*Bioinformatics*, YYYY, 0–0

doi: 10.1093/bioinformatics/xxxxx

Advance Access Publication Date: DD Month YYYY

Manuscript Category

|  |
| --- |
| Systems Biology  metID: A R package for automatable compound annotation for LC-MS-based data  Corresponding Author1,\*, Co-author2 and Co-Author2  1Department of XXXXXXX, Address XXXX etc., 2Department of XXXXXXX, Address XXXX etc.  \*To whom correspondence should be addressed.  Associate Editor: XXXXXXX  Received on XXXXX; revised on XXXXX; accepted on XXXXX  Abstract  **Motivation:** The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.  **Results:** The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.  **Availability:** The quick brown fox jumps over the lazy dog.  **Contact:** example@example.org  **Supplementary information:** Supplementary data are available at *Bioinformatics* online. |

# Introduction

Liquid chromatography coupled to mass spectrometry (LC-MS) is a comprehensive, no-biased technology to research small compounds, which has become more and more popular in food, environment, and biomedical studies(Wishart, 2016)(Fraga-Corral et al., 2020). However, accurate and efficient compound annotation serves as a vital hub prior to data interpretation and further biological hypothesis generation. To present compound annotation in a standardized way, the metabolomics community proposed the annotation levels from 1-4 (Sumner et al., 2007), where the criteria mainly contains mass to charge ratio (m/z), retention time (RT) and MS2 spectra match, which are gold and generally accept standards for compound annotation. Several databases have been made for the public [ref]. Based on these resources, some tools (Chaleckis et al., 2019) were developed for compound annotation. However, the limitations of most of these tools are as follows: (1) these tools usually only use m/z and/or MS2 spectra information, but lack retention time. (2) No single tool so far has combined all possible databases including in-house and public databases in a streamlined fashion; (3) No command line-based tool to allow in-house database construction and automatic compound annotation, which is important to deploy it on cluster servers to take advantage of great computing power.

In this context, we introduce a new R package, “metID”, particularly designed for (1) construct users’ own in-house database and (2) automatable and simple compound annotation. As annotation level 1, metID also provides our in-house databases containing more than 1,000 standards by manual curating (users need use the same LC gradient to use RT). For level 2, metID now provides 6 public MS2 databases (~39,015 compounds and 59,289 spectra) and 6 MS1 databases. metID integrates all these databases and generates annotation scoring metrics and confidence levels for users to assess annotation quality for environment, food/drug and biomedical studies.

# Features and methods

Using metID, users can easily build in-house databases using the standards that were acquired in their own laboratories. The public databases can also be easily organized as the database format for metID. The in-house databases in our laboratory and several public databases are provided. metID can be installed on all platforms (Mac OS, Windows and Linux).

**2.1 Database construction for metID**

If users have in-house standards which have been acquired with MS2 spectra, it is possible to build the in-house database using the “construct\_database()” function (Fig. 1a). Three items are contained in the database: (1) database information, (2) standard information and (3) MS2 spectra for each compound.

**2.2 Compound identification**

RT may shift in different acquisition batches. Therefore, it is necessary to correct the RTs in databases using the “correct\_database\_rt()” function if you spike internal standards into standards and biological samples. Then metID can annotate compounds with different levels according to databases (in-house databases, level 1; public MS2 database, level 2; MS1 databases, level 3). For m/z (ppm), RT (second) and MS2 spectra match (cosine score) scores, they are combined as one total score and scaled to 0-1 (Fig. 1b).

# Results

We applied metID on a public study (Contrepois et al., 2020). The authors reported X annotated metabolites. Using metID, we successfully annotated X compounds (level 1/2). X% of annotations overlapped. Besides these overlapping annotations, metID is able to annotate X new compounds that were not annotated in the original publication. These results indicate that metID is a valid automatable approach for compound annotation. (see Supplementary Material).

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

**Table 1.**Benchmark results of the cascade oscillators model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| |S| | Predicted cost | Timing | Predicted speed | Speed |
| 1 | S219.20(100%) | 68m43s | 1.00 | 1.00 |
| 2 | 29.10+219.10(~50%) | 35m13s | 2.00 | 1.95 |
| 4 | 219.20(100%) | 68m43s | 1.00 | 1.00 |
| 10 | 29.10+219.10(~50%) | 35m13s | 2.00 | 1.95 |
| 20 | 219.20(100%) | 68m43s | 1.00 | 9.5 |

This is table foot note sample text This is table foot note sample text This is table foot note sample text

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.The quick brown fox jumps over the lazy dog.

Acknowledgements

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

Funding

This work has been supported by the …..

*Conflict of Interest:* none declared.

References

Alexandrescu,A. (2001) Modern C++ Design: Generic Programming and Design Patterens Applied. Addision Wesley Professional, Boston.

Dormand,J.R. and Prince,P.J. (1980) A family of embedded Runge–Kutta formulae. *J. Comp. Appl. Math.*, **6**, 19–26.

Alexandrescu,A. (2001) *Modern C++ Design: Generic Programming and Design Patterens Applied.* Addision Wesley Professional, Boston.

Dormand,J.R. and Prince,P.J. (1980) A family of embedded Runge–Kutta formulae. *J. Comp. Appl. Math.*, **6**, 19–26.

Alexandrescu,A. (2001) *Modern C++ Design: Generic Programming and Design Patterens Applied.* Addision Wesley Professional, Boston.

Dormand,J.R. and Prince,P.J. (1980) A family of embedded Runge–Kutta formulae. *J. Comp. Appl. Math.*, **6**, 19–26.

Yoo,M.S. *et al*. (2003) Oxidative stress regulated genes in nigral dopaminergic neurnol cell: correlation with the known pathology in Parkinson’s disease. *Brain Res. Mol. Brain Res.*, **110**(Suppl. 1), 76–84.

Crenshaw, B.,III, and Jones, W.B.,Jr (2003) The future of clinical cancer management: one tumor, one chip. *Bioinformatics*, doi:10.1093/bioinformatics/btn000.

Auhtor,A.B. *et al.* (2000) Chapter title. In Smith, A.C. (ed.), *Book Title*, 2nd edn. Publisher, Location, Vol. 1, pp. ???–???.

Bardet, G. (1920) Sur un syndrome d’obesite infantile avec polydactylie et retinite pigmentaire (contribution a l’etude des formes cliniques de l’obesite hypophysaire). PhD Thesis, name of institution, Paris, France.