Published online 15 April 2016 *Nucleic Acids Research, 2016, Vol. 44, Web Server issue* ***W201–W204***
doi: 10.1093/nar/gkw253

MBROLE 2.0—functional enrichment of chemical compounds

Javier López-Ibáñez, Florencio Pazos and Mónica Chagoyen*

Computational Systems Biology Group, National Centre for Biotechnology (CNB-CSIC), Madrid 28049, Spain

Received January 26, 2016; Revised March 21, 2016; Accepted April 02, 2016

ABSTRACT

Metabolites Biological Role (MBROLE) is a server that performs functional enrichment analysis of a list of chemical compounds derived from a metabolomics experiment, which allows this list to be interpreted in biological terms. Since its release in 2011, MBROLE has been used by different groups worldwide to analyse metabolomics experiments from a variety of organisms. Here we present the latest version of the system, MBROLE2, accessible at http://csbg.cnb.csic.es/mbrole2. MBROLE2 has been supplemented with 10 databases not available in the previous version, which allow analysis over a larger, richer set of vocabularies including metabolite–protein and drug–protein interactions. This new version performs automatic conversion of compound identifiers from different databases, thus simplifying usage. In addition, the user interface has been redesigned to generate an interactive, more intuitive representation of the results.

INTRODUCTION

The final step in the analysis of high-throughput 'omics' experiments is typically a functional interpretation of the results, for which several bioinformatics methods have been developed. Functional enrichment, one of the most commonly used approaches for the interpretation of transcriptomics and proteomics experiments (1), can be examined with the aid of several online tools (2). These techniques analyse the functional annotations assigned to the set of genes or proteins obtained in the experiment (e.g. over- or under-expressed). Comparison of the frequency of these annotations with those of a background set allows extraction of annotations statistically enriched in the genes/protein being studied, which are used to interpret the experimental outcome in biological terms.

Recent technological advances allow analysis of the small chemical compound repertories in biological samples, providing the basis of the field known as metabolomics (3). As with other 'omics' techniques, there is a need to interpret the results of these experiments in functional terms. The goal of functional analysis in metabolomics is to transform a long list of metabolites identified in an experiment into a reduced set of meaningful biological terms that represent, for example, the biological pathways affected (4). A few years ago, we developed MBROLE, a web-based tool for the interpretation of metabolomics experiments by functional enrichment analysis (5). Most existing tools at that time focused on the functional analysis of genes/proteins; in contrast, MBROLE and a few others focused on the analysis of chemical compounds (6–8). The first version of MBROLE contained functional (biological) data on compounds from several public databases, including the Kyoto Encyclopedia of Genes and Genomes (KEGG) (9), the Human Metabolome Database (HMDB) (10), the Chemical Entities of Biological Interest (ChEBI) database (11) and the PubChem database (pubchem.ncbi.nlm.nih.gov). MBROLE has been used to analyse metabolomics experiments carried out in organisms such as human (12), rat (13), *Escherichia coli* (14), *Thermobifida fusca* (15), *Synechococcus elongates* (16), *Phoenix canariensis* (17) or *Cordyceps bassiana* (18).

Here, we present a new version of MBROLE whose main improvements include new metabolite functional annotations, inclusion of compound interactions with emphasis on metabolite–protein and drug–protein interactions, new supported compound identifiers, automatic conversion of identifiers, an improved user interface with a more intuitive presentation of the results and new annotation reports.

MBROLE2 ANALYSIS

MBROLE2 requires users to (i) provide a list of identifiers (IDs) corresponding to a set of chemical compounds (i.e. those found in the experiment), (ii) select the type of annotations they want to analyse and (iii) select a desired background set (e.g. an organism) (Figure 1A). Several input IDs are supported, and a 'Conversion' utility also accepts chemical names, CAS-Registry numbers and PubChem compounds.

MBROLE2 outputs an interactive table with the enriched functional annotations and their statistical significance in terms of *P*-value and false discovery rate (FDR). *P*-values are calculated with the cumulative hypergeometric distri-

*To whom correspondence should be addressed. Tel: +34 915 854 676; Fax: +34 915 854 606; Email: monica.chagoyen@cnb.csic.es

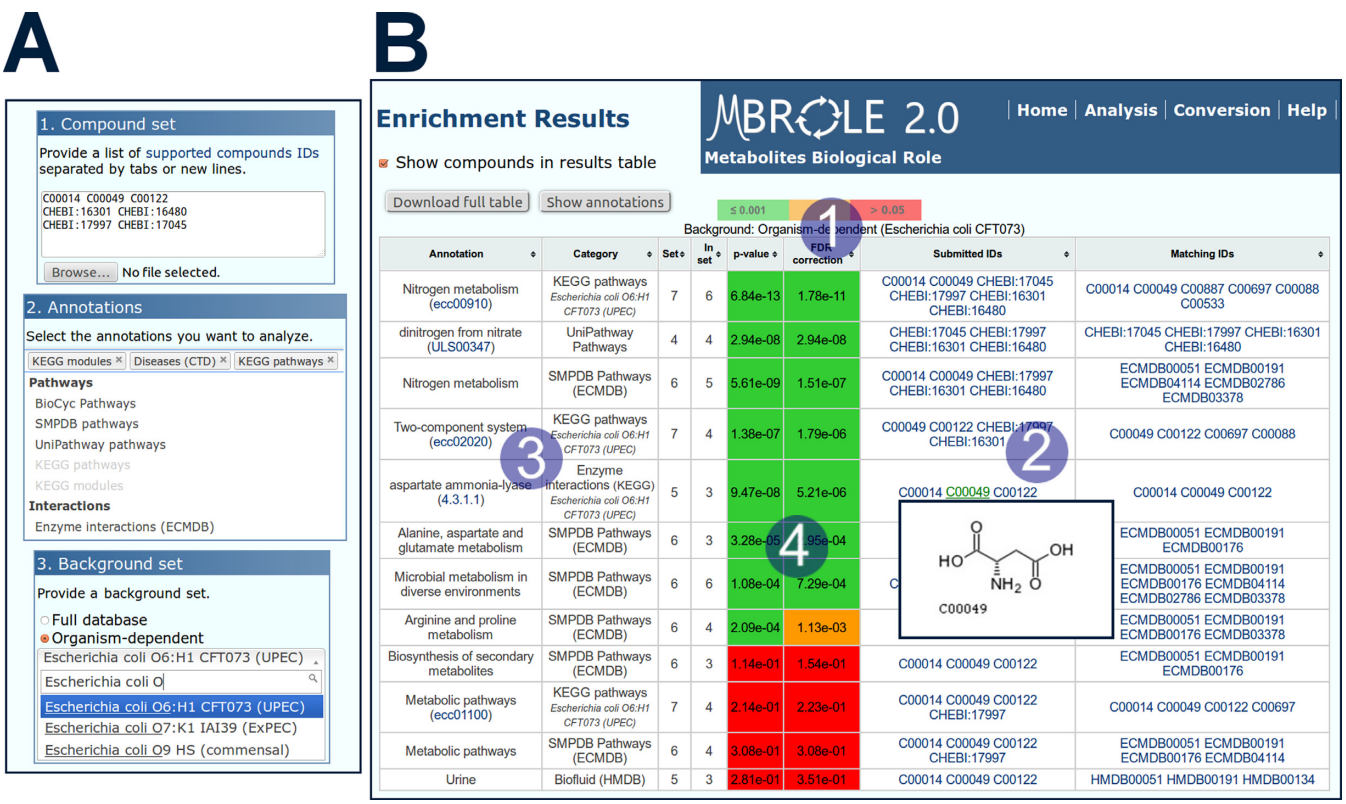

Figure 1. MBROLE2 input and output from enrichment analysis. (**A**) Users input a list of chemical compound IDs, select annotations to analyse and a background set. (**B**) Enriched annotations are shown in a table. This table (1) can be ordered by any column, (2) can show structural representations of compounds, (3) links out to external databases and (4) colours statistics to highlight significant values.

bution by comparing the number of compounds in the set and in the background with a given annotation. Resulting *P*-values are adjusted for multiple testing using the 'FDR' method (19). Users can also retrieve a table with all functional annotations for their input set of metabolites (not only those enriched); MBROLE2 can thus be used as a database of integrated metabolite annotations otherwise disseminated over several resources. In all results tables, link-outs to external resources (for both compounds and annotations) are provided to facilitate navigation and access to additional information. All results can be exported to a spreadsheet program.

Metabolite annotations

MBROLE2 allows analysis of functional annotations of chemical compounds compiled from several public databases. In addition to the databases available in the first release (KEGG, HMDB, ChEBI and PubChem), we have included data from 10 new resources comprising metabolome databases such as the Yeast Metabolome Database (YMDB) (20) and the *E. coli* Metabolome Database (ECMDB) (21), pathway and metabolic reaction databases such as the BioCyc Database Collection (22), Rhea (23) and UniPathway (24), and resources centred on specific classes of compounds including lipids, such as the LIPID MAPS Structure Database (LMSD) (25), and drugs, such as the Comparative Toxicogenomics Database (CTD) (26), Medical Subject Headings (MeSH) (27), the DrugBank database (28), and the Manually Annotated Targets and Drugs Online Resource (MATADOR) (29).

With the addition of these databases, MBROLE2 enables analysis of a wide variety of functional annotations that describe many different aspects of the chemistry and biology of chemical compounds (see Table 1); these include pathways and sub-pathways, interactions with enzymes, proteins and other types of molecules, physiological locations, chemical classifications and taxonomies, and biological roles, uses and applications (e.g. drug indications).

Automatic ID conversion

As input, the previous version of MBROLE required a list of chemical identifiers from the same database for which the annotations were to be analysed (e.g. a list of KEGG compounds for the analysis of KEGG pathways or a list of ChEBI accessions for analysis of ChEBI roles). Users therefore had to convert the IDs (using the MBROLE conversion utility) if they were not supported by the chosen database.

In MBROLE2, we have automated the conversion of chemical IDs, such that the user provides a single ID list, which are automatically converted to those of the appropriate database, depending on the annotations chosen for analysis. This greatly simplifies input and allows users to provide a mixture of chemical IDs in the same list. Both user IDs and converted IDs are shown in the results table.

To ensure unambiguous conversion, only chemical IDs are supported as input for analysis (and not metabo-

Table 1. Functional annotations that can be analysed with MBROLE2

Pathways and sub-pathways	BioCyc, SMPDB, UniPathway, KEGG pathways, KEGG modules
Interactions	With enzymes (ECMDB/YMDB/HMDB, KEGG, BioCyc) With proteins (BioCyc, MATADOR, HMDB/ECMDB/YMDB) Other molecules: bioactive peptides, lipids, etc. (KEGG)
Physiological location	Cellular (HMDB/ECMDB/YMDB) Tissue (HMDB) Biofluid (HMDB)
Chemical classification and taxonomies	HMDB/YMDB taxonomy, KEGG classifications, LIPID MAPS taxonomy
Roles, uses and applications	Biochemical, chemical and application (ChEBI) Biofunctions (HMDB) Roles (KEGG) Drug indications (DrugBank), pharmacological actions (MeSH), Anatomical Therapeutic Chemical (ATC) classification system (for drugs) Diseases (HMDB and CTD)

lite names, for example). Supported input IDs in MBROLE2 include those in the first version (i.e. KEGG compounds, HMDB metabolites, PubChem compounds and ChEBI accessions) plus the new IDs: YMDB and ECMDB metabolites, LIPID MAPS lipids and METLIN metabolite database codes (30). The MBROLE2 conversion utility also accepts chemical names, CAS Registry Numbers and PubChem compound IDs.

Repeated IDs are considered as a single ID in the analysis. Compound IDs with no annotations are not considered in the statistics. Input IDs with more than one correspondence in the automated conversion are considered multiple IDs in the statistics of those annotations (as many as the IDs converted).

Improved interface

Input. Through an intuitive interface, users can now choose which types of annotations to analyse. Annotations are organized into a hierarchical classification so that individual types (e.g. 'BioCyc pathways') or entire groups (e.g. 'Pathways') can be selected/deselected in the input form (Figure 1A).

Annotation search. For those users interested in retrieving a table of functional annotations (without performing enrichment analysis), we have added an 'Annotation search' option to MBROLE2. Those who have already performed an enrichment analysis can access the 'Annotation search' results directly from the enrichment output page.

Output. A single table is provided with the enriched annotations in all the vocabularies selected. Only annotations found for at least three compounds in the input set are included (Figure 1B). Users can sort the table interactively by any column. For example, to focus on annotations for a given vocabulary, the table can be sorted by the 'Category' column; to retrieve the most significant annotations regardless of the vocabulary they belong to, the sort can be by *P*-value or FDR columns, etc. To spot significant annotations conveniently, cells containing *P*-values and FDRs $\leq$ 0.001 are coloured green, yellow (for values between 0.001 and 0.05) and red ($>$0.05). An option is also included to export this results' table to a spreadsheet program, so the user can perform additional calculations, post-processing and graphical representations.

ACKNOWLEDGEMENTS

We thank the members of the Computational Systems Biology Group (CNB-CSIC). Special thanks to David San León (CNB-CSIC) and the members of the Foodomics Laboratory (CIAL-CSIC) for interesting ideas and discussions and the members of the Scripps Center for Metabolomics for providing METLIN ID conversions. We thank C Mark for editorial assistance.

FUNDING

Spanish Ministry of Economy and Competitiveness (MINECO) and European Social Fund Pre-doctoral Contract (to J.L.I.); MINECO [BIO2010-22109].
Conflict of interest statement. None declared.

REFERENCES

1. Khatri,P. and Draghici,S. (2005) Ontological analysis of gene expression data: current tools, limitations, and open problems. *Bioinformatics*, **21**, 3587–3595.
2. Huang da,W., Sherman,B.T. and Lempicki,R.A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.*, **37**, 1–13.
3. Patti,G.J., Yanes,O. and Siuzdak,G. (2012) Innovation: Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.*, **13**, 263–269.
4. Chagoyen,M. and Pazos,F. (2013) Tools for the functional interpretation of metabolomic experiments. *Brief. Bioinform.*, **14**, 737–744.
5. Chagoyen,M. and Pazos,F. (2011) MBRole: enrichment analysis of metabolomic data. *Bioinformatics*, **27**, 730–731.
6. Xia,J. and Wishart,D.S. (2010) MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. *Nucleic Acids Res.*, **38**, W71–W77.
7. Aggio,R.B., Ruggiero,K. and Villas-Boas,S.G. (2010) Pathway Activity Profiling (PAPi): from the metabolite profile to the metabolic pathway activity. *Bioinformatics*, **26**, 2969–2976.
8. Kankainen,M., Gopalacharyulu,P., Holm,L. and Oresic,M. (2011) MPEA–metabolite pathway enrichment analysis. *Bioinformatics*, **27**, 1878–1879.
9. Kanehisa,M., Goto,S., Sato,Y., Kawashima,M., Furumichi,M. and Tanabe,M. (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.*, **42**, D199–D205.
10. Wishart,D.S., Jewison,T., Guo,A.C., Wilson,M., Knox,C., Liu,Y., Djoumbou,Y., Mandal,R., Aziat,F., Dong,E. *et al.* (2013) HMDB 3.0–The Human Metabolome Database in 2013. *Nucleic Acids Res.*, **41**, D801–D807.
11. Hastings,J., de Matos,P., Dekker,A., Ennis,M., Harsha,B., Kale,N., Muthukrishnan,V., Owen,G., Turner,S., Williams,M. *et al.* (2013) The ChEBI reference database and ontology for biologically relevant

chemistry: enhancements for 2013. *Nucleic Acids Res.*, **41**, D456–D463.
12. Garcia-Canaveras,J.C., Donato,M.T., Castell,J.V. and Lahoz,A. (2011) A comprehensive untargeted metabonomic analysis of human steatotic liver tissue by RP and HILIC chromatography coupled to mass spectrometry reveals important metabolic alterations. *J. Proteome Res.*, **10**, 4825–4834.
13. Qi,Y., Gu,H., Song,Y., Dong,X., Liu,A., Lou,Z., Fan,G. and Chai,Y. (2013) Metabolomics study of resina draconis on myocardial ischemia rats using ultraperformance liquid chromatography/quadrupole time-of-flight mass spectrometry combined with pattern recognition methods and metabolic pathway analysis. *Evid. Based Complement Alternat. Med*, 438680.
14. Carneiro,S., Villas-Boas,S.G., Ferreira,E.C. and Rocha,I. (2012) Influence of the RelA activity on E. coli metabolism by metabolite profiling of glucose-limited chemostat cultures. *Metabolites*, **2**, 717–732.
15. Vanee,N., Brooks,J.P., Spicer,V., Shamshurin,D., Krokhin,O., Wilkins,J.A., Deng,Y. and Fong,S.S. (2014) Proteomics-based metabolic modeling and characterization of the cellulolytic bacterium Thermobifida fusca. *BMC Syst. Biol.*, **8**, 86.
16. Diamond,S., Jun,D., Rubin,B.E. and Golden,S.S. (2015) The circadian oscillator in Synechococcus elongatus controls metabolite partitioning during diurnal growth. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E1916–E1925.
17. Giovino,A., Martinelli,F. and Saia,S. (2015) Rhynchophorus ferrugineus attack affects a group of compounds rather than rearranging Phoenix canariensis metabolic pathways. *J. Integr. Plant Biol.*, **58**, 388–396.
18. Hyun,S.H., Lee,S.Y., Sung,G.H., Kim,S.H. and Choi,H.K. (2013) Metabolic profiles and free radical scavenging activity of Cordyceps bassiana fruiting bodies according to developmental stage. *PLoS One*, **8**, e73065.
19. Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B*, **57**, 289–300.
20. Jewison,T., Knox,C., Neveu,V., Djoumbou,Y., Guo,A.C., Lee,J., Liu,P., Mandal,R., Krishnamurthy,R., Sinelnikov,I. *et al.* (2012) YMDB: the yeast metabolome database. *Nucleic Acids Res.*, **40**, D815–D820.
21. Guo,A.C., Jewison,T., Wilson,M., Liu,Y., Knox,C., Djoumbou,Y., Lo,P., Mandal,R., Krishnamurthy,R. and Wishart,D.S. (2013) ECMDB: the E. coli Metabolome Database. *Nucleic Acids Res.*, **41**, D625–D630.
22. Caspi,R., Billington,R., Ferrer,L., Foerster,H., Fulcher,C.A., Keseler,I.M., Kothari,A., Krummenacker,M., Latendresse,M., Mueller,L.A. *et al.* (2016) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.*, **44**, D471–D480.
23. Morgat,A., Axelsen,K.B., Lombardot,T., Alcantara,R., Aimo,L., Zerara,M., Niknejad,A., Belda,E., Hyka-Nouspikel,N., Coudert,E. *et al.* (2015) Updates in Rhea–a manually curated resource of biochemical reactions. *Nucleic Acids Res.*, **43**, D459–D464.
24. Morgat,A., Coissac,E., Coudert,E., Axelsen,K.B., Keller,G., Bairoch,A., Bridge,A., Bougueleret,L., Xenarios,I. and Viari,A. (2012) UniPathway: a resource for the exploration and annotation of metabolic pathways. *Nucleic Acids Res.*, **40**, D761–D769.
25. Sud,M., Fahy,E., Cotter,D., Brown,A., Dennis,E.A., Glass,C.K., Merrill,A.H. Jr, Murphy,R.C., Raetz,C.R., Russell,D.W. *et al.* (2007) LMSD: LIPID MAPS structure database. *Nucleic Acids Res.*, **35**, D527–D532.
26. Davis,A.P., Grondin,C.J., Lennon-Hopkins,K., Saraceni-Richards,C., Sciaky,D., King,B.L., Wiegers,T.C. and Mattingly,C.J. (2015) The Comparative Toxicogenomics Database's 10th year anniversary: update 2015. *Nucleic Acids Res.*, **43**, D914–D920.
27. Lipscomb,C.E. (2000) Medical Subject Headings (MeSH). *Bull. Med. Libr. Assoc.*, **88**, 265–266.
28. Wishart,D.S., Knox,C., Guo,A.C., Shrivastava,S., Hassanali,M., Stothard,P., Chang,Z. and Woolsey,J. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.*, **34**, D668–D672.
29. Gunther,S., Kuhn,M., Dunkel,M., Campillos,M., Senger,C., Petsalaki,E., Ahmed,J., Urdiales,E.G., Gewiess,A., Jensen,L.J. *et al.* (2008) SuperTarget and Matador: resources for exploring drug-target relationships. *Nucleic Acids Res.*, **36**, D919–D922.
30. Smith,C.A., O'Maille,G., Want,E.J., Qin,C., Trauger,S.A., Brandon,T.R., Custodio,D.E., Abagyan,R. and Siuzdak,G. (2005) METLIN: a metabolite mass spectral database. *Ther. Drug. Monit.*, **27**, 747–751.