W368–W371 *Nucleic Acids Research, 2008, Vol. 36, Web Server issue* *Published online 14 May 2008*
doi:10.1093/nar/gkn256

SerbGO: searching for the best GO tool

J. L. Mosquera[1,2,*] and A. Sánchez-Pla[1,2]

[1]Statistics and Bioinformatics Research Group, Departament d'Estadística, Universitat de Barcelona. Av. Diagonal 645, 08028 Barcelona and [2]Statistics and Bioinformatics Unit, IRHUVH. Passeig Vall d'Hebron 119-129, 08035 Barcelona, Spain

Received February 3, 2008; Revised April 10, 2008; Accepted April 20, 2008

ABSTRACT

In recent years, the scientific community has provided many tools to assist with pathway analysis. Some of these programs can be used to manage functional annotation of gene products, others are oriented to exploring and analyzing data sets and many allow both possibilities. Potential users of these tools are faced with the necessity to decide which of the existing programs are the most appropriate for their needs. SerbGO is a user-friendly web tool created to facilitate this task. It can be used (i) to search for specific functionalities and determine which applications provide them and (ii) to compare several applications on the basis of different types of functionalities. Iterating and combining both functionalities can easily lead to selecting an appropriate tool. Data required by SerbGO is either the desired capabilities within a defined *Standard Functionalities* Set or the list of the tools to be compared. The analysis performed carries out a cross-classification that produces an easily readable output with the list of tools that implement the capabilities demanded or a table with the categorization of the GO tools that one wishes to compare. SerbGO is freely available and does not require a login. It can be accessed either directly at our server (http://estbioinfo.stat.ub.es/apli/serbgo) or at the GO Consortium website (http://www.geneontology.org/GO.tools.microarray.shtml#serbgo).

INTRODUCTION

Modern experimental technologies, such as DNA microarrays (1), have become both popular and affordable over the last decade, leading to a considerable increase in experiments and publicly available functional genomic data sets. These high-throughput methodologies pose different challenges: the experiment itself, the statistical analysis of the data and the obtention of biological knowledge from the data. For example, in gene-expression microarray studies, it is very common for the statistical analysis to yield long lists of genes and one of the main challenges is how to give these lists a biological interpretation (2). It might be reasonable to expect that this could be done relying on the information stored in the existing biological databases, which can help to relate the experimental results with previously existing biological knowledge.

A useful resource to achieve both the goal of interpretation and the need of automation is the Gene Ontology (GO) (3). The GO is a cooperative project, which was set in motion in the late 90s, developed and maintained by the GO Consortium. Briefly, it is an annotation database originated 'to provide a controlled vocabulary to describe gene and gene product attributes in any organism'. It consists of three independent ontologies: Biological Process (BP), Molecular Function (MF) and Cellular Component (CC). Each of them is represented as a directed acyclic graph (DAG) (4) with two kinds of relationships ('is-a' and 'part-of') and whose nodes are the GO terms arranged from the most specific ones at the bottom to the only one at the top which is the most general term. The gene products may be linked to one or more GO terms in these ontologies. Thus, when a given gene has been annotated to a GO term it is also linked to its related nodes.

In recent years, many tools have been developed to assist analysis of experimental results based on the GO. Some of these tools are intended to manage functional annotations while others are specific for analyzing gene lists and many allow both possibilities (5). The scientific community has rapidly moved from lacking the appropriate GO tools to having a wide range of applications with, seemingly, very similar capabilities. It seems reasonable to ask ourselves whether it is worthwhile to keep developing new variants of the same programs. We may have reached the point where most needs may be solved by an already existing tool and the problem is simply deciding between those tools available.

This article presents a web-based application called SerbGO (Searching for the best GO tool), intended to help users to select the tools which best suit their needs as well

*To whom correspondence should be addressed. Tel: +34 93 402 15 60; Fax: +34 93 411 17 33; Email: jlmosquera@ir.vhebron.net; jlmosquera@gmail.com

as to easily compare the capabilities of various applications in the context of their experiments.

GO TOOLS AND THE STANDARD FUNCTIONALITIES SET

Due to the high heterogeneity among different types of tools it was decided to focus only on 'Tools for Gene Expression/Microarray Analysis' (http://www.geneontology.org/GO.tools.microarray.shtml).

To build SerbGO, a long list of applications available at the GO website (microarray tools) was reviewed from the existing literature. These tools use either the ontologies or the gene associations provided by the GO Consortium to facilitate the analysis of gene expression data.

The review yielded a substantial number of heterogeneous features, which were grouped into a potential set of functionalities. After several iterations, the features initially selected were converted into specific functionalities once redundancies were excluded. This process resulted in a set of features arranged in 205 standard functionalities.

The capabilities of the GO tools analyzed were classified *in situ* according to the *Standard Functionalities Set* and taking the following criteria into account:

(1) The functionality was available in the GO tool.
(2) The functionality was mentioned in the publication but it could not be validated.
(3) The functionality was not found in the paper or the application.

The list of applications which was finally included with their references is provided as Supplementary Material.

These tools use either the ontologies or the gene associations provided by the GO Consortium to facilitate the analysis of gene expression data. It must be noted that inclusion in the GO website does not imply approval by the GO Consortium and does not mean the tool has been tested or has been found to use information accurately. It can be said that this list 'is provided to promote an exchange of information between users and software developers'.

APPLICATION OUTLINE

Inputs

SerbGO is a web-based application designed to (i) facilitate researchers the task of determining which of the existing tools are appropriate for their needs and (ii) to enable a comparison between some of the available tools.

(1) The input needed to select those tools with the desired set of capabilities is a list of functionalities from the Standard Functionalities Set.
(2) The input needed to compare several tools is the list of programs to be compared.

Both actions can be performed interactively using the *Query Form* or the *Compare Tools* menu options (Figure 1).

Table 1. Number of standard functionalities per section

Section	No. of functionalities
Tools for	2
Type of experiment	7
Interface	7
Availability	4
Supported species	26
Data	40
Annotation	70
Statistical analysis	26
Output	23

Tools analyzed were classified according to a set of 205 standard functionalities arranged in nine sections.

Tool selection

The Query Form menu option at the top of the page allows the user to select different functionalities and to get the most appropriate tools to provide them. This form contains the *Standard Functionalities Set* arranged in nine sections (Table 1) and spread out over six pages.

To find the 'right tool' a user selects the desired functionalities by checking the appropriate fields at the specific sections (Figure 1A–C). Once the choices have been made for a page it is required to validate the query by clicking on the 'Next' button at the bottom of the page, which allows the user to move on the following one. The next page will show the new sections and the remaining tools will appear at the top-right corner. At the last selection page a 'Find' button will appear instead of 'Next' button. This new button allows users to move on to the outputs after validation.

Nonavailable features are shown as shaded colors. They can be activated by switching the corresponding radio button. In such cases, the user could have access to this option by switching on the previous radio button.

Queries are implemented with the logical operator AND. That is, the more capabilities are selected, the less tools will be available.

During the process of navigation over the pages, and at any time, it is possible to start a new query if the user clicks on the Query Form menu option at the top of the page.

Tool comparison

By checking any of the tools in the *Compare Tools* form, a list of their capabilities according to the *Standard Functionalities Set* can be obtained (Figure 1F).

Outputs

The output for the *Query Form* is a table with an alphabetically sorted list of the tools performing the functionalities demanded, the name of the developer and the name of the tool linked to its corresponding site (Figure 1D). The programs shown can be compared by clicking the Find button at the bottom of the results page (Figure 1E).

The output page for the Compare Tools form shows a table where rows contain the categorized functionalities and columns contain the GO tools names, which are linked to their respective sites (Figure 1E).

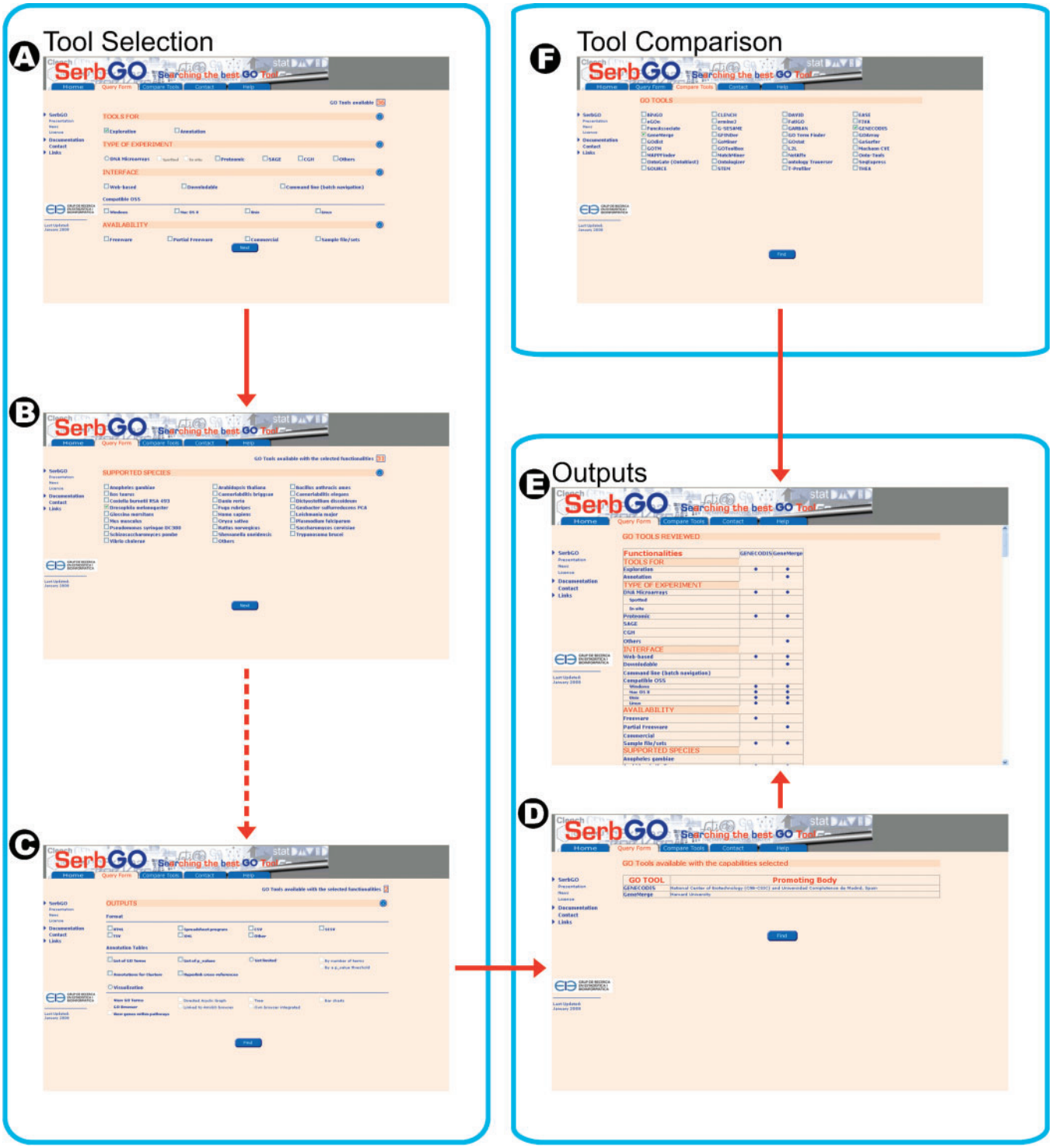

Figure 1. SerbGO workflow. (**A**) First page of the *Query Form* shows the *Standard functionalities* for the following sections: TOOL FOR, TYPE OF EXPERIMENT, INTERFACE and AVAILABILITY. (**B**) After the first validation a user selects the SUPPORTED SPECIES required and follows with the query until the last page. On the top-right corner is shown the number of tools available. (**C**) By clicking on the 'Find' button at the bottom of the page, the programs that fit the capabilities selected will be shown. (**D**) This screenshot shows the output for a list of tools and their developers. They can be compared if a user clicks on the 'Find' button. (**E**) A cross-tabulation for functionalities available in each tool is shown when the researcher requires a comparison of them. It can be attained either by comparing the output list of a *Query Form* or by selecting a set of tools at the *Compare Tools* form. (**F**) This page shows the entire collection of tools included in SerbGO. These programs can be compared by selecting the desired ones which query has to be validated on the button displayed at the bottom of the page.

Example

To illustrate the concept of how to determine which GO tools for gene-expression analysis provide the features required by a potential user the following example can be considered.

A potential SerbGO user has a list of *Drosophila melanogaster* genes. He/she would like to know which tools are available to (i) do a GO enrichment analysis (ii) that allow FlyBase Ids and (iii) correction for multiple testing for hypergeometric distribution tests. In such a situation, the user should click on the *Query Form* menu option and selects 'Exploration' at the TOOLS FOR section (Figure 1A). After that, move on the next page and selects 'Drosophila melanogaster' option (Figure 1B). When validation is made, there are 19 tools available. In DATA section, the user checks 'FlyBase ID' identifiers. He/she has to follow until the STATISTICAL ANALYSIS section, where will select 'Enrichment of GO Terms', 'Hypergeometric' test and 'Correction for Multiple Tests'. When the user gets the last query page, after clicking on the Find button the outputs are shown (Figure 1C). The researcher can see that there are two tools implementing the capabilities desired: GENECODIS and GeneMerge (Figure 1D). Now, if he/she wishes to compare the tools, it can be done by simply clicking on the new 'Find' button. This comparison will show a cross-tabulation of the capabilities available in GENECODIS and GeneMerge (Figure 1E).

IMPLEMENTATION AND AVAILABILITY

SerbGO is a web tool developed in PHP 4.3.3 on Windows using the ADOdb Database Abstraction Library for PHP and the Javascript language increased interactivity. It runs accurately on Mozilla Firefox, Internet Explorer and Konqueror browsers.

The information about tools and their functionalities has been stored in a database implemented in the open source relational database management system MySQL.

SerbGO is freely available under a Common Creative license and does not require a login. It can be accessed directly at our server (http://estbioinfo.stat.ub.es/apli/serbgo). The tool was submitted to the GO Consortium and is also available at their site (http://www.geneontology.org/GO.tools.microarray.shtml#serbgo).

BENCHMARK

SerbGO has been running since June 2006. During the testing period, most of the tools available at the GO Consortium website were included in the beta version. This version was used by several people outside the authors. SerbGO was also tested by the developers of some of the tools such as FatiGO or GARBAN who suggested some improvements that were incorporated into the testing version and validated at the first stable version.

DISCUSSION

Whether because of a lack of information about what GO tools do or because of the large number of applications available, it has long seemed reasonable for researchers to implement their own tools to 'provide' biological meaning for their experiments. This has resulted in many, and often very similar programs, which has surfaced the need for an application such as SerbGO that can be used to explore and differentiate amongst the ever-growing set of GO tools.

Thanks to the Standard Functionalities Set, a GO tool can be easily classified to determine which capabilities it implements. This greatly facilitates the task of choosing a tool that adapts to the specific interest of a user. SerbGO is intended to be used by experimental biologists without any previous training in bioinformatics. However, it should be taken into account that the best search approach is to start by checking few capabilities and in subsequent iterations gradually increase the features of interest until a satisfying list of tools is obtained. In other words, the main idea is not to check all the capabilities required at once, since this may result in a null output.

SerbGO is the only web tool to proceed in such a way and after 2 years we have observed that it is highly flexible to obtain an application or a set of applications that allow the researcher to attain their goals. In order to keep SerbGO useful, it is updated periodically (twice a year at least) and accurately. Users, especially GO tool developers, are welcome to help us implement improvements to SerbGO.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank all members of the Statistics and Bioinformatics Research Group, Department of Statistics, University of Barcelona, for their useful suggestions and comments. We are also grateful to Dr Fátima Nuńez from the IRHUVH for her suggestions. Funding to pay the Open Access publication charges for this article was provided by the project no. 303483 FBG-University of Barcelona.

Conflict of interest statement. None declared.

REFERENCES

1. Simon,R.M., Korn,E.L., McShane,L.M., Radmacher,M.D., Wright,G.W. and Zhao,Y. (2004) *Design and Analysis of DNA Microarray Investigations*. Springer-Verlag, New York.
2. Sánchez-Pla,A. and Mosquera,J.L. (2008) The quest for biological significance. In Bonilla,L.L., Moscoso,M., Platero,G., Vega,J.M. (ed.), *Progress in Industrial Mathematics at ECMI 2006*. Series Mathematics in Industry. Subseries The European Consortium for Mathematics in Industry. Springer, Heidelberg, vol. 12, pp. 566–570.
3. Asburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
4. Diestel,R. (2000) *Graph Theory*. Springer-Verlag, New York.
5. Khatri,P. and Drăghici,S. (2005) Ontological analysis of gene expression data: current tools, limitations, and problems. *Bioinformatics*, **18**, 3587–3595.