Published online 29 November 2022 *Nucleic Acids Research, 2023, Vol. 51, Database issue* ***D621–D628***
https://doi.org/10.1093/nar/gkac1069

NPASS database update 2023: quantitative natural product activity and species source database for biomedical research

Hui Zhao[1,†], Yuan Yang[1,†], Shuaiqi Wang[1,†], Xue Yang[1,†], Kaicheng Zhou[1], Caili Xu[1], Xuyao Zhang[1], Jiajun Fan[1], Dongyue Hou[1], Xingxiu Li[1], Hanbo Lin[1], Ying Tan[2], Shanshan Wang[3], Xin-Yi Chu[3], Dongzhi Zhuoma[4], Fengying Zhang[5], Dianwen Ju[1,*], Xian Zeng[1,*] and Yu Zong Chen[6,*]

[1]Department of Biological Medicines & Shanghai Engineering Research Center of Immunotherapeutics, Fudan University School of Pharmacy, Shanghai 201203, China, [2]The State Key Laboratory of Chemical Oncogenomics & Key Laboratory of Chemical Biology, Tsinghua University Shenzhen Graduate School, Shenzhen Kivita Innovative Drug Discovery Institute, China, [3]Qian Xuesen Collaborative Research Center of Astrochemistry and Space Life Sciences, Institute of Drug Discovery Technology, Ningbo University, Ningbo 315211, China, [4]Medical College, Tibet University, Lhasa 850000, China, [5]Key Lab of Agricultural Products Processing and Quality Control of Nanchang City, Jiangxi Agricultural University, Nanchang 330045, China and [6]Shenzhen Bay Laboratory, Shenzhen 518000, China

Received September 15, 2022; Revised October 16, 2022; Editorial Decision October 17, 2022; Accepted October 26, 2022

ABSTRACT

Quantitative activity and species source data of natural products (NPs) are important for drug discovery, medicinal plant research, and microbial investigations. Activity values of NPs against specific targets are useful for discovering targeted therapeutic agents and investigating the mechanism of medicinal plants. Composition/concentration values of NPs in individual species facilitate the assessments and investigations of the therapeutic quality of herbs and phenotypes of microbes. Here, we describe an update of the NPASS natural product activity and species source database previously featured in NAR. This update includes: (i) new data of ∼95 000 records of the composition/concentration values of ∼1 490 NPs/NP clusters in ∼390 species, (ii) extended data of activity values of ∼43 200 NPs against ∼7 700 targets (∼40% and ∼32% increase, respectively), (iii) extended data of ∼31 600 species sources of ∼94 400 NPs (∼26% and ∼32% increase, respectively), (iv) new species types of ∼440 co-cultured microbes and ∼420 engineered microbes, (v) new data of ∼66 600 NPs without experimental activity values but with estimated activity profiles from the established chemical similarity tool ChemIcal Checker, (vi) new data of the computed drug-likeness properties and the absorption, distribution, metabolism, excretion and toxicity (ADMET) properties for all NPs. NPASS update version is freely accessible at http://bidd.group/NPASS.

INTRODUCTION

Natural products (NPs) are important sources of modern drug discovery (1,2). NP research facilitates the mechanistic investigation of herbal medicines (3,4). In particular, the knowledge of activity values of NPs and their derivatives against specific targets is important for structure-activity studies in drug discovery (5). This knowledge is also needed to assess the therapeutic effects of individual NPs (6), NP combinations (7) and medicinal herbs (8). The composition/concentration values of NPs in individual species facilitate the assessments and investigations of foods, medicinal herbs, and microbes. These include the nutritional, flavor, and beneficial quality of foods (9) and teas (10), the therapeutic potential of medicinal herbs (11), and the phenotype of microbial communities (12).

These investigations can be facilitated by the establishment of the quantitative NP activity and content databases, which include the NP activity and species sources database NPASS (13), the USDA Database for the Flavonoid Content of Selected Foods (https://data.nal.usda.gov/dataset/usda-database-flavonoid-

*To whom correspondence should be addressed. Tel: +86 755 26032094; Email: chenyuzong@sz.tsinghua.edu.cn
Correspondence may also be addressed to Xian Zeng. Tel: +86 51980035; Email: zengxian@fudan.edu.cn
Correspondence may also be addressed to Dianwen Ju. Tel: +86 51980037; Email: dianwenju@fudan.edu.cn
†The authors wish it to be known that, in their opinion, first four authors should be regarded as Joint First Authors.

Table 1. Accumulation of NPs, source organisms, quantitative composition and biological activities in the latest and previous version of NPASS database

		NPASS-2018	NPASS-2023
Natural products (NPs)	NP (with activity values)	**30 926**	**43 285** (increased by 40.0%)
	NP (without activity values)	–	**51 128** (new)
	Total NPs	**30 926**	**94 413** (increased by 205.3%)
Organisms	Natural organisms	**25 041**	**31 690** (increased by 26.6%)
	Co-culture organisms	–	**444** co-culture combinations (new)
	Engineered organisms	–	**427** engineered species (new)
	Total organisms	**25 041**	**32 561** (increased by 30.0%)
Organism-NP pairs		**298 106**	**872 723** (increased by 192.8%)
Activity records		**446 552**	**958 866** (increased by 114.7%)
Targets		**5 863**	**7 753** (increased by 32.2%)
The composition/concentration of NPs in individual species		–	**95 004** quantity records (new) - **398** organisms - **1 490** NP/NP Clusters - **15 292** NP-Organism pairs
Bio- and Chem-properties of NPs		**1 Category** - Physicochemical property	**7 Categories** - Physicochemical properties - Medicinal chemistry properties (new) - ADMET properties (new)
NP similarity metrics		**1 Category** - Chemically structural similarity	**2 Categories** - Chemically structural similarity - Biological similarity (new)

content-selected-foods-release-32-november-2015), the Unified Food Composition Database for the European Project 'Stance4Health' (14) and Flavonoid Database Based on Indonesian Foods (15). The usefulness of these databases can be further enhanced by expanded data coverage of activity values, species sources, and composition/content values for more variety of NPs and species. Moreover, there are a large number of NPs for which experimental activity values are not yet available. Nonetheless, computational methods such as chemical similarity have been established for estimating molecular activity values (16). Methods for evaluating the drug-likeness and the absorption, distribution, metabolism, excretion and toxicity (ADMET) properties have been developed to assess the drug development potential of active molecules (17–19). These tools may be employed for providing estimated activity values and drug-likeness scores of NPs.

Expanding the diversity of NP-producing sources is the fundamental driving force for the discovery of novel NPs (20). To this end, intensive efforts have been made for decades to explore new natural organisms (e.g. organisms of underexplored taxonomic space or novel organisms from extreme living environments like deep-sea sediments) (20). In recent years, new strategies are emerging as attractive lines of NP research, and these include: (i) co-cultures of multiple organisms for mimicking their natural living ecosystems (also termed as microbial consortia or co-cultivation), e.g. fermenting producer and inducer species together to mimic cell-cell communications for induction of novel NPs (21); (ii) genetic engineering of microorganisms through synthetic biology strategies, e.g. introducing synthetic genes to model strains, reprogramming their metabolic pathways, and activating silenced synthetic pathways (22,23). These new NP-producing sources provide either highly-expanded NP chemical diversity or optimized metabolic profiles, contributing to the discovery of many novel NPs or increased production yields of high-value NPs (e.g. biofuel, pharmaceuticals and industrial chemicals). However, current bioinformatics resources basically focus on NPs isolated from naturally individual species. There is a lack of an open-access database for data on NPs from species co-cultivations and engineered microorganisms.

We therefore conducted several major updates of the NPASS database (Table 1). The first is the addition of ∼95 000 new records of the composition/concentration values of ∼1 500 NPs/NP clusters in ∼400 species. The second is the extended data of activity values of ∼43 200 NPs against ∼7 700 targets (∼40% and ∼32% expansion, respectively). The third is the expanded data of ∼31 600 species sources of ∼94 400 NPs (∼26% and ∼32% increase, respectively). The fourth is the new data of species types of ∼440 co-cultured microbes and ∼420 engineered microbes. The fifth is the new data of ∼67 000 NPs without experimental activity values but with estimated activity profiles from the established bioactivity similarity tool Chemical Checker (16). The sixth is the computed drug-likeness properties and the absorption, distribution, metabolism, excretion and toxicity (ADMET) properties for all NPs using the established tool ADMETlab 2.0 (17). The key features of these updates were illustrated in Figure 1. A comparison of NPASS with several representative NP databases, including StreptomeDB (24), The Natural Products Atlas (25), CMNPD (26) and COCONUT (27) was provided in Table 2. NPASS is freely accessible at http://bidd.group/NPASS.

COMPOSITION/CONCENTRATION VALUES OF NATURAL PRODUCTS

Knowledge of the composition/concentration values of natural products is valuable in assessing the nutrition and health-beneficial quality of foods and medicinal plants. For instance, the plant family *Brassicaceae* includes well-known vegetables such as broccoli, cabbage, cauliflower, mustard, rapeseed, rocket and turnip, which were among the first plants cultivated and domesticated by humanity (28). These

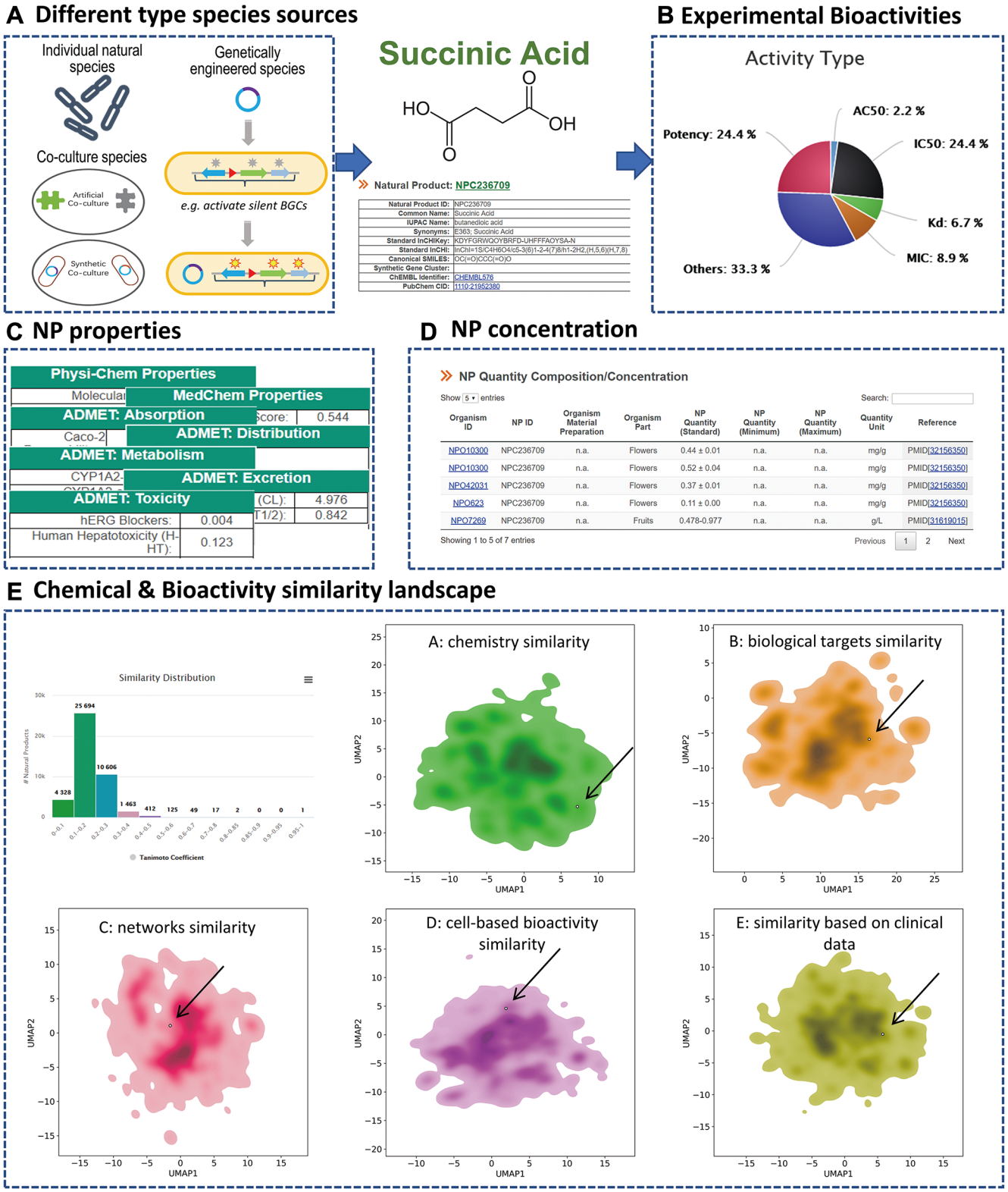

Figure 1. Illustration of key features of NPASS-2023 was shown using the example NP succinic acid. (**A**) Several types of species sources are included: individual natural species, co-cultured species and genetically engineered species. (**B**) Experiments-derived biological activities of NPs were significantly expended in this update. (**C**) In addition to physico-chemical properties, new property layers were calculated by ADMETlab 2.0. (**D**) Quantitative composition/concentration data of NPs in their producing species was included in this update. (**E**) In addition to traditional structure-based similarity, a five-level bioactivity similarity distribution was built based on Chemical Checker descriptors and visualized in 2D UMAP projection. The current NP was labeled with a small circle (indicated by those long arrows) in UMAP projections.

Table 2. Different features of representative NP databases

Database name		NPASS	StreptomeDB	The natural products atlas	CMNPD	COCONUT
Coverage of organism category		Diverse (plants, bacteria, metazoan, fungi)	Specialized (Streptomyces genus)	Specialized (bacteria, fungi)	Specialized (marine species, e.g. bacteria, fungi, algae, sponges)	Diverse (plants, bacteria, fungi, marine species, animals)
No. of organisms	Natural organisms	31 690	3302	1200	3354	n.a.
	Co-culture Organisms	444	n.a.*	n.a.	n.a.	n.a.
	Engineered Organisms	427	n.a.	n.a.	n.a.	n.a.
	Total organisms	32 561	3302	1200	3354	n.a.
No. of NPs		94 413	6524	32 552	31 561	1 136 517
No. of activity records		958 866	1 031	n.a.	72 349	n.a.
No. of quantity records		95 004	n.a.	n.a.	n.a.	n.a.
No. of organism-NP pairs		872 723	10 912	n.a.	n.a.	n.a.
No. of targets		7753	n.a.	n.a.	2652	n.a.
Properties of NPs	physicochemical properties	Y*	Y	N*	Y	N
	Medicinal chemistry properties	Y	N	N	N	N
	ADMET properties	Y	Y	N	Y	N
Similarity metrics	Chemically structural similarity	Y	N	N	N	N
	Biological similarity	Y	N	N	N	N
Latest version		2.0 (2023)	3.0 (2021)	2.0 (2022)	1.0 (2021)	1.0 (2022)

*Note: n.a. means not available; Y means this data category was included; N means this data category was not included.

vegetables contain chemical ingredients of high nutritional and health beneficial value. In particular, glucosinolates are one of the most important classes of secondary metabolites in the *Brassicaceae* family, and the compositions of these metabolites influence the level of protective and preventive effects against several cancer (28). Moreover, *Brassicaceae* plants contain an important inorganic micronutrient, Se, which serves as part of the active site of the antioxidant enzyme glutathione peroxidase, and the content of Se affects the antioxidant effects of the vegetables (28). Therefore, the experimentally determined composition/content of individual NPs in plants and microbes was searched from the PubMed database (29). These search results were obtained by using the keywords 'composition', 'content', 'chemical characterization', 'biochemical characterization', 'phytochemical characterization', 'abundance' or 'profile' in combination with 'plant', 'herb', 'vegetable', 'fruit', 'microbe' and 'bacterium'.

CO-CULTURED AND ENGINEERED ORGANISMS FOR ENHANCED NP-PRODUCTION

In the genomic and synthetic biology era, strategies for the discovery of novel NPs are rapidly evolving (1). Optimization of cultivation systems and genetic modification of species are emerging strategies to expand the sources of NP discovery from traditional models (individual natural organisms) to more sophisticated ones (co-cultivation of multiple species and unnatural organisms). Naturally, microorganisms grow and function (e.g. producing specialized metabolites) in concert within their surrounding microbial ecosystem, rather than as individual/isolated species. Species communities can exert positive microbial interactions that benefit community members through several mechanisms such as syntrophy, cooperation, mutualism, and commensalism (30). By recovering or designing specialized positive microbial interactions, co-culture systems can produce novel NPs or increase the yield of existing NPs through reshaping metabolic profiles of individual organisms (21,31). In addition, many synthetic gene clusters for high-value NPs biosynthesis are silenced in natural organisms. By activating silenced synthetic pathways or introducing synthetic enzyme genes, natural species can be engineered to produce high-value NPs more efficiently (22,32).

These strategies represent promising trends in NP research, which leverages large-scale mining of synthetic gene clusters to exploit biosynthetic potential and manipulate key genes or metabolic parameters to enable the production of novel NPs or increased NP yields (33–35). Therefore, we compiled NP yields or yield changes of NPs in co-cultured species and engineered species through an extensive search of PubMed. Finally, NPASS included 444 co-culture species combinations and 427 engineered species/strains manually extracted from 221 and 382 publications, respectively. During data curation, key manipulation parameters, such as the taxonomic background of parent species, the synthetic genes regulated/introduced, the fold-changes of NP yields and other information, were also extracted from the original references when available. As shown in Figure 2, diverse co-culture combinations were curated in NPASS. Phylogenetic trees were generated based on NCBI taxonomy (36) and visualized using the iTOL web server (37) to present the landscape of co-culture combinations. These curated engineered species (Figure 2B) and co-culture combinations (Figure 2C) contributed to both the discovery of novel NPs (4% and 36%, respectively) and yield increase of existing high-value NPs (89% and 40%, respectively).

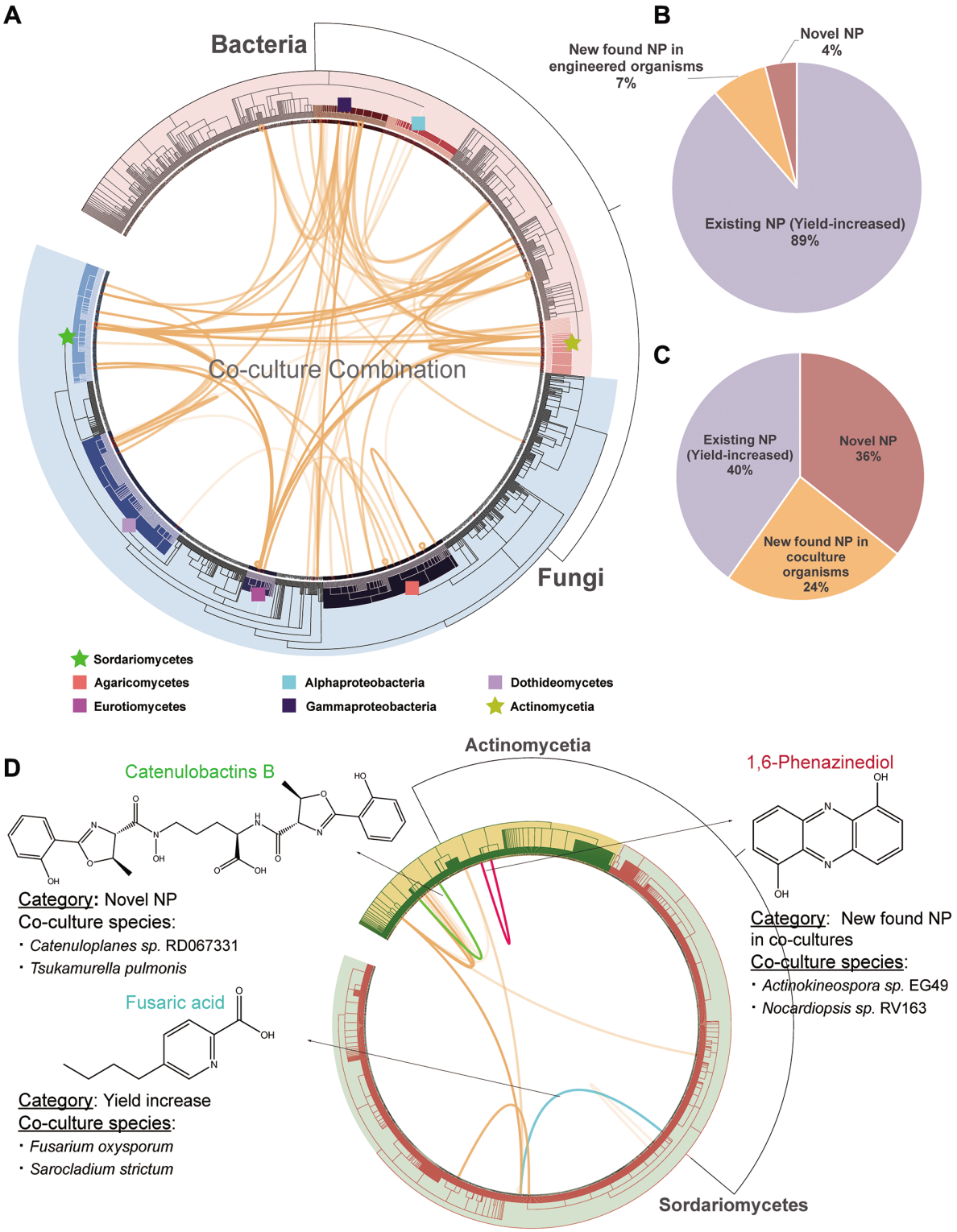

Figure 2. An overview of co-cultured and engineered species for optimized NP production. (**A**) An overview of co-culture combinations was projected on a family-level phylogenetic tree of bacteria and fungi. Taxonomic classes that enriched co-cultured species were highlighted with different colors and labeled with different shapes. The phylogenetic tree was built using phyloT-v2 (https://phylot.biobyte.de/) based on NCBI taxonomic identifiers and visualized using iTOL. Connection lines between taxonomic tree leaves indicate co-culture combinations. NPs produced by engineered or co-cultured species were classified into three categories: (i) novel NP: novel structure that reported for the first time; (ii) new found NP: NPs that produced by engineered species but not found in wild type parent species; (iii) existing NP (Yield-increased): NPs produced by both parent species and engineered species but have increased yield in engineered species. The percentages of NPs produced by engineered species (**B**) and co-cultured species (**C**) were shown in pie charts. (**D**) The family-level phylogenetic tree in (A) was zoomed in at genus-level for *Actinomycetia* and *Sordariomycetes* class to illustrate typical examples of NP production in co-culture combinations. NCBI Taxonomic IDs used to generate phylogenetic trees were provided in Supplementary Table S1.

ESTIMATING ACTIVITY PROFILES OF NATURAL PRODUCTS WITH THE MOLECULAR SIMILARITY PRINCIPLE

NPs are the most significant source for drug discovery, while the majority of isolated NPs have not been subjected to bioactivity evaluation against broad therapeutic targets. Characterizing the bioactive space of NP is critical for more efficient NP-based research and drug discovery. The molecular similarity principle has become a driving force of drug discovery, particularly in molecular activity prediction and activity profile estimation (16,38). This principle has also been applied for charting and navigating NP chemical space (39), probing clustered patterns of NP drugs against privileged molecular target sites (40), and developing NP likeness scores (41).

Recently, these chemical structure-centric similarity metrics have been extended to a broad level of biology, aiming to capture more comprehensive representations of biological properties/behaviors of chemical compounds at different biological levels. For example, a five-levels (chemistry, targets, networks, cells, and clinics) Chemical Checker bioactivity descriptor was proposed (16), and it exhibits enhanced performance in predicting the biological activity of compounds that lack experimental bioactivity evaluation (42). This provides clues for revealing the bioactive space of NPs. In the updated NPASS database, we employed the Chemical Checker (16) to generate bioactivity descriptors for all NPs and subsequently constructed the NPs' bioactivity similarity space for NPASS. In detail, continuous descriptor vectors generated by Chemical Checker were subjected to Uniform Manifold Approximation and Projection (UMAP) manifold learning for dimension reduction. Five 2D sub-spaces A, B, C, D and E (represent chemistry, targets, networks, cells, and clinics levels, respectively) were built (based on the first two dimensions learned by UMAP) and visualized as density maps generated by ChemPlot Python package (43). For each NP, its locations in density maps were highlighted using a small circle. Biological similarities of all NP pairs were calculated using Euclidean distance based on UMAP projection coordinates (UMAP distance) to navigate the NP biological similarity space. A smaller UMAP distance between NPs indicated a closer bioactivity similarity, which allowed us to explore top-N similar NPs of an individual NP in each bioactivity sub-spaces. Due to the inherent differences between Chemical Checker sub-spaces, for a specific NP, the lists of top-N similar NPs in different bioactivity sub-spaces are not necessarily consistent (16). This may be useful for inferring activity profiles of those NPs which have no experimental activity record via exploring their top similar active NPs in each sub-space.

DRUG-LIKENESS AND ADMET PROFILING OF NATURAL PRODUCTS

The drug development potential of active molecules can be partly judged by the drug-likeness properties (18,19). In particular, drug-likeness rules have been developed and widely used in drug discovery (19,44,45). These rules exploit the drug's distinguished physicochemical properties, such as molecular weight and the number of hydrogen bond donors, as the basis for drug-likeness evaluations (46). The therapeutic effectiveness and safety of drugs are also strongly influenced by the absorption, distribution, metabolism, excretion (ADME) and toxicity (T) profiles of these drugs (47). Early evaluation of ADMET properties of molecules, including NPs, is useful for selecting promising candidates in drug discovery (48). We therefore used the ADMETlab 2.0 (17) web server for computing physical-chemical properties, medicinal chemistry properties, and ADMET properties for each NP.

CONCLUDING REMARKS

There has been increasing interest in NP drug discovery (1), new nutrition sources from plant diversity (49), and the exploration of beneficiary effects of herbs (50). These extensive investigations are generating data on a wider variety of NP activity (51,52) and on the composition/concentration of NPs in various species (53,54). The rich information generated from these and future investigations can be incorporated into the NPASS database and other established databases (https://data.nal.usda.gov/dataset/usda-database-flavonoid-content-selected-foods-release-32-november-2015) (14,15). These databases with extended data collectively provide enriched resources for facilitating drug discovery (1), functional food development (55), and investigations and explorations of herbal medicines (3).

DATA AVAILABILITY

No new data were generated or analysed in support of this research.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Shanghai Sailing Program [20YF1402700]; National Natural Science Foundation of China [32000479]; State Key Laboratory of Chemical Oncogenomics and the National Key R&D Program of China, Synthetic Biology Research [2019YFA0905900]; Scientific Research Grant of Ningbo University [215-432000282]; Ningbo Top Talent Project [215-432094250]; Shenzhen Bay Laboratory [201901]; China Postdoctoral Science Foundation [2021M690714]. Funding for open access charge: Shanghai Sailing Program [20YF1402700].
Conflict of interest statement. None declared.

REFERENCES

1. International Natural Product Sciences Taskforce, Atanasov,A.G., Zotchev,S.B., Dirsch,V.M. and Supuran,C.T. (2021) Natural products in drug discovery: advances and opportunities. *Nat. Rev. Drug Discov.*, **20**, 200–216.
2. Tao,L., Zhu,F., Qin,C., Zhang,C., Xu,F., Tan,C.Y., Jiang,Y.Y. and Chen,Y.Z. (2014) Nature's contribution to today's pharmacopeia. *Nat. Biotechnol.*, **32**, 979–980.

3. Li,F.S. and Weng,J.K. (2017) Demystifying traditional herbal medicine with modern approach. *Nat. Plants*, **3**, 17109.
4. Zeng,X., Zhang,P., Wang,Y., Qin,C., Chen,S., He,W., Tao,L., Tan,Y., Gao,D., Wang,B. *et al.* (2019) CMAUP: a database of collective molecular activities of useful plants. *Nucleic Acids Res.*, **47**, D1118–D1127.
5. Itoh,H. and Inoue,M. (2019) Comprehensive structure-activity relationship studies of macrocyclic natural products enabled by their total syntheses. *Chem. Rev.*, **119**, 10002–10031.
6. He,Q.R., Tang,J.J., Liu,Y., Chen,Z.F., Liu,Y.X., Chen,H., Li,D., Yi,Z.F. and Gao,J.M. (2021) The natural product trienomycin a is a STAT3 pathway inhibitor that exhibits potent in vitro and in vivo efficacy against pancreatic cancer. *Br J. Pharmacol.*, **178**, 2496–2515.
7. Qin,C., Tan,K.L., Zhang,C.L., Tan,C.Y., Chen,Y.Z. and Jiang,Y.Y. (2012) What does it take to synergistically combine sub-potent natural products into drug-level potent combinations? *PLoS One*, **7**, e49969.
8. Wang,Y., Zeng,X., Zhao,Y., Chen,W. and Chen,Y.Z. (2020) The pros and cons of traditional chinese medicines in the treatment of COVID-19. *Pharmacol. Res.*, **157**, 104873.
9. Hounsome,N., Hounsome,B., Tomos,D. and Edwards-Jones,G. (2008) Plant metabolites and nutritional quality of vegetables. *J. Food Sci.*, **73**, R48–R65.
10. Ramdani,D., Chaudhry,A.S. and Seal,C.J. (2013) Chemical composition, plant secondary metabolites, and minerals of green and black teas and the effect of different tea-to-water ratios during their extraction on the composition of their spent leaves as potential additives for ruminants. *J. Agric. Food Chem.*, **61**, 4961–4967.
11. Ni,H., Liu,J., Dai,O., Feng,R., Liu,F., Cao,X.Y., Peng,C. and Xiong,L. (2021) Chemical composition and uterine smooth muscle relaxant activity of essential oils from 10 kinds of blood-activating and stasis-resolving chinese medicinal herbs. *J. Ethnopharmacol.*, **269**, 113713.
12. Chen,L., Collij,V., Jaeger,M., van den Munckhof,I.C.L., Vich Vila,A., Kurilshikov,A., Gacesa,R., Sinha,T., Oosting,M., Joosten,L.A.B. *et al.* (2020) Gut microbial co-abundance networks show specificity in inflammatory bowel disease and obesity. *Nat. Commun.*, **11**, 4018.
13. Zeng,X., Zhang,P., He,W., Qin,C., Chen,S., Tao,L., Wang,Y., Tan,Y., Gao,D., Wang,B. *et al.* (2018) NPASS: natural product activity and species source database for natural product research, discovery and tool development. *Nucleic Acids Res.*, **46**, D1217–D1222.
14. Hinojosa-Nogueira,D., Perez-Burillo,S., Navajas-Porras,B., Ortiz-Viso,B., de la Cueva,S.P., Lauria,F., Fatouros,A., Priftis,K.N., Gonzalez-Vigil,V. and Rufian-Henares,J.A. (2021) Development of an unified food composition database for the european project "Stance4Health". *Nutrients*, **13**, 4206.
15. Sefrina,L.R., Briawan,D., Sinaga,T. and Permaesih,D. (2020) Flavonoid database based on indonesian foods. *J. Nutr. Sci. Vitaminol (Tokyo)*, **66**, S251–S255.
16. Duran-Frigola,M., Pauls,E., Guitart-Pla,O., Bertoni,M., Alcalde,V., Amat,D., Juan-Blanco,T. and Aloy,P. (2020) Extending the small-molecule similarity principle to all levels of biology with the chemical checker. *Nat. Biotechnol.*, **38**, 1087–1096.
17. Xiong,G., Wu,Z., Yi,J., Fu,L., Yang,Z., Hsieh,C., Yin,M., Zeng,X., Wu,C., Lu,A. *et al.* (2021) ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Res.*, **49**, W5–W14.
18. Bickerton,G.R., Paolini,G.V., Besnard,J., Muresan,S. and Hopkins,A.L. (2012) Quantifying the chemical beauty of drugs. *Nat. Chem.*, **4**, 90–98.
19. Leeson,P.D. and Springthorpe,B. (2007) The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discov.*, **6**, 881–890.
20. Pye,C.R., Bertin,M.J., Lokey,R.S., Gerwick,W.H. and Linington,R.G. (2017) Retrospective analysis of natural products provides insights for future discovery trends. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 5601–5606.
21. Arora,D., Gupta,P., Jaglan,S., Roullier,C., Grovel,O. and Bertrand,S. (2020) Expanding the chemical diversity through microorganisms co-culture: current status and outlook. *Biotechnol. Adv.*, **40**, 107521.
22. Song,H., Ding,M.Z., Jia,X.Q., Ma,Q. and Yuan,Y.J. (2014) Synthetic microbial consortia: from systematic analysis to construction and applications. *Chem. Soc. Rev.*, **43**, 6954–6981.
23. Liu,J., Wang,X., Dai,G., Zhang,Y. and Bian,X. (2022) Microbial chassis engineering drives heterologous production of complex secondary metabolites. *Biotechnol. Adv.*, **59**, 107966.
24. Moumbock,A.F.A., Gao,M., Qaseem,A., Li,J., Kirchner,P.A., Ndingkokhar,B., Bekono,B.D., Simoben,C.V., Babiaka,S.B., Malange,Y.I. *et al.* (2021) StreptomeDB 3.0: an updated compendium of streptomycetes natural products. *Nucleic Acids Res.*, **49**, D600–D604.
25. van Santen,J.A., Poynton,E.F., Iskakova,D., McMann,E., Alsup,T.A., Clark,T.N., Fergusson,C.H., Fewer,D.P., Hughes,A.H., McCadden,C.A. *et al.* (2022) The natural products atlas 2.0: a database of microbially-derived natural products. *Nucleic Acids Res.*, **50**, D1317–D1323.
26. Lyu,C., Chen,T., Qiang,B., Liu,N., Wang,H., Zhang,L. and Liu,Z. (2021) CMNPD: a comprehensive marine natural products database towards facilitating drug discovery from the ocean. *Nucleic Acids Res.*, **49**, D509–D515.
27. Sorokina,M., Merseburger,P., Rajan,K., Yirik,M.A. and Steinbeck,C. (2021) COCONUT online: collection of open natural products database. *J. Cheminform.*, **13**, 2.
28. Avato,P. and Argentieri,M.P. (2015) Brassicaceae: a rich source of health improving phytochemicals. *Phytochem. Rev.*, **14**, 1019–1033.
29. Sayers,E.W., Beck,J., Bolton,E.E., Bourexis,D., Brister,J.R., Canese,K., Comeau,D.C., Funk,K., Kim,S., Klimke,W. *et al.* (2021) Database resources of the national center for biotechnology information. *Nucleic Acids Res.*, **49**, D10–D17.
30. Canon,F., Nidelet,T., Guedon,E., Thierry,A. and Gagnaire,V. (2020) Understanding the mechanisms of positive microbial interactions that benefit lactic acid bacteria Co-cultures. *Front. Microbiol.*, **11**, 2088.
31. Kong,W., Meldgin,D.R., Collins,J.J. and Lu,T. (2018) Designing microbial consortia with defined social interactions. *Nat. Chem. Biol.*, **14**, 821–829.
32. Shi,Y.M., Hirschmann,M., Shi,Y.N., Ahmed,S., Abebew,D., Tobias,N.J., Grun,P., Crames,J.J., Poschel,L., Kuttenlochner,W. *et al.* (2022) Global analysis of biosynthetic gene clusters reveals conserved and unique natural products in entomopathogenic nematode-symbiotic bacteria. *Nat. Chem.*, **14**, 701–712.
33. Hemmerling,F. and Piel,J. (2022) Strategies to access biosynthetic novelty in bacterial genomes for drug discovery. *Nat. Rev. Drug Discov.*, **21**, 359–378.
34. Scherlach,K. and Hertweck,C. (2021) Mining and unearthing hidden biosynthetic potential. *Nat. Commun.*, **12**, 3864.
35. Medema,M.H., de Rond,T. and Moore,B.S. (2021) Mining genomes to illuminate the specialized chemistry of life. *Nat. Rev. Genet.*, **22**, 553–571.
36. Schoch,C.L., Ciufo,S., Domrachev,M., Hotton,C.L., Kannan,S., Khovanskaya,R., Leipe,D., McVeigh,R., O'Neill,K., Robbertse,B. *et al.* (2020) NCBI taxonomy: a comprehensive update on curation, resources and tools. *Database (Oxford)*, **2020**, baaa062.
37. Letunic,I. and Bork,P. (2021) Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.*, **49**, W293–W296.
38. Franco,P., Porta,N., Holliday,J.D. and Willett,P. (2017) Molecular similarity considerations in the licensing of orphan drugs. *Drug Discov. Today*, **22**, 377–381.
39. Lachance,H., Wetzel,S., Kumar,K. and Waldmann,H. (2012) Charting, navigating, and populating natural product chemical space for drug discovery. *J. Med. Chem.*, **55**, 5989–6001.
40. Tao,L., Zhu,F., Qin,C., Zhang,C., Chen,S., Zhang,P., Zhang,C., Tan,C., Gao,C., Chen,Z. *et al.* (2015) Clustered distribution of natural product leads of drugs in the chemical space as influenced by the privileged target-sites. *Sci. Rep.*, **5**, 9325.
41. Jayaseelan,K.V., Moreno,P., Truszkowski,A., Ertl,P. and Steinbeck,C. (2012) Natural product-likeness score revisited: an open-source, open-data implementation. *BMC Bioinformatics*, **13**, 106.
42. Bertoni,M., Duran-Frigola,M., Badia,I.M.P., Pauls,E., Orozco-Ruiz,M., Guitart-Pla,O., Alcalde,V., Diaz,V.M., Berenguer-Llergo,A., Brun-Heath,I. *et al.* (2021) Bioactivity descriptors for uncharacterized chemical compounds. *Nat. Commun.*, **12**, 3932.
43. Cihan Sorkun,M., Mullaj,D., Koelman,J.M.V.A. and Er,S. (2022) ChemPlot, a python library for chemical space visualization**. *Chemistry–Methods*, **2**, e202200005.

44. Lipinski,C.A., Lombardo,F., Dominy,B.W. and Feeney,P.J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.*, **46**, 3–26.
45. Lipinski,C.A. (2004) Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov. Today Technol.*, **1**, 337–341.
46. Lipinski,C.A. (2000) Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods*, **44**, 235–249.
47. van de Waterbeemd,H. and Gifford,E. (2003) ADMET in silico modelling: towards prediction paradise?*Nat. Rev. Drug Discov.*, **2**, 192–204.
48. Wishart,D.S. (2007) Improving early drug discovery through ADME modelling: an overview. *Drugs R D*, **8**, 349–362.
49. Cantwell-Jones,A., Ball,J., Collar,D., Diazgranados,M., Douglas,R., Forest,F., Hawkins,J., Howes,M.R., Ulian,T., Vaitla,B. *et al.* (2022) Global plant diversity as a reservoir of micronutrients for humanity. *Nat. Plants*, **8**, 225–232.
50. doValle,I.F., Roweth,H.G., Malloy,M.W., Moco,S., Barron,D., Battinelli,E., Loscalzo,J. and Barabási,A.-L. (2021) Network medicine framework shows that proximity of polyphenol targets and disease proteins predicts therapeutic effects of polyphenols. *Nat. Food*, **2**, 143–155.
51. Tavakkoli,A., Iranshahi,M., Hasheminezhad,S.H., Hayes,A.W. and Karimi,G. (2019) The neuroprotective activities of natural products through the nrf2 upregulation. *Phytother Res.*, **33**, 2256–2273.
52. Liu,Y., Jing,S.X., Luo,S.H. and Li,S.H. (2019) Non-volatile natural products in plant glandular trichomes: chemistry, biological activities and biosynthesis. *Nat. Prod. Rep.*, **36**, 626–665.
53. Grymel,M., Mazurkiewicz,R., Bajkacz,S., Bilik,J. and Kowalczyk,S. (2022) Extraction, purification, quantification, and stability of bioactive spilanthol from acmella oleracea. *Planta Med.*, https://doi.org/10.1055/a-1903-2226.
54. D'Angiolo,M., De Leo,M., Camangi,F., Magliocca,G., De Tommasi,N., Braca,A. and Marzocco,S. (2022) Chemical constituents of ulmus minor subsp. minor fruits used in the italian phytoalimurgic tradition and their Anti-inflammatory activity evaluation. *Planta Med.*, **88**, 762–773.
55. Hunter,P.M. and Hegele,R.A. (2017) Functional foods and dietary supplements for the management of dyslipidaemia. *Nat. Rev. Endocrinol.*, **13**, 278–288.