SIDERITE: Unveiling Hidden Siderophore Diversity in the Chemical Space Through Digital Exploration

Ruolin He[1#], Shaohua Gu[1,2#], Jiazheng Xu[3], Xuejian Li[4], Haoran Chen[4], Zhengying Shao[3], Fanhao Wang[1], Jiqi Shao[1], Wen-Bing Yin[5,6], Long Qian[1]*, Zhong Wei[3]*, Zhiyuan Li[1,2]*

[1]Center for Quantitative Biology, Academy for Advanced Interdisciplinary Studies, Peking University
[2]Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University
[3]Educational Ministry Engineering Center of Resource-saving fertilizers, Nanjing Agricultural University
[4]Beyond Flux Technology Co., Ltd.
[5]State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences
[6]Savaid Medical School, University of Chinese Academy of Sciences

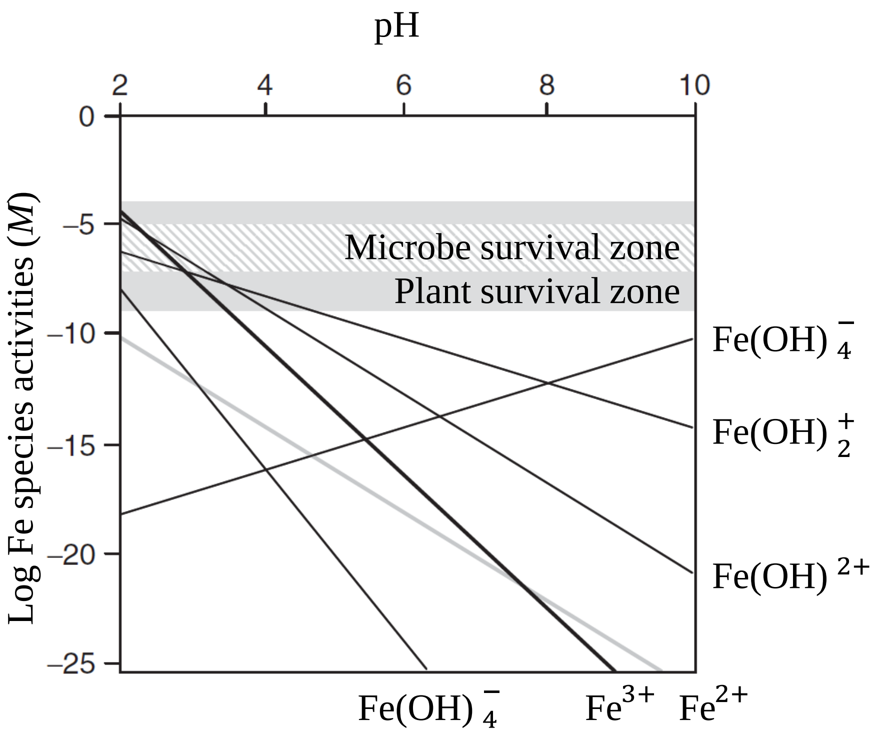
Ruolin He, Shaohua Gu, Jiazheng Xu, Xuejian Li, Haoran Chen, Zhengying Shao, Fanhao Wang, Jiqi Shao, Wen-Bing Yin, Long Qian, Zhong Wei, Zhiyuan Li. 2024. SIDERITE: Unveiling Hidden Siderophore Diversity in the Chemical Space Through Digital Exploration. ***iMeta*** 3: e192. https://doi.org/10.1002/imt2.192

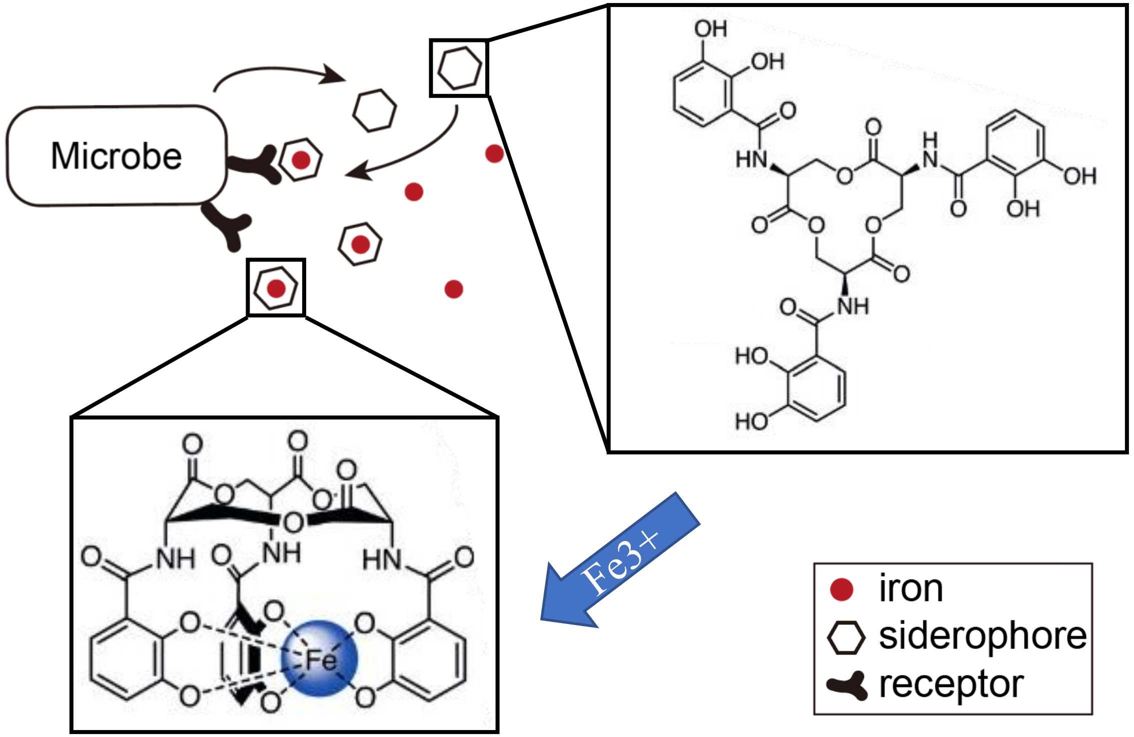
Introduction

Nature lacks the iron necessary for microbial survival.

Microbes can produce siderophores to acquire iron in the environment.

Siderophores are molecules with highly structural diversity.

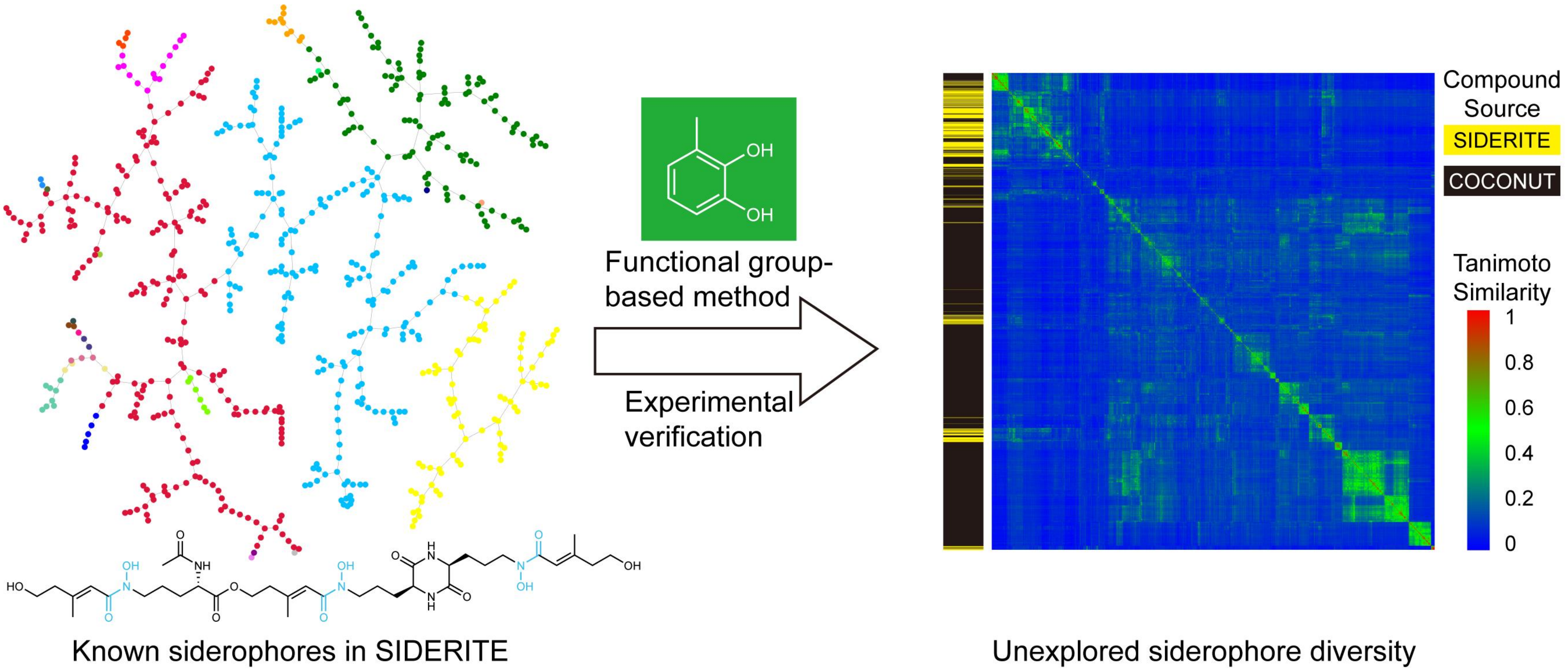




Siderophore Fe-complex

Robin, Agnès, et al. 2008 *Advances in agronomy*

Introduction



SIDERITE: Siderophore information database
COCONUT: the collection of open natural products database

Highlights

SIDERITE is the most comprehensive digitized integrated siderophore information database.

It covers all 649 known siderophores from 872 records as of May 2023. (Update at least once a year)

The confusion of siderophore was solved by merging siderophores with the same structure but different names.

Bacillibactin / Corynebactin

Siderophore ID: SID00195

Theoretical Denticity: 6

Functional Group: Catecholate (3)

Biosynthetic Type: NRPS

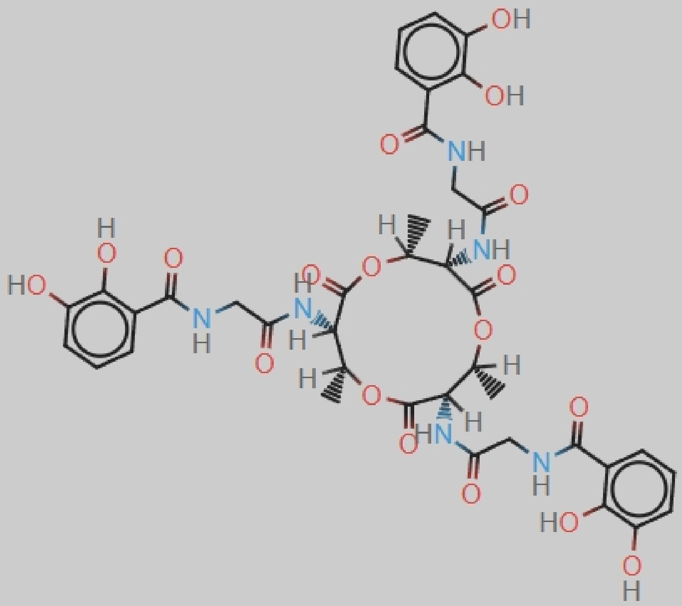
Canonical SMILES: C[C@H]1OC(=O)[C@@H](NC(=O)CNC(=O)c2cccc(O)c2O)[C@@H](C)OC(=O)[C@@H](NC(=O)CNC(=O)c2cccc(O)c2O)[C@@H](C)OC(=O)[C@H]1NC(=O)CNC(=O)c1cccc(O)c1O

Monomers Number: 9

Monomers Name: aThr/Thr, aThr/Thr, Gly, diOH-Bz, aThr/Thr, Gly, diOH-Bz, Gly, diOH-Bz

Coverage of Monomers: 1

Molecular Formula: $C_{39}H_{42}N_6O_{18}$



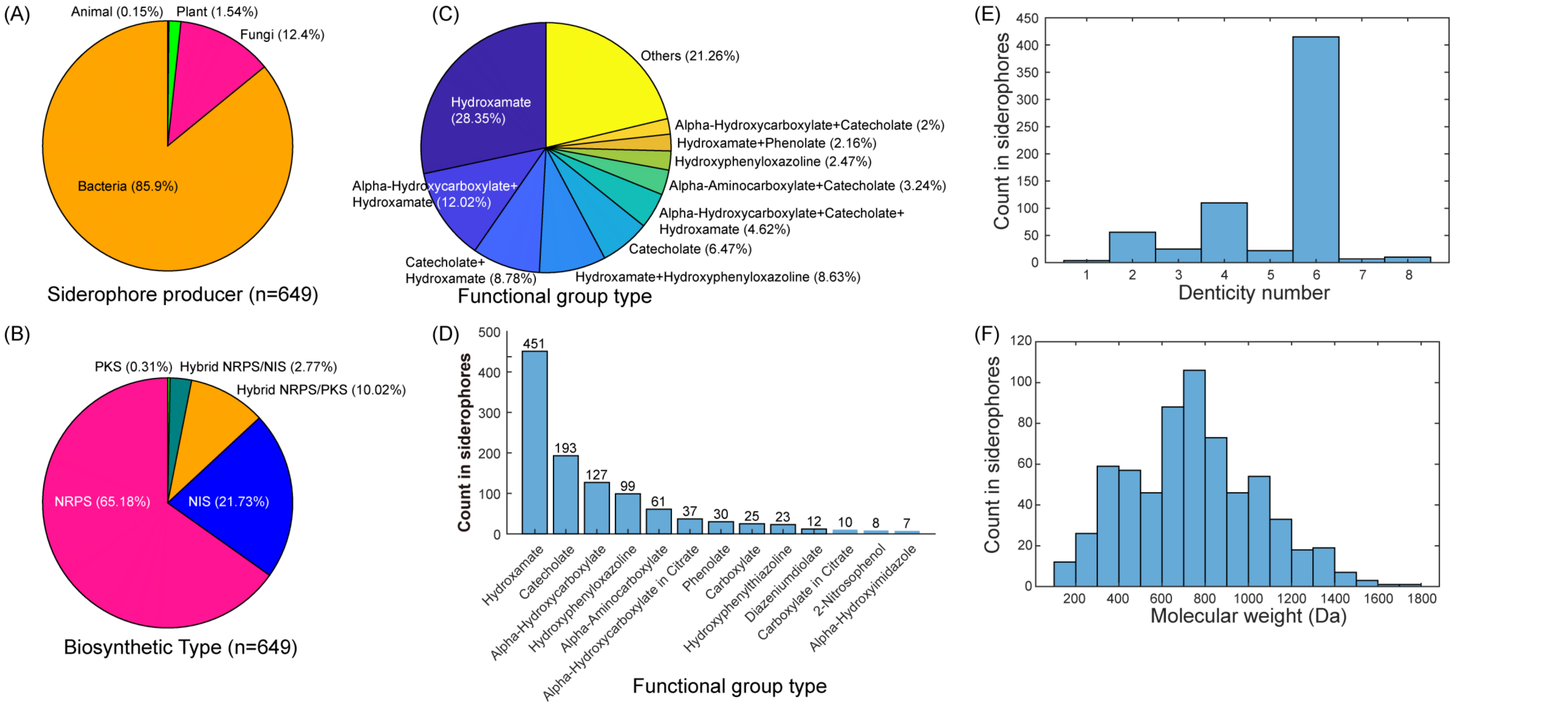
Generated by Smiles Drawer

References

1. Melissa K. Wilson, Rebecca J. Abergel, Kenneth N. Raymond, Jean E.L. Arceneaux, B. Rowe Byers, "Siderophores of Bacillus anthracis, Bacillus cereus, and Bacillus thuringiensis", Biochemical and Biophysical Research Communications,2006, 348(1), 320-325

Overview of SIDERITE

The SIDERITE consists mainly of bacterial, six-denticity, hydroxamate, NRPS siderophores.



Clustering of siderophores by structural similarities

649 siderophores are clustered into 25 clusters and 102 groups.

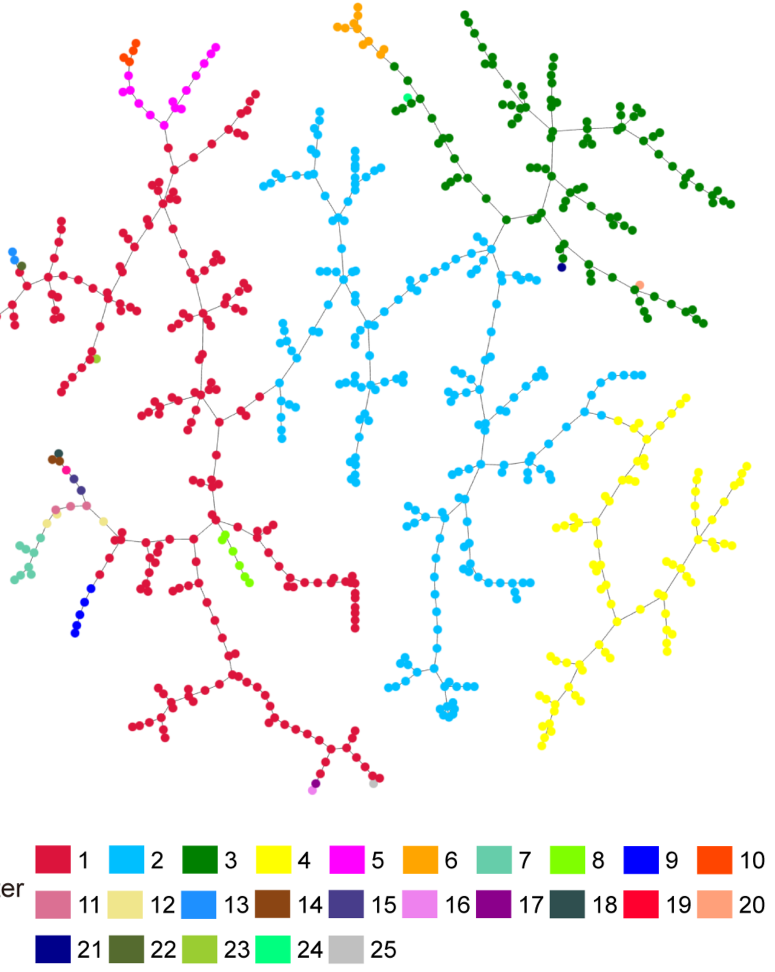
4 largest clusters account for 89.37% known siderophores.

Cluster 1 (201, 30.97%):
siderophores with phenyl ring structures

Cluster 2 (197, 30.35%):
NRPS siderophore with hydroxamate or α-hydroxycarboxylate

Cluster 3 (103, 15.87%):
NIS siderophore with hydroxamate or α-hydroxycarboxylate

Cluster 4 (79, 12.17%):
NRPS siderophores with chromophores



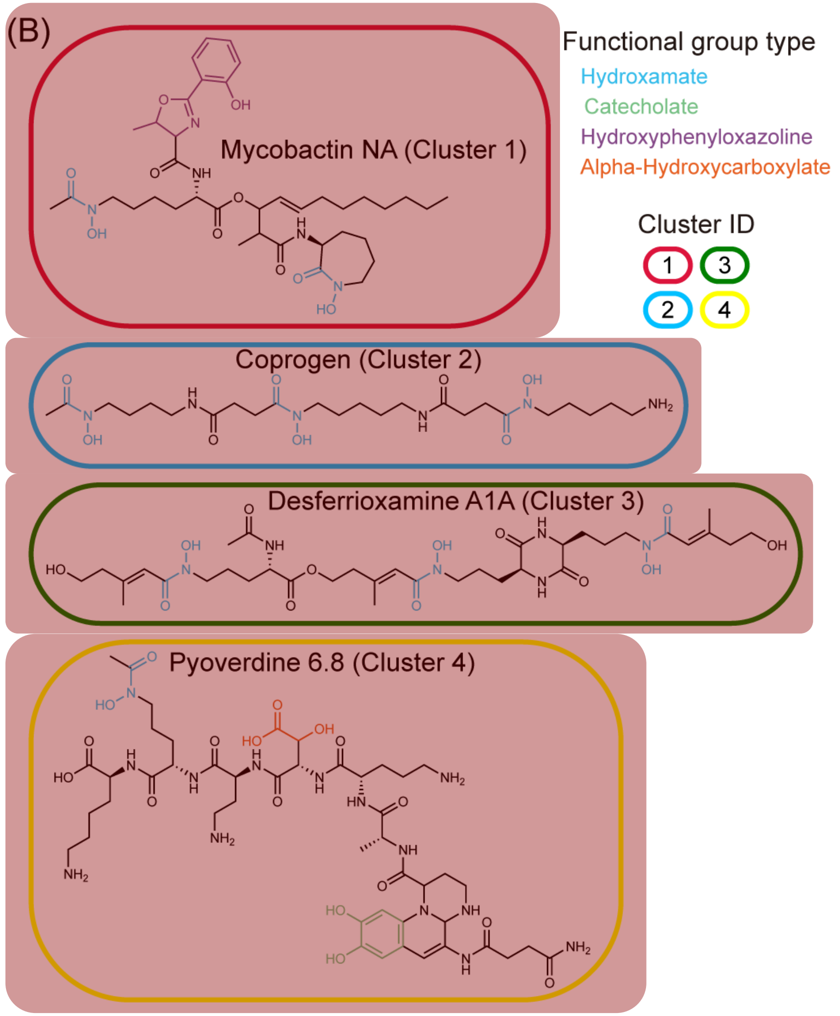


Figure 1 The visualization of 25 siderophore clusters in siderophore information database (SIDERITE).

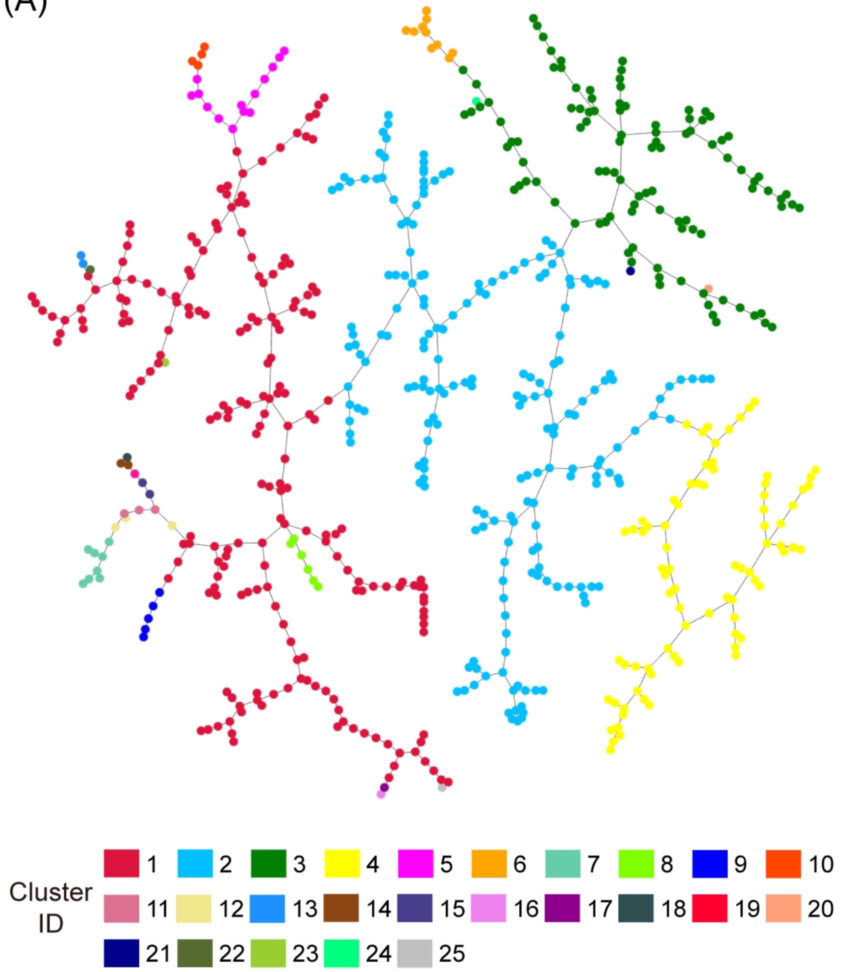
NRPS: non-ribosomal peptide synthetase
NIS : NRPS-independent siderophore

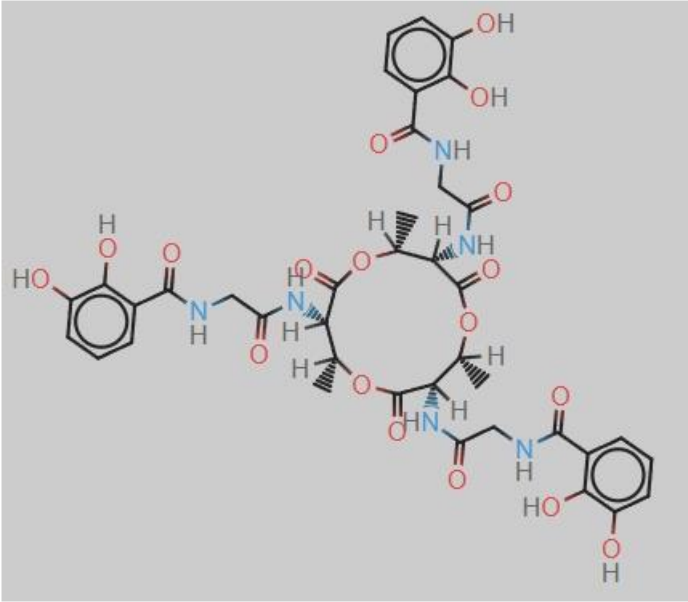
Naming siderophore by clusters and groups

Each siderophore is assigned a unique ID *x.y.z* (Class).

Infer biosynthetic type from other members with known biosynthetic types.

(A)





Example: Bacillibactin

Class: 1.22.3

x. *y* .*z*

Biosynthetic types of siderophores in the same group are same.

x: cluster ID

y: group ID

z: the z-th record

Discovering potential siderophores by the functional group-based method

Rules:
Containing functional groups and
don't contain modified functional groups

Result:
3199 siderophores candidates from COCONUT

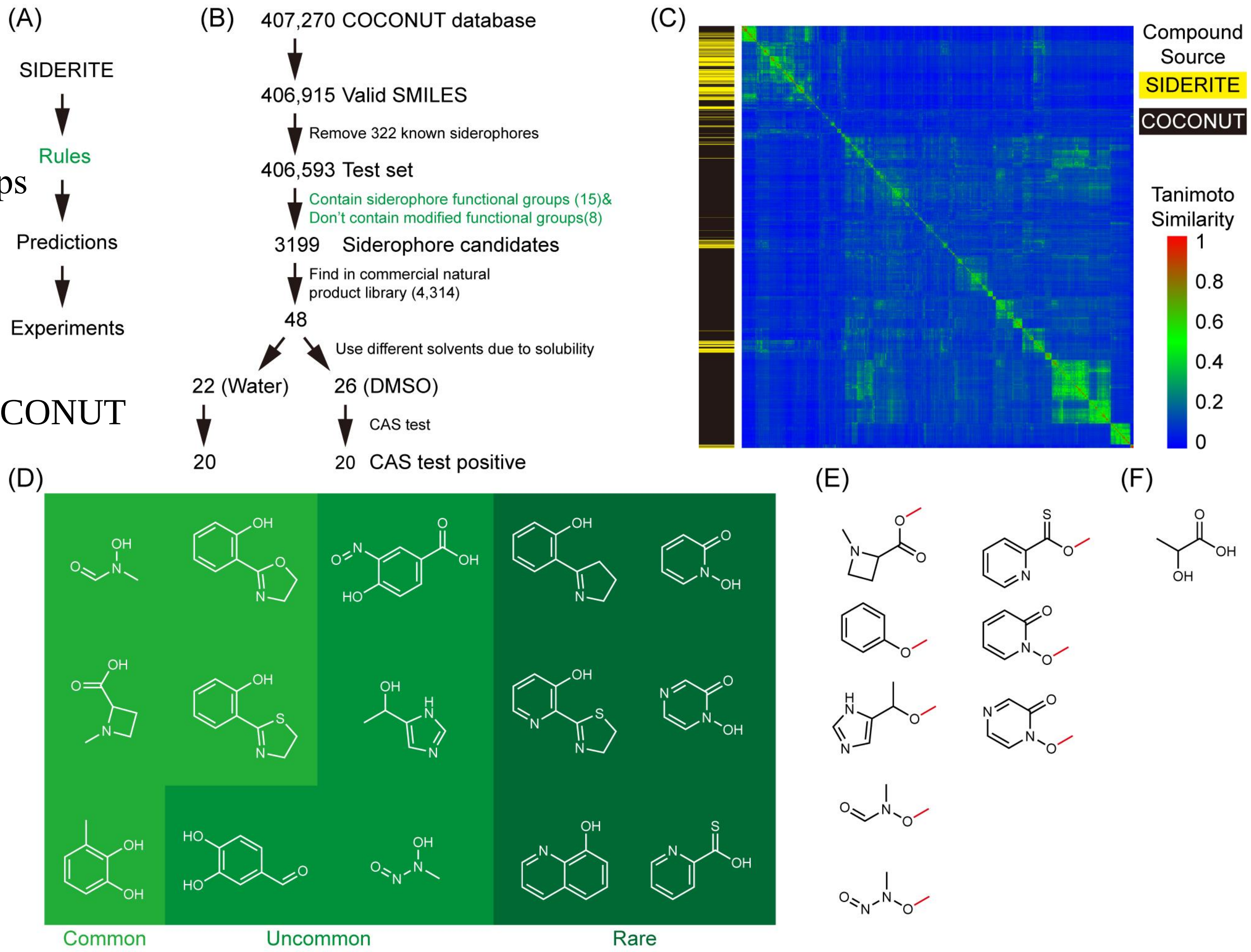


Figure 2 The rule-based siderophore discovery approach and the result of chrome azurol S (CAS) test experiments.

Verifying potential siderophores by CAS assay

48 purchasable molecules from 3199 candidates were verified by CAS assay.

83.3% (40/48) molecules exhibited iron-binding activities.

The high positivity rate from the CAS assay supports the effectiveness of our functional group-based method.

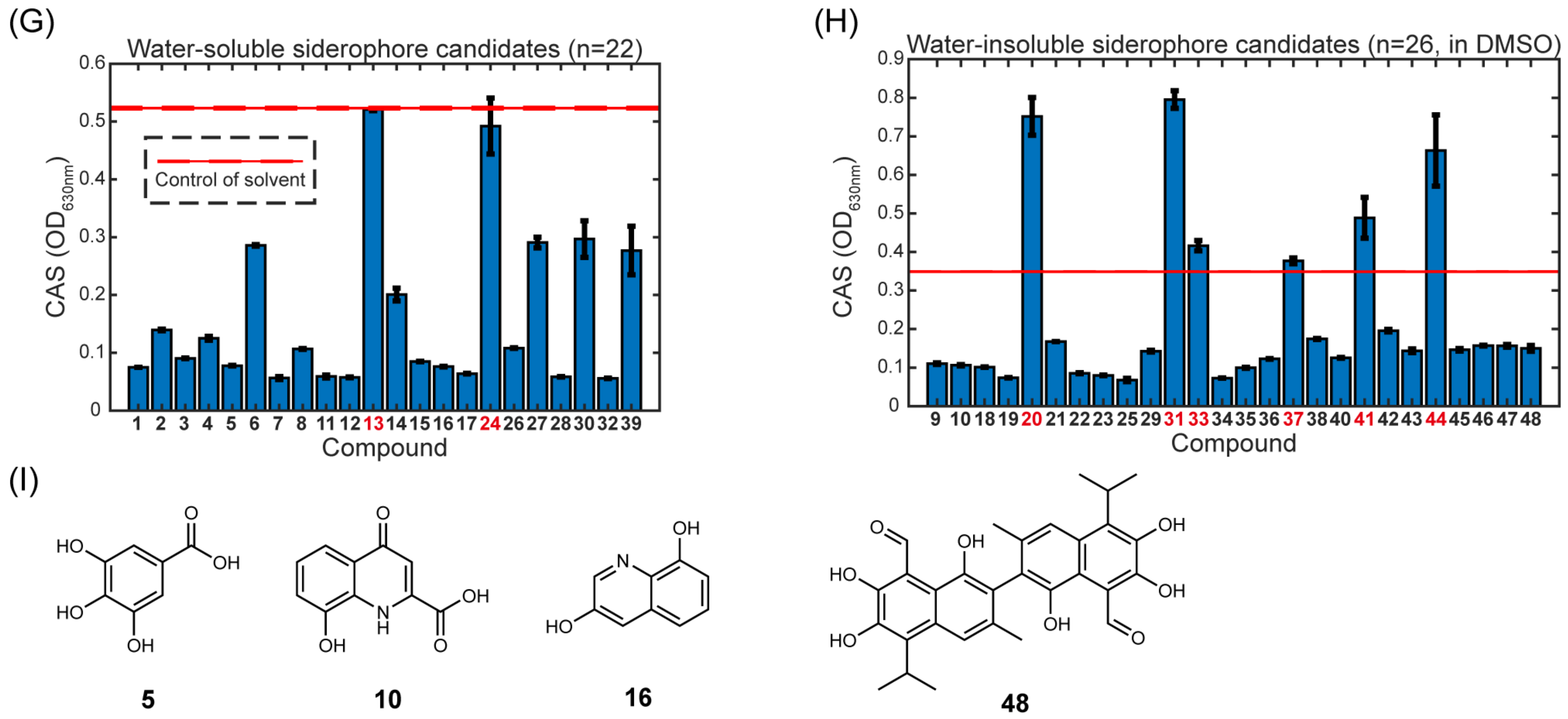


Figure 2 The rule-based siderophore discovery approach and the result of chrome azurol S (CAS) test experiments.

Usage of SIDERITE

BDA Informatics Suite

SIDERITE
Siderophore Information Database

Document: https://github.com/RuolinHe/SIDERITE/wiki
Report bug: https://github.com/RuolinHe/SIDERITE/issues
Forum for discussion: https://groups.google.com/g/siderite-database
Tutorial video: https://www.youtube.com/watch?v=q0EoiRi1zVE

Database | Similarity Search | TMAP Index

Search Fast | Advanced

Total: 649

- 2-deoxymugineic acid
- 2-hydroxybenzoic acid
- 2-HydroxyNicotianamine
- 2-N-hydroxy-3, 4-dihydroisoquinol-2…
- 2-N-methylcoprogen
- 2-N-methylcoprogen B
- 2,3-dihydroxybenzoic acid
- 2,3-dihydroxybenzoylglycine
- 2,3-dihydroxybenzoylserine

Welcome to

SIDERITE - Siderophore Information Database

SIDERITE is a digitized integrated siderophore information database. It covers all 649 known siderophores from 872 records as of May 2023.

There are three ways to access the siderophore resource in the SIDERITE, which are as follows:

1. **Search Based on Conditions**

- **Fast Search:** Provides global real-time fuzzy search
- **Advanced Search:** Provides precise search with support for composite conditions

To initiate your search, please enter your queries on the search sidebar located on the left-hand side of the screen.

2. **Similarity Search**

Find similar siderophores in the SIDERITE with the desired molecule.

3. **Indexing Interface**

A visualized indexing interface of SIDERITE. You can set desired property of siderophores to label all siderophores in the database and select the specific siderophore with a single click.

Summary

- In our work, we introduced the most comprehensive Siderophore Information Database (SIDERTE), the first digitized siderophore repository with 649 unique structures in the SMILES format.

- Based on these digitized structures, a computational method was developed for discovering novel iron-binding molecules with high accuracy and found remarkable structural diversity largely uncharted in the realm of siderophore research.

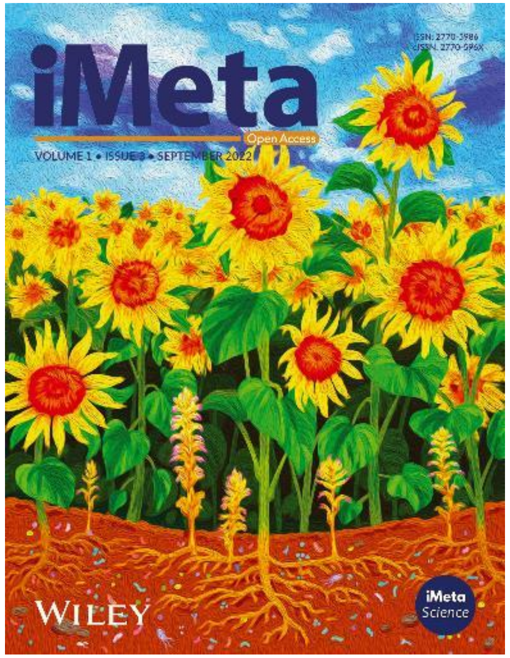
- We provide tutorial materials and feedback channels in the database or the GitHub page, and are committed to maintaining the SIDERITE database continually and updating it based on the feedback received from our users.

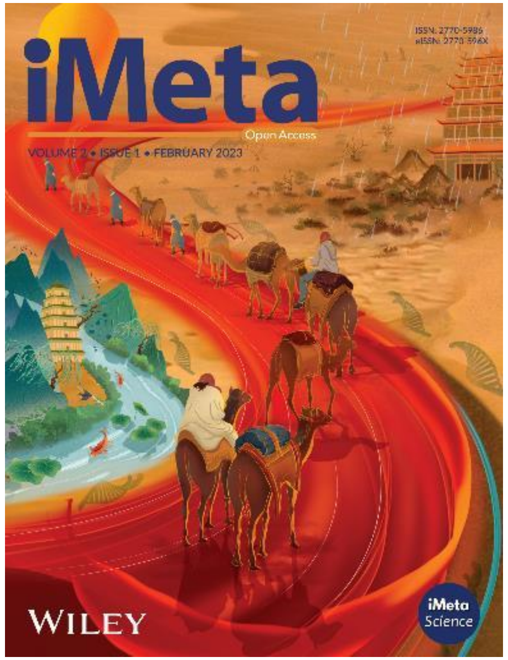
- Website: https://siderite.bdainformatics.org/

Ruolin He, Shaohua Gu, Jiazheng Xu, Xuejian Li, Haoran Chen, Zhengying Shao, Fanhao Wang, Jiqi Shao, Wen-Bing Yin, Long Qian, Zhong Wei, Zhiyuan Li. 2024. SIDERITE: Unveiling Hidden Siderophore Diversity in the Chemical Space Through Digital Exploration. ***iMeta*** 3: e192. https://doi.org/10.1002/imt2.192

iMeta: Integrated meta-omics to change the understanding of the biology and environment

WILEY





"***iMeta***" is an open-access Wiley partner journal launched by iMeta Science Society consist of scientists in bioinformatics and metagenomics world-wide. iMeta aims to promote microbiome, and bioinformatics research by publishing research, methods/protocols, and reviews. The goal is to publish high-quality papers (top 10%, IF>20) targeting a broad audience. Unique features include video submission, reproducible analysis, figure polishing, bilingual, and promotion by social media with 500,000 followers. Since 2022 have been published 160 papers and cited > 2300 times. Index by **ESCI**, Google Scholar, DOAJ and Scopus.

Society: http://www.imeta.science

Publisher: https://wileyonlinelibrary.com/journal/imeta

Submission: https://wiley.atyponrex.com/journal/IMT2

office@imeta.science

Promotion Video

iMetaScience

iMetaScience