Published online 2 November 2017 *Nucleic Acids Research, 2018, Vol. 46, Database issue* ***D1261–D1265***
doi: 10.1093/nar/gkx1016

mVOC 2.0: a database of microbial volatiles

Marie C. Lemfack[1,†], Bjoern-Oliver Gohlke[2,3,†], Serge M. T. Toguem[1], Saskia Preissner[2,3], Birgit Piechulla[1] and Robert Preissner[2,3,*]

[1]University of Rostock, Institute for Biological Sciences, Albert-Einstein-Strasse 3, 18059 Rostock, Germany, [2]Structural Bioinformatics Group, Institute of Physiology & Experimental and Clinical Research Center (ECRC), Charité—University Medicine Berlin, Philippstr. 12 / Lindenberger Weg 80, 10115 / 13125 Berlin, Germany and [3]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg 69120, Germany

Received September 15, 2017; Editorial Decision October 17, 2017; Accepted October 18, 2017

ABSTRACT

Metabolic capabilities of microorganisms include the production of secondary metabolites (e.g. antibiotics). The analysis of microbial volatile organic compounds (mVOCs) is an emerging research field with huge impact on medical, agricultural and biotechnical applied and basic science. The mVOC database (v1) has grown with microbiome research and integrated species information with data on emitted volatiles. Here, we present the mVOC 2.0 database with about 2000 compounds from almost 1000 species and new features to work with the database. The extended collection of compounds was augmented with data regarding mVOC-mediated effects on plants, fungi, bacteria and (in-)vertebrates. The mVOC database 2.0 now features a mass spectrum finder, which allows a quick mass spectrum comparison for compound identification and the generation of species-specific VOC signatures. Automatic updates, useful links and search for mVOC literature are also included. The mVOC database aggregates and refines available information regarding microbial volatiles, with the ultimate aim to provide a comprehensive and informative platform for scientists working in this research field. To address this need, we maintain a publicly available mVOC database at: http://bioinformatics.charite.de/mvoc.

INTRODUCTION

Microorganisms appear in almost all habitats on earth. Survival and growth is at least partially due to the extreme flexibility and adaptability of microbial metabolism. It is well known that microorganisms produce a wealth of secondary metabolites under nutrient limitation and/or in competition to other organisms. Macromolecules including antibiotics, toxins, siderophores and communication signals are common compounds released to structure communities and populations in respective habitats. Besides these diffusible substances it became apparent that microorganisms also release a plethora of small metabolites, with low molecular mass, often with low boiling points and high vapor pressures (mVOCs) (1,2). Such volatiles are predestined infochemicals for inter- and intraspecific communication acting in the atmosphere as well as below ground over long distances (3). These mVOCs derive from different biosynthetic pathways and are grouped into aromatic compounds, fatty acid derivatives, terpenoids, nitrogen- and sulphur-containing compounds, to name the most abundant (1). These VOCs were registered from more than 500 bacteria and fungi (4). Many more including unique, unusual and novel compounds will be identified in the future since 10^{18} microbial species are expected to exist on earth (5). This vision demonstrates the incredible potential of this emerging research field.

Despite the ubiquitous and often massive appearance of microorganisms, only limited information on biological/ecological roles and modes of actions of these compounds are available. It was shown that mVOC bouquets affect other organisms in co-cultivation assays and therefore it was concluded that they play a role in plant/fungi-microbe and animal/human-microbe interactions (6–10). However, such investigations are not always suitable to assign a defined bioactivity to a specific compound. To fill this gap of information, individual compounds or defined mixtures need to be studied in appropriate assays. At present, only <10% of discrete microbial VOCs are associated with a function (11,12). Subsequently, studies on fungal and bacterial VOCs lag behind the knowledge on plant and animal derived VOCs.

Recently, increasing attention was given to mVOCs because they are important for basic as well as applied sciences (13–15). Microbial VOCs offer chances to provide sustainable and eco-friendly alternatives to pesticides and fertilizers in the agriculture and the environment, and non-invasive diagnostic tools for animal and human health (16).

*To whom correspondence should be addressed. Tel: +49 30 450 528 495; Fax: +49 30 450 540 955; Email: robert.preissner@charite.de
†These authors contributed equally to the paper as first authors.

Furthermore, mVOCs are already used in biotechnological processes to control, for example, foodstuff, buildings and other hardware.

So far most of the chemical information on mVOCs and volatilomes appears however scattered in the literature and remains underexploited. Three years ago we initiated the attempt to summarize the volatilomes of bacteria and fungi resulting in the first mVOC database (4). Quickly, it became obvious that more information, tools, statistics and links would be useful for the community working in this research field. Here, we present mVOC 2.0, which now provides an extended collection of mVOCs, increased number of species, and improved usability (e.g. for spectra).

MATERIALS AND METHODS

Data collection

The data were acquired by an automatic text mining procedure using the PubMed-API (https://www.ncbi.nlm.nih.gov/home/develop/api/). The resulting set of publications was manually screened by experts. mVOCs were extracted and validated by a second expert before taken into this database. Through this manual procedure so far 1860 unique mVOCs could be extracted from the literature, which are obtained from 604 bacterial and 340 fungi species. The number of included mVOCs as well as species has more than doubled compared to the first version of this database.

For further extension, scientific literature will be continuously screened (via PubMed API) and checked by experts before including into the database on a monthly basis.

Data processing

To get structural information for the extracted mVOCs from the literature, the compounds were mapped to the PubChem database (17). Only exact matches were accepted in a first step. All non-matched compounds were evaluated by a biochemist, and mapped to existing compounds from the database. Additional information from NIST (18), KEGG (19) and the human metabolome database (20) was mapped to the mVOCs. Especially, volatile specific information like boiling point, vapor pressure or soil adsorption was included.

Server, database and system requirements

The mVOC database is implemented as a relational database on a MySQL server (http://www.mysql.com/). PHP and javascript have been used to build the website. As an applet for sketching compounds for the 'Structure Search' and 'Add a new mVOC' function, the open-source web-component ChemDoodle is implemented on the mVOC website. Web access is enabled by Apache HTTP Server 2. We strongly recommend the use of a recent version of Mozilla Firefox, Google Chrome or Safari browser, with JavaScript options enabled, to access the website.

DATABASE SEARCH OPTIONS

The database was extended to enable further and more specific search options for the user. It is now possible to search directly for mass spectra peaks or for emitter and receiver of specific mVOCs. Additional volatile-specific information is now available.

Search for mass spectra

As a new feature of this database, a search for the three most intensive peaks in mass spectra is included. It is also possible to search for compounds by name, PubChem-ID or their molecular formula.

Search emitter/receiver of discrete mVOCs

In co-cultivation assays the receiver organism is exposed to a complex and rather undefined mixture of emitted VOCs. The underlying bioactive compound(s) are not obvious. An up-to-date dataset of discrete/individual compounds causing effects in plants, fungi, bacteria, invertebrates and vertebrates is provided. This subset can be searched by different physico-chemical properties, species information as well as compound synonyms.

Search mVOCs

This database offers different entry points for the user. It is possible to browse the content of the database by using ('browse mVOC') button. This feature allows browsing by the first letter of the compound name as well as the chemical classification of the compounds.

To search for specific compounds, the user can choose to search by PubChem-ID, name or the molecular formula. Furthermore, an option to search for physico-chemical properties like the molecular weight and logP value is provided. In addition, the search for chemical classification and certain species within the kingdoms bacteria and fungi is supported.

Compounds can also be searched by structural similarity ('Structure Search') including substructure search. Users can easily search by name, smiles string or by drawing the structure on their own. To provide a maximum range of compounds, the search for name and smiles is using the PubChem's PUG interface to get the structure file of the compound. Drawing of a compound can be done directly in the embedded ChemDoodle interface (http://www.chemdoodle.com) which also provides the option to upload a user-defined MOL file.

Biological interpretation

To investigate the biological function of mVOCs from the database, two features are included by the 'Features/Tools' dropdown menu. To put the compound-target relations contained in the database into a cellular context, this information is mapped onto metabolic pathways (KEGG). This feature is available via the 'KEGG Pathways' button. Here, the user has to choose the organism and pathway of interest and will get displayed all mVOCs for which compound-target relations could be figured out. This allows the user to investigate possible effects of specific compounds or can serve as starting point for further analysis of metabolism or signalling in general.

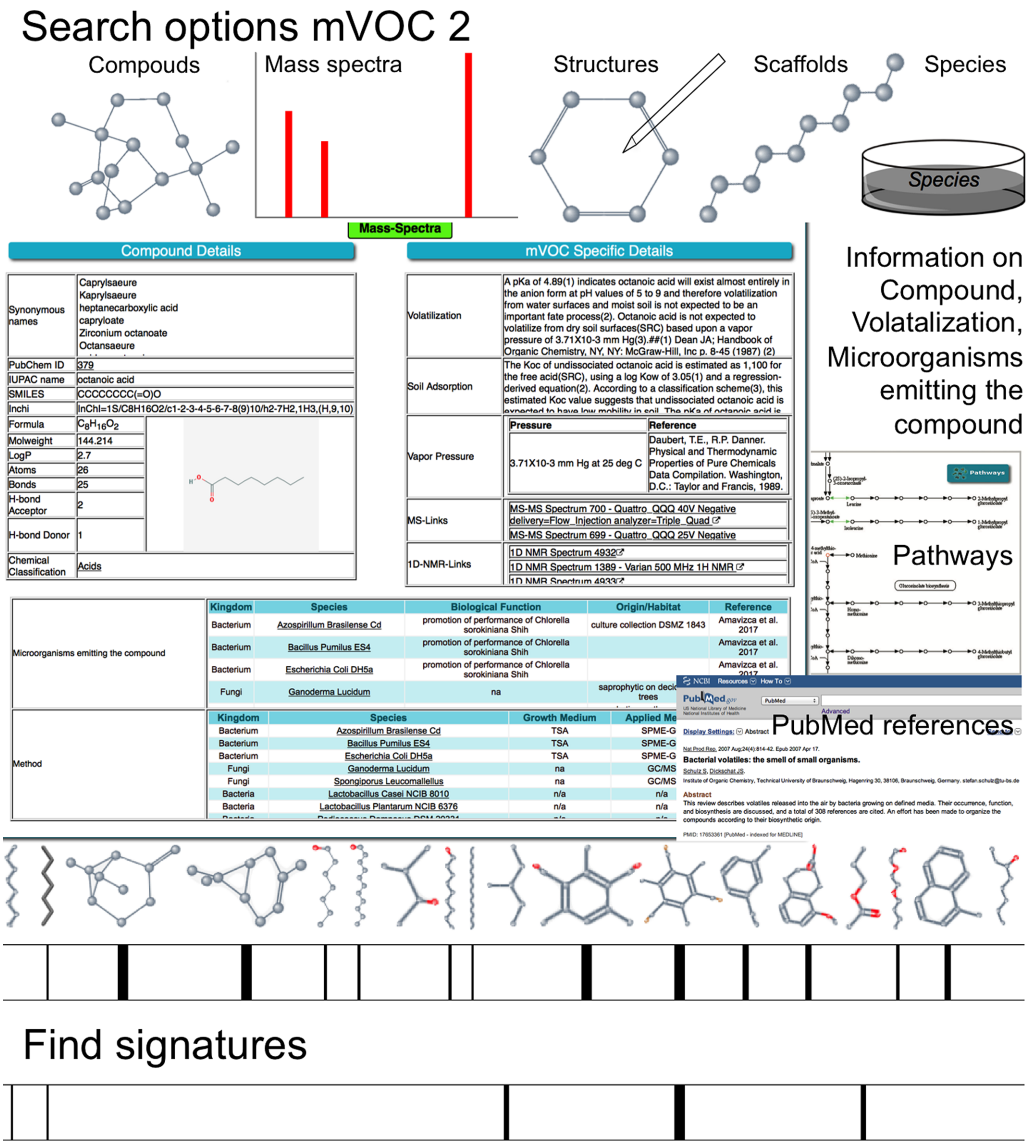

Figure 1. Functionalities of the mVOC 2.0 database. Five different search options enable users to use the database for either research on specific VOCs, chemical structures, scaffolds or mass spectra, microorganisms and fungi. Information on compounds and their volatilization, as well as pathways and species-specific signatures is provided along with relevant literature.

With the search for signature volatiles ('Signatures' button), the database can be used to show the occurrence of unique mVOCs in different species (fingerprinting). Therefore a bacterial or fungal species has to be selected, for which the emitted mVOCs will be displayed. If a VOC is only emitted by one of the listed species it will be highlighted in green. After the determination of a set of volatiles (e.g. from a mass spectrum) this allows the identification of the putative emitter species (see the use case for an example).

Search results

The resulting information is divided into three main parts. The first one consists of structural details and links to further investigate the compound. The structure is also displayed in 2D, by clicking onto the structure, a 3D-visualisation is provided, to illustrate the conformational flexibility of the compound.

Second box: new volatile-specific information like boiling point, vapor pressure, volatilization, soil adsorption as well as links to further mass spectra/nuclear magnetic resonance spectra are provided.

Mass spectra from the NIST database were directly included into the database and can be accessed by the ('mass spectra') button on the top.

The third box displays information about the VOC-emitting microorganisms, their biological function, the origin/habitat, the respective methods for extracting, analysing and confirming the compounds are complemented by the respective references.

The resulting mass spectra are included from the NIST database. The spectra shows the three most intensive peaks. For detailed information, a link to the NIST database as

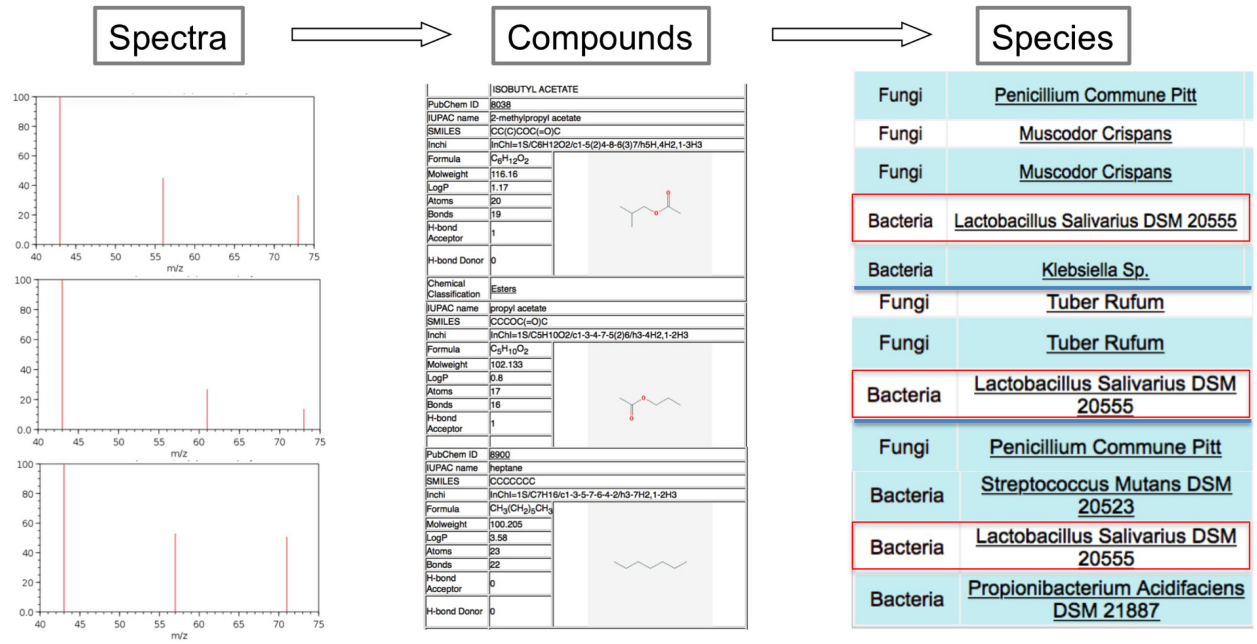

Figure 2. Use case of mVOC 2.0 database. A query for the base peak at 43 in the mass spectrum was performed. Compounds with this base peak are displayed and can be further narrowed to find possible emitter spiecies.

well as the NIST identifier are provided and offers the user the opportunity for further evaluation.

The results page for the search for emitter and receiver of discrete mVOCs comprises of a table with detailed species information. It shows the causing effects in plants, fungi, bacteria, invertebrates and vertebrates of discrete/individual compounds. Structural and volatile specific information is equally displayed as provided for the mVOC-search.

USE CASE

The mVOC 2.0 database offers different ways to search for VOCs (Figure 1). Users are enabled to look up certain compounds via name, PubChem-ID, chemical formula or properties. Chemical structures can be drawn and a similarity search will be performed or specific scaffolds can be searched for.

Another approach to find VOCs is to look up characteristic top or lower peaks in mass spectra or to find all mVOCs emitted by a certain species of microorganisms or fungi.

Each compound in the database is stored along with detailed information about the compound and mVOC specific details, such as volatilization. Pathways can be viewed if available and all references can be found at the end of the page.

With the help of the various search options, specific VOCs or signatures for e.g. microorganisms can be found, determined or evaluated. One example for a database query could be the search for a base peak of '43', (second '56' and third '73') from the upper mass spectrum in Figure 2. This search leads to all compounds with the base peak at 43 m/z, referring to isobutyl acetate (PubChem-ID: 8038). Repeating this with the peaks from the second and third spectrum leads to propyl acetate (PubChem-ID: 7997) and heptane (PubChem-ID: 8900). Merging this with the information of microorganisms emitting the three compounds reveals that VOCs emitted by *Lactobacillus salivarius* are likely to be responsible for the peaks created in the mass spectra.

CONCLUSION AND PERSPECTIVE

The feedback of users helped to identify at least three aspects to be further developed for a next version of the mVOC database: (i) 'VOC-signatures' to allow species identification from spectra; (ii) 'mass spectra analysis' for quick compound identification in mixtures and co-cultivation assays; and (iii) information on 'emitter/receiver species' to start understanding the complex interactions within a microbiome, during microbe-host interactions or in various other habitats.

While the number of microorganisms investigated so far is still small compared to the expected 10^{18} species on earth (5) it is nevertheless worth mentioning that the collected mVOC 2.0 dataset is significant because frequently occurring or pathogenic microbes are well represented. Its strength becomes clear when compared with the Scent-Base (21) or the Pherobase (22), which were established a long time ago without automatic update-procedures, while the mVOC database was only launched in 2014 and grew continuously. For plants and animals, additional databases are available, however the mVOC database is unique and presently the only one summarizing respective data on microorganisms. In the future, an integration of volatiles from all kingdoms will be useful for a better understanding of quorum sensing and putative medical applications (23).

ACKNOWLEDGEMENTS

The authors thank Andrea Mellin, Claudia Dinse and Thoralf Bernhardt for data acquisition.

FUNDING

German Cancer Consortium (DKTK), Heidelberg, Germany; German Federal Ministry of Education and Research (e:Top); Immunotox (BMBF); European Commission (SynSys); German Research Foundation DFG [Pi153/36–1]. Funding for open access charge: Charite - Universitätsmedizin Berlin.
Conflict of interest statement. None declared.

REFERENCES

1. Schulz,S. and Dickschat,J.S. (2007) Bacterial volatiles: the smell of small organisms. *Nat. Prod. Rep.*, **24**, 814–842.
2. Effmert,U., Kalderas,J., Warnke,R. and Piechulla,B. (2012) Volatile mediated interactions between bacteria and fungi in the soil. *J. Chem. Ecol.*, **38**, 665–703.
3. Wheatley,R.E. (2002) The consequences of volatile organic compound mediated bacterial and fungal interactions. *Antonie Van Leeuwenhoek*, **81**, 357–364.
4. Lemfack,M.C., Nickel,J., Dunkel,M., Preissner,R. and Piechulla,B. (2014) mVOC: a database of microbial volatiles. *Nucleic Acids Res.*, **42**, D744–D748.
5. Stolz,J.F. (2017) Gaia and her microbiome. *FEMS Microbiol. Ecol.*, **93**, fiw247.
6. Bailly,A. and Weisskopf,L. (2012) The modulating effect of bacterial volatiles on plant growth: current knowledge and future challenges. *Plant Signal. Behav.*, **7**, 79–85.
7. Schenkel,D., Lemfack,M.C., Piechulla,B. and Splivallo,R. (2015) A meta-analysis approach for assessing the diversity and specificity of belowground root and microbial volatiles. *Front. Plant Sci.*, **6**, 707.
8. Lemfack,M.C., Ravella,S.R., Lorenz,N., Kai,M., Jung,K., Schulz,S. and Piechulla,B. (2016) Novel volatiles of skin-borne bacteria inhibit the growth of Gram-positive bacteria and affect quorum-sensing controlled phenotypes of Gram-negative bacteria. *Syst. Appl. Microbiol.*, **39**, 503–515.
9. Tyc,O., Zweers,H., de Boer,W. and Garbeva,P. (2015) Volatiles in inter-specific bacterial interactions. *Front. Microbiol.*, **6**, 1412.
10. Kanchiswamy,C.N., Malnoy,M. and Maffei,M.E. (2015) Chemical diversity of microbial volatiles and their potential for plant growth and productivity. *Front. Plant Sci.*, **6**, 151.
11. Piechulla,B., Lemfack,M.C. and Kai,M. (2017) Effects of discrete bioactive microbial volatiles on plants and fungi. *Plant Cell Environ.*, **40**, 2042–2067.
12. Ledger,T., Rojas,S., Timmermann,T., Pinedo,I., Poupin,M.J., Garrido,T., Richter,P., Tamayo,J. and Donoso,R. (2016) Volatile-mediated effects predominate in paraburkholderia phytofirmans growth promotion and salt stress tolerance of Arabidopsis thaliana. *Front. Microbiol.*, **7**, 1838.
13. Hertel,M., Hartwig,S., Schutte,E., Gillissen,B., Preissner,R., Schmidt-Westhausen,A.M., Paris,S., Kastner,I. and Preissner,S. (2016) Identification of signature volatiles to discriminate Candida albicans, glabrata, krusei and tropicalis using gas chromatography and mass spectrometry. *Mycoses*, **59**, 117–126.
14. Hertel,M., Preissner,R., Gillissen,B., Schmidt-Westhausen,A.M., Paris,S. and Preissner,S. (2016) Detection of signature volatiles for cariogenic microorganisms. *Eur. J. Clin. Microbiol. Infect. Dis.*, **35**, 235–244.
15. Casarrubia,S., Sapienza,S., Fritz,H., Daghino,S., Rosenkranz,M., Schnitzler,J.P., Martin,F., Perotto,S. and Martino,E. (2016) Ecologically different fungi affect arabidopsis development: contribution of soluble and volatile compounds. *PLoS One*, **11**, e0168236.
16. Hartwig,S., Raguse,J.D., Pfitzner,D., Preissner,R., Paris,S. and Preissner,S. (2017) Volatile organic compounds in the breath of oral squamous cell carcinoma patients: a pilot study. *Otolaryngol. Head Neck Surg.*, **21**, doi:10.1177/0194599817711411.
17. Kim,S., Thiessen,P.A., Bolton,E.E., Chen,J., Fu,G., Gindulyte,A., Han,L., He,J., He,S., Shoemaker,B.A. *et al.* (2016) PubChem substance and compound databases. *Nucleic Acids Res.*, **44**, D1202–D213.
18. Babushok,V.I., Linstrom,P.J., Reed,J.J., Zenkevich,I.G., Brown,R.L., Mallard,W.G. and Stein,S.E. (2007)Development of a database of gas chromatographic retention properties of organic compounds. *J. Chromatogr A*, **20**, 414–421.
19. Kanehisa,M., Sato,Y., Kawashima,M., Furumichi,M. and Tanabe,M. (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.*, **44**, D457–D462.
20. 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.
21. Knudsen,J.T., Eriksson,R., Gershenzon,J. and Ståhl,B. (2006) Diversity and distribution of floral scent. *Botanical Rev.*, **72**, 1–120.
22. El-Sayed,A.M. (2009) The pherobase: database of insect pheromones and semiochemicals. www.pherobase.com.
23. Wynendaele,E., Verbeke,F., D'Hondt,M., Hendrix,A., Van De Wiele,C., Burvenich,C., Peremans,K., De Wever,O., Bracke,M. and De Spiegeleer,B. (2015) Crosstalk between the microbiome and cancer cells by quorum sensing peptides. *Peptides*, **64**, 40–48.