

秦紫嫣10.20周报

1.收集171个物种的代谢产物信息



input type = all , input word = **Tripterygium wilfordii**

Number of matched data :280

C_ID	CAS ID	Metabolite	Molecular formula	Mw	Organism or InC
C00000207	65-85-0	Benzoic acid	C7H6O2	122.03677944	Tripterygium wilfordii
C00000219	1211-29-6	Methyl jasmonate	C13H20O3	224.1412445	Tripterygium wilfordii
C00000609	7770-78-7	Arctigenin	C21H24O6	372.1572885	Tripterygium wilfordii
C00000627	6216-81-5	Lirioresinol B	C22H26O8	418.16276781	Tripterygium wilfordii
C00000659	40957-99-1	Medioresinol	C21H24O7	388.15220312	Tripterygium wilfordii
C00000674	491-70-3	5,7,3',4'-Tetrahydroxyflavone	C15H10O6	286.04773805	Tripterygium wilfordii
C00000846	763-10-0	Geranyl diphosphate	C10H20O7P2	314.06842597	Tripterygium wilfordii
C00000848	358-71-4	Isopentenyl diphosphate	C5H12O7P2	246.00582572	Tripterygium wilfordii
C00000876	21738-30-7	Copalyl diphosphate	C20H36O7P2	450.19362648	Tripterygium wilfordii
C00000907	13058-04-3	Farnesyl pyrophosphate	C15H28O7P2	382.13102623	Tripterygium wilfordii
C00000908	6699-20-3	Geranylgeranyl diphosphate	C20H36O7P2	450.19362648	Tripterygium wilfordii
C00000977	578-86-9	Liquiritigenin	C15H12O4	256.07355887	Tripterygium wilfordii
C00000992	480-39-7	Dihydrochrysin	C15H12O4	256.07355887	Tripterygium wilfordii
C00001076	478-01-3	Nobiletin	C21H22O8	402.13146768	Tripterygium wilfordii
C00001176	64-19-7	Acetic acid	C2H4O2	60.02112937	Tripterygium wilfordii
C00001195	150-97-0	Mevalonic acid	C6H12O4	148.07355887	Tripterygium wilfordii
C00001528	479-61-8	Chlorophyll a	C55H72MgN4O5	892.53531331	Tripterygium wilfordii
C00001550	79-83-4	Pantothenic acid	C9H17NO5	219.11067266	Tripterygium wilfordii
C00001987	37239-51-3	Wilfordin	C43H49NO19	883.28987839	Tripterygium wilfordii
C00002322	53938-08-2	Celabenzine	C23H29N3O2	379.22597719	Tripterygium wilfordii
C00002485	495-32-9	Nodakenetin	C14H14O4	246.08920894	Tripterygium wilfordii
C00002526	446-72-0	Genistein	C15H10O5	270.05282343	Tripterygium wilfordii
C00002609	564-73-8	Hinokiol	C20H30O2	302.2245802	Tripterygium wilfordii
C00002631	21453-69-0	(+)-Lirioresinol B	C22H26O8	418.16276781	Tripterygium wilfordii
C00002637	642-71-7	3,4,5-Trimethoxyphenol	C9H12O4	184.07355887	Tripterygium wilfordii
C00003345	20554-84-1	Parthenolide	C15H20O3	248.1412445	Tripterygium wilfordii
C00003493	38647-10-8	Triptolide	C20H24O7	376.15220312	Tripterygium wilfordii
C00003494	38748-32-2	Triptolide	C20H24O6	360.1572885	Tripterygium wilfordii
C00003672	83-46-5	(-)-beta-Sitosterol	C29H50O	414.38616622	Tripterygium wilfordii
C00003737	638-95-9	alpha-Amyrin	C30H50O	426.38616622	Tripterygium wilfordii
C00003738	559-70-6	beta-Amyrin	C30H50O	426.38616622	Tripterygium wilfordii
C00003755	111-02-4	All-trans-squalene	C30H50	410.39125159	Tripterygium wilfordii
C00003759	50802-21-6	Maytenin	C28H36O3	420.26644502	Tripterygium wilfordii
C00004565	520-18-3	Kaempferol	C15H10O6	286.04773805	Tripterygium wilfordii
C00004631	117-39-5	Quercetin	C15H10O7	302.04265268	Tripterygium wilfordii

C_ID	CAS ID	Metabolite	Molecular formula	Mw	Species	C_ID_url
C00001411	87-52-5	Gramine	C11H14N2	174.115698	Acer saccharinum	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00001411
C00002728	458-36-6	Coniferaldehyde	C10H10O3	178.062994	Acer saccharinum	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00002728
C00002775	4206-58-0	Sinapaldehyde	C11H12O4	208.073559	Acer saccharinum	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00002775
C00006798	140447-90-1	Cyanidin 3-(2"-galloyl)-glucoside	C28H25O15	601.119345	Acer saccharinum	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00006798
C00000152	7400-08-0	p-Coumaric acid	C9H8O3	164.047344	Acorus calamus	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00000152
C00000619	97-53-0	Eugenol	C10H12O2	164.08373	Acorus calamus	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00000619
C00000633	73036-51-8	Acoradin	C24H32O6	416.219889	Acorus calamus	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00000633
C00000805	80-56-8	alpha-Pinene	C10H16	136.125201	Acorus calamus	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00000805
C00000816	127-91-3	beta-Pinene	C10H16	136.125201	Acorus calamus	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00000816
C00000819	76-22-2	(+)-Camphor	C10H16O	152.120115	Acorus calamus	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00000819
C00000823	138-86-3	Limonene	C10H16	136.125201	Acorus calamus	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00000823
C00000843	3338-55-4	cis-beta-Ocimene	C10H16	136.125201	Acorus calamus	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00000843
C00000844	106-22-9	Citronellol	C10H20O	156.151415	Acorus calamus	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00000844
C00000845	106-24-1	(E)-geraniol	C10H18O	154.135765	Acorus calamus	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00000845
C00000862	3779-61-1	trans-beta-Ocimene	C10H16	136.125201	Acorus calamus	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00000862
C00002715	5273-86-9	beta-Asarone	C12H16O3	208.109944	Acorus calamus	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00002715
C00002731	458-37-7	Curcumin	C21H20O6	368.125988	Acorus calamus	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00002731
C00002739	487-11-6	Elemicin	C12H16O3	208.109944	Acorus calamus	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00002739
C00002760	6379-72-2	Methylisoeugenol	C11H14O2	178.09938	Acorus calamus	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00002760
C00002902	537-42-8	E-Pterostilbene	C16H16O3	256.109944	Acorus calamus	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00002902
C00002903	501-36-0	Resveratrol	C14H12O3	228.078644	Acorus calamus	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00002903
C00003028	507-70-0	Borneol	C10H18O	154.135765	Acorus calamus	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00003028
C00003029	79-92-5	Camphene	C10H16	136.125201	Acorus calamus	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00003029
C00003035	141-27-5	trans-Citral	C10H16O	152.120115	Acorus calamus	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00003035
C00003036	106-26-3	cis-Citral	C10H16O	152.120115	Acorus calamus	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00003036
C00003046	105-87-3	Geranyl acetate	C12H20O2	196.14633	Acorus calamus	https://www.knapsackfamily.com/knapsack_core/top.php?input_type=all&input_word=Tripterygium_wilfordii&C_ID=C00003046

- 通过在Knapsackfamily数据库 (https://www.knapsackfamily.com/knapsack_core/top.php) 上依次传入物种名，可以直接获取该物种的所有代谢产物相关信息，包括代谢物名称，分子式，物种名以及该条代谢物的详情页链接等。通过爬取该数据库，共收集了8836条代谢物基本信息。

Basic Information

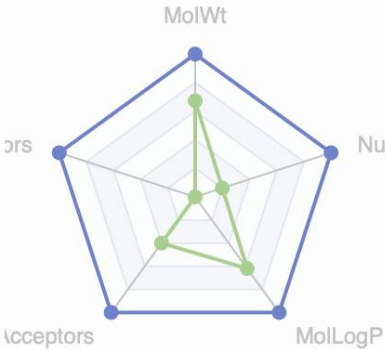
ADMET_Solubility	-5.541
ADMET_Solubility_Level	2
ADMET_BBB	0.261
ADMET_BBB_Level	1
ADMET_EXT_CYP2D6	0.868871
ADMET_EXT_CYP2D6#Prediction	1
ADMET_EXT_CYP2D6_Applicability#MD	16.6417
ADMET_EXT_Hepatotoxic	4.36137
ADMET_EXT_Hepatotoxic#Prediction	1
ADMET_EXT_Hepatotoxic_Applicability#MD	14.7925
ADMET_EXT_BBB	0.70015

- 随后想根据TCMBank数据库中提供的ADMET相关指标，来筛选代谢产物，但是因为目前缺乏统一的评价标准来综合解读这些指标，且该数据库中仅仅提供的数据，也并没有告知该数据高低的具体意义或影响，所以无法通过ADMET来筛选代谢产物。

Drug-likeness

Lipinski rule-of-five ? :

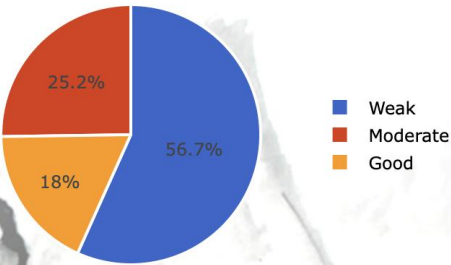
Term	Value	Judgment
The average molecular weight (MolWt)	336.367	Yes
Number of Hydrogen Bond Donors (NumHDonors)	0	Yes
Number of Hydrogen Bond Acceptors (NumHAcceptors)	4	Yes
Wildman-Crippen LogP value (MolLogP)	3.096	Yes
Number of Rotatable Bonds (NumRotatableBonds)	2	Yes



Quantitative estimation of drug-likeness ? : 0.674

Drug-likeness Grading

(Click each sector to see related infos)



Search



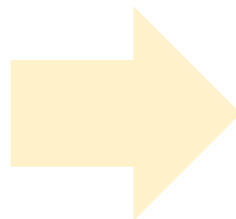
Ingredient Name in English 1↓	Molecular Formula 1↓	Drug-likeness Weight 1↓	Drug-likeness Grading 1↓
Dimethyl Trisulfide	C2H6S3	0.531	Moderate
Di-2-Butyl Disulfide	C8H18S2	0.623	Moderate
Gummosin	C24H30O5	0.464	Weak
Polyanthinin	C27H34O5	0.354	Weak

➤ 随后我继续调研多个数据库，想要找到一个能够有效筛选代谢产物的综合性指标。结果发现，“类药性”是多个数据库中普遍存在的参数，但是该指标主要基于口服小分子的经验规则建立。若将其作为统一筛选标准，则会排除大量结构独特且生物活性显著的天然产物，所以还是不能以类药性作为标准。

2. 收集8836个代谢产物的SMILES(一种线性表示分子结构的方法)

Metabolite Information

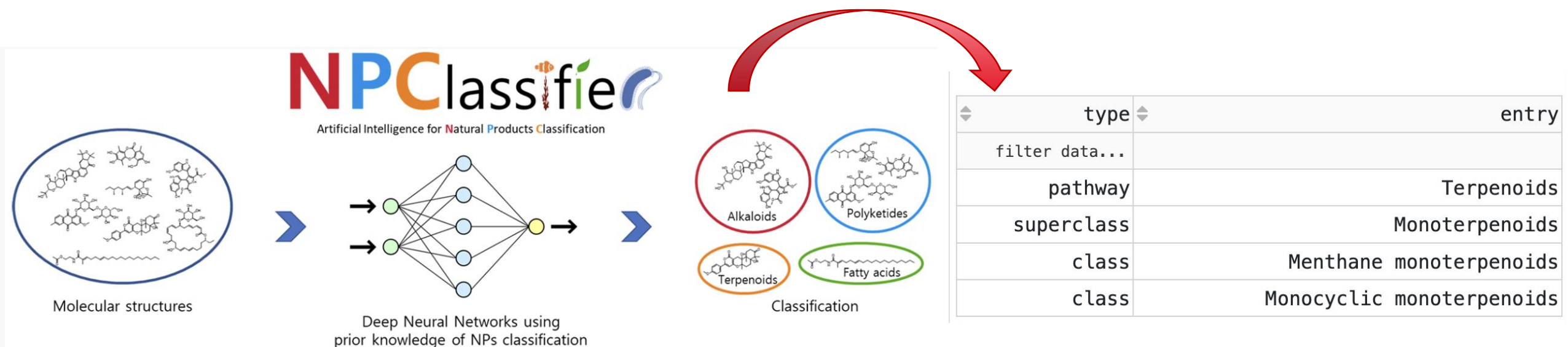
Name	Benzoic acid
Formula	C7H6O2
Mw	122.03677944
CAS RN	65-85-0
C_ID	C00000207 , Twins (%)
InChIKey	WPYMKLBDIGXBTP-UHFFFAOYSA-N
InChICode	InChI=1S/C7H6O2/c8-7(9)6-4-2-1-3-5-6/h1-5H,(H,8,9)
SMILES	<chem>O=C(O)c1ccccc1</chem>
Start Substs in Alk. Biosynthesis (Prediction)	



1	Metabolite	SMILES
2	Gramine	<chem>CN(C)Cc1c[nH]c2ccccc12</chem>
3	Coniferaldehyde	<chem>COc1cc(/C=C/C=O)ccc1O</chem>
4	Sinapaldehyde	<chem>COc1cc(/C=C/C=O)cc(OC)c1O</chem>
5	Cyanidin 3-(2"-galloylglucoside)	<chem>O=C(O)[C@@H]1C(O)[C@H](O)C(CO)O[C@H]1O</chem>
6	p-Coumaric acid	<chem>O=C(O)/C=C/c1ccc(O)cc1</chem>
7	Eugenol	<chem>C=CCc1ccc(O)c(OC)c1</chem>
8	Acoradin	<chem>COc1cc(OC)c(C2C(C)C(C)C2c2cc(OC)c(OC)cc2OC)</chem>
9	alpha-Pinene	<chem>CC1=CCC2CC1C2(C)C</chem>
10	beta-Pinene	<chem>C=C1CCC2CC1C2(C)C</chem>
11	(+)-Camphor	<chem>CC1(C)[C@@H]2CC[C@@]1(C)C(=O)C2</chem>
12	Limonene	<chem>C=C(C)C1CC=C(C)CC1</chem>
13	cis-beta-OCimene	<chem>C=C/C(C)=CCC=C(C)C</chem>
14	Citronellol	<chem>CC(C)=CCCC(C)CCO</chem>
15	(E)-geraniol	<chem>CC(C)=CCC/C(C)=C/CO</chem>
16	trans-beta-OCimene	<chem>C=C/C(C)=C/CC=C(C)C</chem>
17	beta-Asarone	<chem>C/C=Cc1cc(OC)c(OC)cc1OC</chem>
18	Curcumin	<chem>COc1cc(/C=C/C(=O)/C=C(O)/C=C/c2ccc(O)c(O)c2)cc1</chem>
19	Elemicin	<chem>C=CCc1cc(OC)c(OC)c(OC)c1</chem>
20	Methylisoeugenol	<chem>C/C=C/c1ccc(OC)c(OC)c1</chem>
21	E-Pterostilbene	<chem>COc1cc(/C=C/c2ccc(O)cc2)cc(OC)c1</chem>
22	Resveratrol	<chem>Oc1ccc(/C=C/c2cc(O)cc(O)c2)cc1</chem>
23	Borneol	<chem>CC1(C)C2CCC1(C)C(O)C2</chem>
24	Camphene	<chem>C=C1C2CCC(C2)C1(C)C</chem>
25	trans-Citral	<chem>CC(C)=CCC/C(C)=C/C=O</chem>
26	cis-Citral	<chem>CC(C)=CCC/C(C)=CC=O</chem>
27	Geranyl acetate	<chem>CC(=O)OC/C=C(C)CCC=C(C)C</chem>
28	(+)-Linalool	<chem>C=C[C@](C)(O)CCC=C(C)C</chem>

- 和王猛老师讨论后，我们决定先把筛选标准放一边，先对这些代谢产物进行系统分类。于是我先继续在KKnapsackfamily数据库中收集了这些代谢产物的SMILES信息，该信息将作为后续分类工作的数据基础。

3. 查询8836个代谢产物参与的路径以及分类



- 依次向NPCClassifier(<https://npclassifier.ucsd.edu/#>)中传入代谢产物对应的SMILE值，可以得到该代谢产物对应的通路，分类等信息；
- NPCClassifier中的pathway主要包括七类：脂肪酸、聚酮、莽草酸-苯丙素、萜类、生物碱、氨基酸/肽和碳水化合物；
- NPCClassifier中的superclasses代表了生源途径Pathway中的子类别，目前提出了70类，主要体现了化合物的化学性质或分类学相关信息；
- Superclasses进一步被细分为Class。主要代表一些特定的化合物种类、特征官能团等。

8836个代谢产物的分类情况

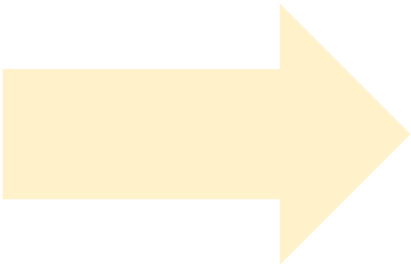
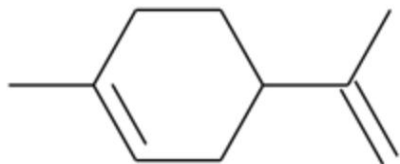
1	Metabolite	SMILES	pathway	superclass	class
2	Gramine	CN(C)Cc1c[nH]c2ccccc12	Shikimates and Phenylpropanoids	Phenylpropanoids (C6-C3)	Cinnamic acids and derivatives
3	Coniferaldehyde	COc1cc(/C=C/C=O)ccc1O	Shikimates and Phenylpropanoids	Phenylpropanoids (C6-C3)	Cinnamic acids and derivatives
4	Sinapaldehyde	COc1cc(/C=C/C=O)cc(OC)c1O	Alkaloids	Tryptophan alkaloids	Simple indole alkaloids
5	Cyanidin 3-(2"-galloylglucoside)	O=C(O[C@@H]1C(O)[C@H](O)C(CO)O[C@H]1O	Shikimates and Phenylpropanoids	Phenylpropanoids (C6-C3)	Cinnamic acids and derivatives
6	p-Coumaric acid	O=C(O)/C=C/c1ccc(O)cc1	Shikimates and Phenylpropanoids	Flavonoids	Anthocyanidins
7	Eugenol	C=CCc1ccc(O)c(OC)c1	Shikimates and Phenylpropanoids	Phenylpropanoids (C6-C3)	Cinnamic acids and derivatives
8	Acoradin	COc1cc(OC)c(C2C(C)C(C)C2c2cc(OC)c(OC)cc2OC)	Terpenoids	Monoterpenoids	Pinane monoterpenoids
9	alpha-Pinene	CC1=CCC2CC1C2(C)C	Terpenoids	Monoterpenoids	Pinane monoterpenoids
10	beta-Pinene	C=C1CCC2CC1C2(C)C	Terpenoids	Monoterpenoids	Camphane monoterpenoids
11	(+)-Camphor	CC1(C)[C@@H]2CC[C@@]1(C)C(=O)C2	Shikimates and Phenylpropanoids	Lignans	Minor lignans Neolignans
12	Limonene	C=C(C)C1CC=C(C)CC1	Terpenoids	Monoterpenoids	Menthane monoterpenoids Monocycl
13	cis-beta-Ocimene	C=C/C(C)=CCC=C(C)C	Terpenoids	Monoterpenoids	Acyclic monoterpenoids
14	Citronellol	CC(C)=CCCC(C)CCO	Terpenoids	Monoterpenoids	Acyclic monoterpenoids
15	(E)-geraniol	CC(C)=CCC/C(C)=C/CO	Shikimates and Phenylpropanoids	Diarylheptanoids	Linear diarylheptanoids
16	trans-beta-Ocimene	C=C/C(C)=C/CC=C(C)C	Terpenoids	Monoterpenoids	Acyclic monoterpenoids
17	beta-Asarone	C/C=Cc1cc(OC)c(OC)cc1OC	Polyketides		
18	Curcumin	COc1cc(/C=C/C(=O)/C=C(O)/C=C/c2ccc(O)c(O	Terpenoids	Monoterpenoids	Acyclic monoterpenoids
19	Elemicin	C=CCc1cc(OC)c(OC)c(OC)c1	Shikimates and Phenylpropanoids	Phenylpropanoids (C6-C3)	Cinnamic acids and derivatives
20	Methylisoeugenol	C/C=C/c1ccc(OC)c(OC)c1	Shikimates and Phenylpropanoids	Phenylpropanoids (C6-C3)	Cinnamic acids and derivatives
21	E-Pterostilbene	COc1cc(/C=C/c2ccc(O)cc2)cc(OC)c1	Shikimates and Phenylpropanoids	Stilbenoids	Monomeric stilbenes
22	Resveratrol	Oc1ccc(/C=C/c2cc(O)cc(O)c2)cc1	Terpenoids	Monoterpenoids	Camphane monoterpenoids Fenchan
23	Borneol	CC1(C)C2CCC1(C)C(O)C2	Terpenoids	Monoterpenoids	Acyclic monoterpenoids
24	Camphene	C=C1C2CCC(C2)C1(C)C	Terpenoids	Monoterpenoids	Camphane monoterpenoids
25	trans-Citral	CC(C)=CCC/C(C)=C/C=O	Shikimates and Phenylpropanoids	Stilbenoids	Monomeric stilbenes
26	cis-Citral	CC(C)=CCC/C(C)=CC=O	Terpenoids	Monoterpenoids	Acyclic monoterpenoids

统计各个Pathway的出现频率

NP Classifier

Version - 1.5

C=C(C)C1CC=C(C)CC1



pathway	count
terpenoids	2799
shikimates and phenylpropanoids	2532
fatty acids	1236
alkaloids	1141
amino acids and peptides	410
carbohydrates	278
polyketides	272

type	entry
filter data...	
pathway	Terpenoids
superclass	Monoterpenoids
class	Menthane monoterpenoids
class	Monocyclic monoterpenoids

➤ 8836个代谢产物的pathway有7条，分别是：terpenoids; shikimates and phenylpropanoids; fatty acids; alkaloids; amino acids and peptides; carbohydrates; polyketides

统计各个Superclass的出现频率并绘制词云图

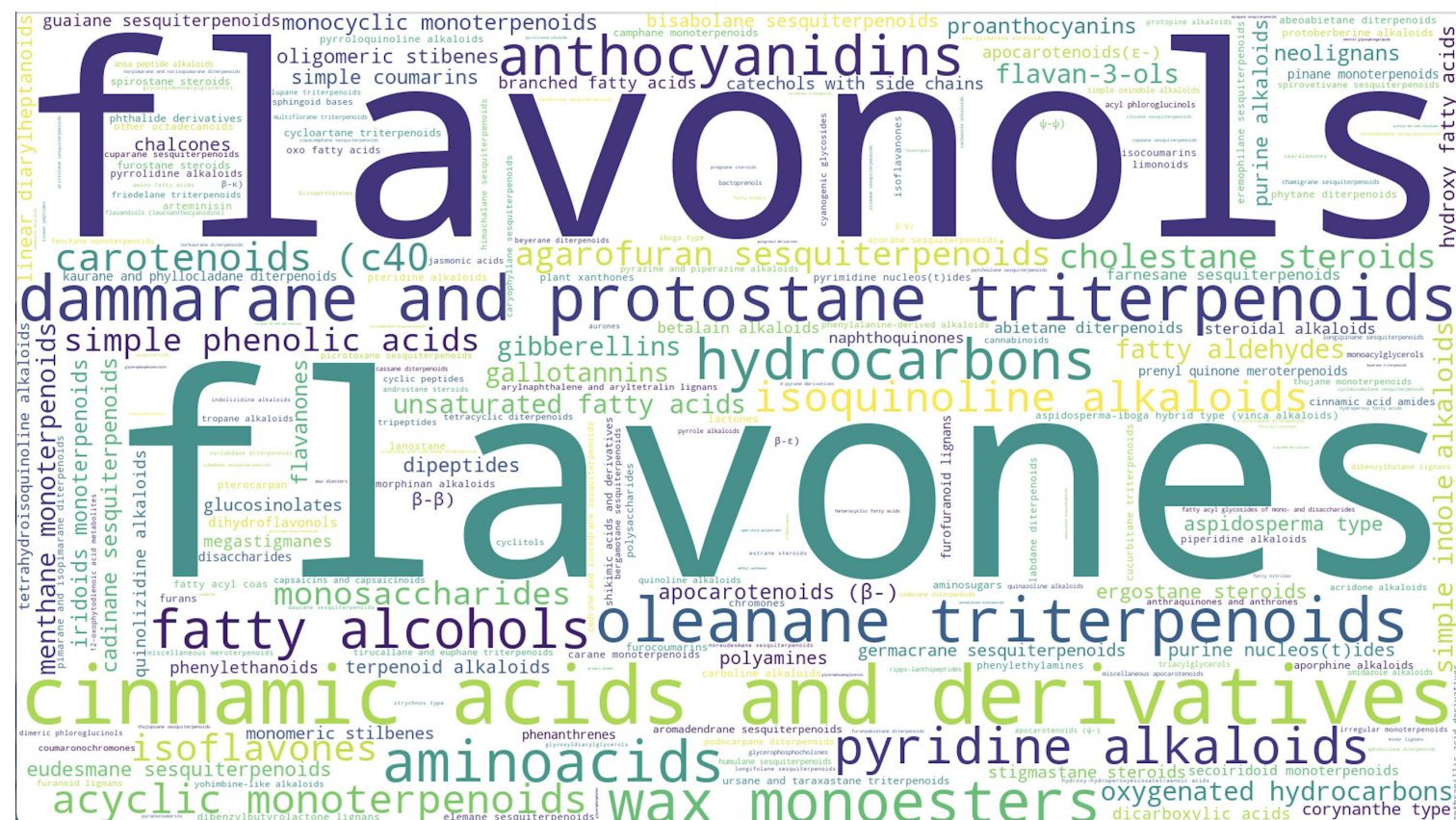
superclass	count
flavonoids	1176
sesquiterpenoids	728
triterpenoids	614
fatty acyls	574
monoterpenoids	501
tryptophan alkaloids	292
diterpenoids	291
phenylpropanoids (c6-c3)	285
fatty acids and conjugates	269
small peptides	267
fatty esters	264
steroids	259
tyrosine alkaloids	209
pseudoalkaloids	201
phenolic acids (c6-c1)	199
isoflavonoids	174
lignans	173
apocarotenoids	161
nicotinic acid alkaloids	150
saccharides	143
carotenoids (c40)	123
stilbenoids	122
coumarins	97
ornithine alkaloids	87
lysine alkaloids	76
nucleosides	76
amino acid glycosides	65



- 出现频率最高的10个superclass是：flavonoids; sesquiterpenoids; triterpenoids; fatty acyls; monoterpenoids; tryptophan alkaloids; diterpenoids; phenylpropanoids; fatty acids and conjugates

统计各个Class的出现频率并绘制词云图

A	B
class	count
flavones	315
flavonols	306
cinnamic acids and derivatives	252
dammarane and protostane triterpenoids	248
oleanane triterpenoids	231
wax monoesters	210
aminoacids	194
hydrocarbons	189
anthocyanidins	185
fatty alcohols	172
pyridine alkaloids	149
isoquinoline alkaloids	139
acyclic monoterpeneoids	138
carotenoids (c40)	125
isoflavones	116
agarofuran sesquiterpenoids	112
cholestane steroids	112
simple phenolic acids	104
monosaccharides	104
oxygenated hydrocarbons	101
flavan-3-ols	100
gibberellins	99
unsaturated fatty acids	99
menthane monoterpeneoids	97
fatty aldehydes	96
simple indole alkaloids	95
gallotannins	93



- 出现频率最高的10个class是：flavones; flavonols; cinnamic acids and derivatives; dammarane and protostane triterpenoids; oleanane triterpenoids; wax monoesters; aminoacids; hydrocarbons; anthocyanidins; fatty alcohols

目前存在的问题

- ▶目前虽然收集了8836个代谢产物，但是还没有对这些代谢产物进行筛选，还没找到比较好的筛选方法；
- ▶用NPClassfier工具虽然可以对代谢产物进行分类，但其结果基于SMILES字符串的算法推断，存在一定的不确定性，最好进行验证。

下周计划

- ▶查阅文献或者调研数据库，看看有没有和代谢产物合成通路相关的数据库，或者有没有介绍相关通路的文献