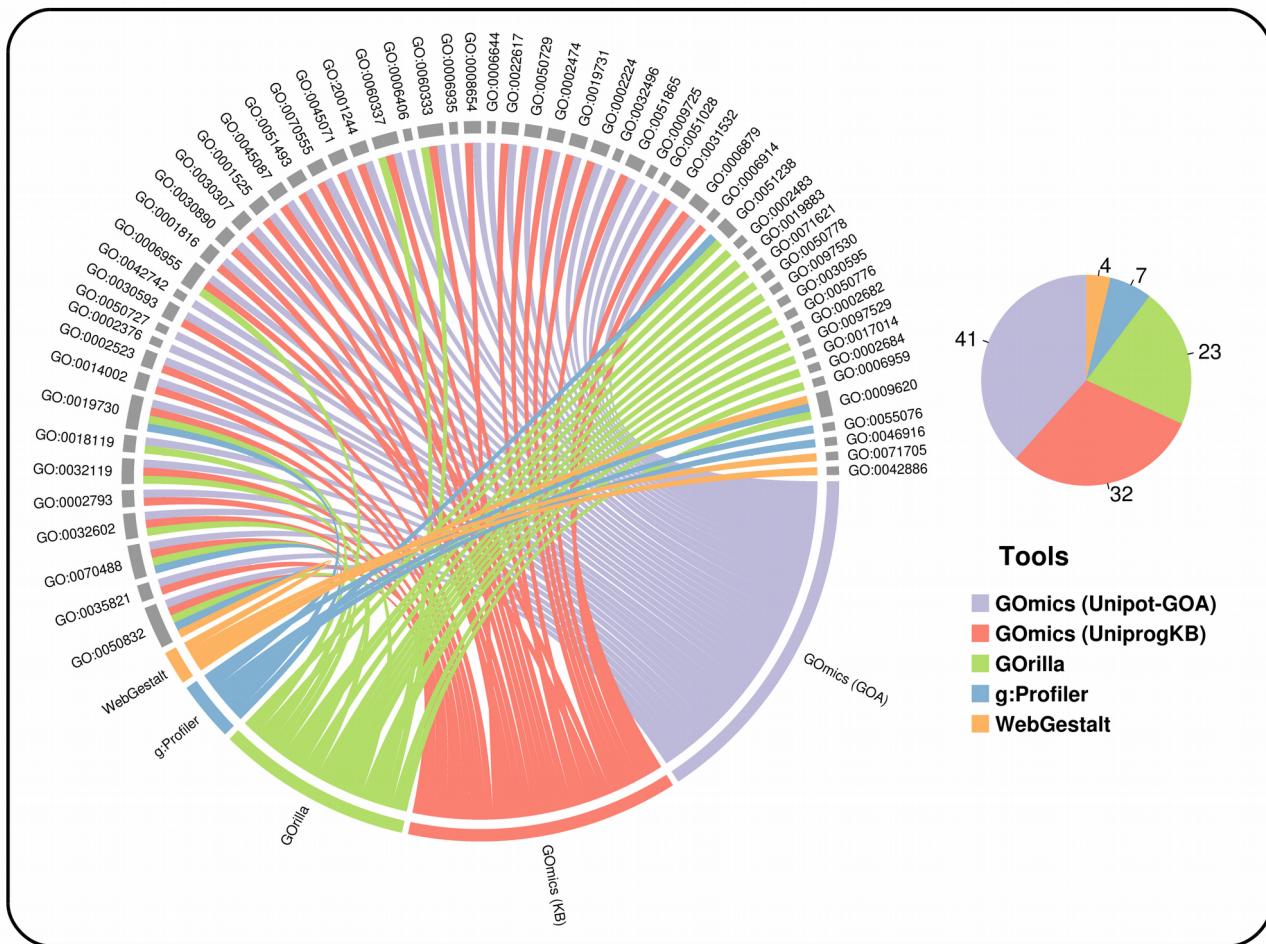


Figure 5. Comparison of GOMics with other similar tools. This analysis shows a comparison of GO terms related to Biological Process using the same data of Case study 1 previously described with a FDR 0.05 in all cases. GOEAST also was included in this analysis, however, did not recognize the identifiers and generated an error message. The mini Pie Chart show total GO terms identified by each tool.

4. Discussion

We provided a ORA-based tool named GOMics to facilitate the analysis and interpretation of high-throughput data for a large number organisms. GOMics can be applied to several “Omics” approaches such as proteome, phosphoproteome, secretome, sub-proteome, acetylome , methylome and transcriptome). The versatility of GOMics allows to analyze from five to thousands of identifiers with any of three analysis included, although the analysis time depends on internet connection and amount of data entered. GOMics also allows to include expression data or abundance, or any value related with genes/proteins, this to aid interpreting and understand the role of genes/proteins down-up regulated under different biological conditions. As proof of the concept, we perform a functional enrichment with one protein set using our tool. In this test GOMics showed greater coverage of GO terms identified compared to others similar tools. On the other hand, in addition to functional annotation by enrichment analysis, the graphical visualizations are important. The goal of graphical visualization is to analyze, explore, discover, illustrate, and communicate information in well understandable form. GOMics is able of builds networks-based different graphical representations from enrichment results, this feature makes it an integral complete. GOMics supports the three categories included in Gene Ontology (Biological Process, Molecular Function and Cellular Component) in their analysis. Our next version may include Reactome (<https://reactome.org/>) and NCG (<http://ncg.kcl.ac.uk/>) databases (only for *Homo sapiens*), also may include domains enrichment and networks-based on protein-protein interactions.

Supplementary Materials: Supplementary Data are available at <https://github.com/bioinfproject/bioinfo/>.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used “Conceptualization, X.X. and Y.Y.; Methodology, X.X.; Software, X.X.; Validation, X.X., Y.Y. and Z.Z.; Formal Analysis, X.X.; Investigation, X.X.; Resources, X.X.; Data Curation, X.X.; Writing-Original Draft Preparation, X.X.; Writing-Review & Editing, X.X.; Visualization, X.X.; Supervision, X.X.; Project Administration, X.X.; Funding Acquisition, Y.Y.”, please turn to the [CRediT taxonomy](#) for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

Funding: Please add: “This research received no external funding” or “This research was funded by [name of funder] grant number [xxx]” and “The APC was funded by [XXX]”. Check carefully that the details given are accurate and use the standard spelling of funding agency names at <https://search.crossref.org/funding>, any errors may affect your future funding.

Acknowledgments: In this section you can acknowledge any support given which is not covered by the author contribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments).

Conflicts of Interest: Declare conflicts of interest or state “The authors declare no conflict of interest.” Authors must identify and declare any personal circumstances or interest that may be perceived as inappropriately influencing the representation or interpretation of reported research results. Any role of the funders in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript, or in the decision to publish the results must be declared in this section. If there is no role, please state “The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results”.

References

1. P. Khatri, M. Sirota, and A. J. Butte, “Ten years of pathway analysis: Current approaches and outstanding challenges,” *PLoS Comput. Biol.*, vol. 8, no. 2, 2012.
2. X. Ma and L. Gao, “Biological network analysis: insights into structure and functions,” *Brief. Funct. Genomics*, vol. 11, no. 6, pp. 434–442, Nov. 2012.
3. R. P. Huntley *et al.*, “The GOA database: gene Ontology annotation updates for 2015.,” *Nucleic Acids Res.*, vol. 43, no. Database issue, pp. D1057–63, Jan. 2015.
4. E. Camon *et al.*, “The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology,” *Nucleic Acids Res.*, vol. 32, no. 90001, p. 262D–266, Jan. 2004.
5. M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, “KEGG as a reference resource for gene and protein annotation.,” *Nucleic Acids Res.*, vol. 44, no. D1, pp. D457–62, Jan. 2016.
6. C. I. Castillo-Davis and D. L. Hartl, “GeneMerge - Post-genomic analysis, data mining, and hypothesis testing,” *Bioinformatics*, vol. 19, no. 7, pp. 891–892, 2003.
7. P. Shannon *et al.*, “Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks.”
8. M. Bastian, S. Heymann, and M. Jacomy, “Gephi? An Open Source Software for Exploring and Manipulating Networks Visualization and Exploration of Large Graphs.”
9. Z. Gu, L. Gu, R. Eils, M. Schlesner, and B. Brors, “Genome analysis circlize implements and enhances circular visualization in R,” vol. 30, no. 19, pp. 2811–2812, 2014.
10. J. R. Conway, A. Lex, and N. Gehlenborg, “UpSetR: An R package for the visualization of intersecting sets and their properties,” *Bioinformatics*, vol. 33, no. 18, pp. 2938–2940, 2017.
11. S. Sabrkhany *et al.*, “Exploration of the platelet proteome in patients with early-stage cancer,” *J. Proteomics*, vol. 177, pp. 65–74, Apr. 2018.
12. Z.-K. Wang, J. Wang, J. Liu, S.-H. Ying, X.-J. Peng, and M.-G. Feng, “Proteomic and Phosphoproteomic Insights into a Signaling Hub Role for Cdc14 in Asexual Development and Multiple Stress Responses in *Beauveria bassiana*,” 2016.
13. L. J. Gay and B. Felding-Habermann, “Contribution of platelets to tumour metastasis,” *Nat. Rev. Cancer*, vol. 11, no. 2, pp. 123–134, Feb. 2011.
14. R. Kellner, “Proteomics. Concepts and perspectives.,” *Fresenius. J. Anal. Chem.*, vol. 366, no. 6–7, pp. 517–24.

15. Z.-K. Wang, J. Wang, J. Liu, S.-H. Ying, X.-J. Peng, and M.-G. Feng, "Proteomic and Phosphoproteomic Insights into a Signaling Hub Role for Cdc14 in Asexual Development and Multiple Stress Responses in *Beauveria bassiana*," *PLoS One*, vol. 11, no. 4, p. e0153007, 2016.
16. J. Reimand *et al.*, "g:Profiler-a web server for functional interpretation of gene lists (2016 update).," *Nucleic Acids Res.*, vol. 44, no. W1, pp. W83-9, 2016.
17. E. Eden, R. Navon, I. Steinfeld, D. Lipson, and Z. Yakhini, "GORilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists.," *BMC Bioinformatics*, vol. 10, p. 48, Feb. 2009.
18. Q. Zheng and X.-J. Wang, "GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis.," *Nucleic Acids Res.*, vol. 36, no. Web Server issue, pp. W358-63, Jul. 2008.
19. B. Zhang, S. Kirov, and J. Snoddy, "WebGestalt: an integrated system for exploring gene sets in various biological contexts.," *Nucleic Acids Res.*, vol. 33, no. Web Server issue, pp. W741-8, Jul. 2005.



© 2018 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).