Doctoral Dissertation

Prediction of metabolite activities by repetitive clustering of the structural similarity based networks

Nobutaka Wakamatsu

June 4, 2020

Division of Information Science
Graduate School of Science and Technology
Nara Institute of Science and Technology

A Doctoral Dissertation submitted to the Division of Information Science,
Nara Institute of Science and Technology
In partial fulfillment of the requirements for the degree of
Doctor of Engineering.

Thesis Committee:

Professor Shigehiko Kanaya	(Supervisor)
Professor Keiichi Yasumoto	(Co-supervisor)
Associate Professor Md. Altaf-Ul-Amin	(Co-supervisor)
Associate Professor Naoaki Ono	(Co-supervisor)
Assistant Professor Ming Huang	(Co-supervisor)

Prediction of metabolite activities by repetitive clustering of the structural similarity based networks

Nobutaka Wakamatsu

Abstract

A number of studies have investigated the relations between structures and activities of metabolites. It has been proposed that structural similarity between metabolites implies activity similarity between them. In light of this fact we propose a method for activity prediction of secondary metabolites based on association philosophy. First we determined the structural similarity scores between targeted metabolite pairs using COMPLIG algorithm. To increase the possibility of clusters over represented with known metabolites, we calculated structural similarity between metabolite pairs for which activities of both metabolites or at least one metabolite are known and then selected the metabolite pairs for which the similarity score is higher than a threshold ($s > 0.95$). The network of such metabolite pairs was then clustered using the DPClusO algorithm. Statistically significant cluster-activity pairs were then selected using the hypergeometric test. Then biological activities of unannotated metabolites were predicted from the activity of metabolites included in the statistically overrepresented clusters. After the first round of biological activity prediction, we considered the predicted metabolites as known metabolites and repeated the prediction process until no new prediction was made. Finally, we were able to predict biological activity of 945 metabolites with unknown biological activity. Furthermore, we investigated the activity-activity relationship by examining the activity prediction trends of this study.

*Doctor's Thesis, Department of Information Science, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD1761021, May 25, 2020

Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1. Background

Numerous chemical reactions of various types called metabolic reactions are continuously occurring in living organisms. It is to gain energy, and to obtain molecules that make up cell, and to relate to the environment to increase the survival probability.

Molecules biosynthesized by these metabolic reactions are called metabolites.

Metabolites are divided into primary and secondary metabolites according to their role.

Primary metabolites are metabolites essential for maintaining normal physiological processes such as cell growth, development, and reproduction. The primary metabolites are amino acids, sugars, nucleic acids, etc. Many organisms biosynthesize primary metabolite by primary metabolic pathway.

Secondary metabolites are metabolites that are not classified as primary metabolites. Secondary metabolites are mainly biosynthesized by bacteria, fungi, and plants, and many are endemic to the species. Deficiencies in secondary metabolites do not result in immediate death, but affect the survival and health of the organism over an extended period of time. In some case, no major change is observed. Secondary metabolites are important to adapt to the environment and increase the probability of survival.

Through evolution process, organisms acquired pathways that biosynthesize diverse metabolites. Evolutionary pressures have led production of various secondary metabolites and constructed pathways to improve fitness for survival of organisms. Metabolites that are involved in growth, development, and reproduction of an organism are called “primary metabolites” and their production is referred as the “central metabolic pathway” which are key components in maintaining normal physiological processes. On the other hand, “secondary metabolites” are remaining compounds, that, although important, are not essential for the survival of an organism under suitable environment.

Secondary metabolites are commonly utilized for survival strategies under adverse conditions and for other purposes. For example, some organisms produce and use secondary metabolites that are repellents against bacteria, fungi, amoeba, plants, insects, and large animals. Also, some organisms produce hormones for symbiosis with

microorganisms, plants, nematodes, insects, and higher animals.

Secondary metabolites are also used in signaling, or, regulation of various metabolic pathways. Secondary metabolites often help an organism maintain various interactions with the environment, adapting for the environmental needs [1,2]. Some secondary metabolites in plants serves as pigments, e.g., flowers display various colors so that they can attract pollinators and also certain coloring of leaves or stems can defend against attack by animals. Therefore, secondary metabolites can have biological and pharmaceutical activities that have various effects on different organisms. In this paper we refer secondary metabolites functions as metabolite activities or activities.

In species-metabolite relational database KNApSAcK, currently there are 111,199 records involving 50,899 metabolites and 22,350 species. However, the activities of only 3,210 metabolites are recorded in KNApSAcK metabolite activity database. Given the importance of metabolites in agriculture, ecology and healthcare it would be of great signigicance to predict the functions of the other metabolites included in the KNApSAcK database. In this regard a computational method would be less expensive and fast approach.

In the current work we propose to develop a computational method to predict the functions of a huge set of metabolites whose functions are unknown, utilizing structural similarity networks. Such networks contain relations between metabolites concerning chemical similarity of metabolites.

1.2. Objectives

Research into secondary plant metabolism gained steam in the latter half of the 19th century. Earlier researches only focused on categorizing the secondary plant metabolites but did not give much emphasis on the actual function of the secondary plant metabolites. In the first half of the 20th century, the main research around secondary plant metabolism was dedicated to the formation of secondary metabolites in plants and related pathways. One of the most ground breaking ideas of 1970s argued that plant secondary metabolites evolved in relation to environmental conditions and therefore they play an indispensable role in the survival of a plant in its environment. To understand the interaction of plants with environment, the interdependence of various organisms and ecology in general, it is first necessary to know the functions of metabolites, specially the secondary metabolites. Knowing the functions of secondary metabolites will also help to analyze the evolution of metabolic pathways and diversity of enzymes in plants.

Conventional agricultural industry relies on a wide use of chemical pesticides and fertilizers. However, increased demand for organic products shows that consumers prefer reduced chemical use. Volatile organic compounds (VOCs) are special type of metabolites. VOCs emitted by bacteria and fungi might have the potential to be alternatives to the use of chemical pesticides to protect plants from pests and pathogens. microbial VOCs are seen as biocontrol agents to control various phytopathogens and as biofertilizers for plant growth promotion. Furthermore, metabolites are deeply involved in human health care. The use of VOCs as biomarkers to detect human diseases is rapidly increasing. Plant metabolites are a major sources of drugs and they are widely used in pharmaceutical industries. The above discussion implies the importance of function prediction of huge set of metabolites.

Studies have been conducted to detect complete sets of metabolites in tissues and cells as well as to understand species-species relationships based on secondary metabolites [3-6]. Other studies focused on the biological interactions between organisms in ecosystems

[7] and plant interactions in the environment, such as pest and pathogen defense compounds, and in UV-B sunscreens [8-10].

Metabolome studies have also been extended from model species such as crops and medicinal plants to health care and medicinal systems based on the interactions between organisms and humans. Secondary metabolites are important components in the drug discovery process, as such biodiversity produces a variety of small molecules and provides a source of new chemicals that are potential drug leads. The secondary metabolites synthesized by plants, fungi and microorganisms are diverse, e.g. there are at least 30,000 terpenoids[11], 9,000 flavonoids, 1,600 isoflavonoids and 12,000 alkaloids[12]. These metabolites have been deposited in the species-metabolite relational database(DB) known as the KNApSAcK Core DB, which contains 111,199 species-metabolite relationships involving altogether 50899 metabolites and 22,350 species [13]. Out of 50,899 metabolites of the KNApSAcK database the activity of only 3,161 metabolites are recorded. Given the importance of metabolites in agriculture, ecology and healthcare it would be of great help to various sectors to predict the activities of the unknown metabolites by using some computational means. However, a computational method would be less expensive and fast approach. Therefore, as a first step it is wise to develop and apply a computational approach which is the focus of our work.

Previously it was reported that structural similarity between metabolites imply similar activities. Several study used primary, secondary and tertiary structure similarities between proteins for predicting protein functions (*e.g. Nature Reviews Molecular Cell Biology 8.12 (2007): 995-1005*). At metabolite level also there have been a number of researches showing activity structure relationships (*e.g. Free radical biology and medicine 20.7 (1996): 933-956*). It has been reported that structural similarity between metabolites implies similar functions. Ohtana et al. showed that structurally similar groups of metabolites are rich with similar activity metabolites [14]. Volatile organic compounds are special types of metabolites. Azian et al. linked the structures of VOCs

with their activities [15]. In this work, we propose a network based approach for prediction of metabolite activities using structural similarity.

Network clustering algorithms have been successfully applied in different fields in recent years. In systems biology, network clustering has been applied in protein-protein interaction networks, gene expression networks and metabolic pathways. In text mining, content related document searching is done using clustering of bipartite networks consisting of different document types and frequently used keywords in the documents. In a social network, finding a common group of people with different shared interest are analyzed by network biclustering. Many more such applications of network clustering can be mentioned. In the present work, we applied network clustering for activity prediction of metabolites.

In summary, the purpose of the present proposal is to develop a computational method for function prediction of metabolites by clustering the structural similarity based network utilizing the concepts of network clustering and False Discovery Rate (FDR).

1.3 Outline of the dissertation

The rest of this dissertation is organized as follows. Chapter 2 explains in detail the methods adopted in this thesis for activity prediction of metabolites. Chapter 3 presents the results of the experiments conducted for activity prediction. Chapter 4 discusses on activity-activity relations in the context of structural similarity of metabolites. Finally chapter 5 provides conclusions of the present thesis.

Chapter 2 Materials and Methods

2.1. Dataset

We collected molecular structure data of 50,037 metabolites recorded in KNApSAcK Core database [16]. We obtained molecular structures from KNApSAcK database as MOL formats and input to COMPLIG algorithm [17] for calculating structural similarity between metabolites. The MOL format file contains information such as the coordinates of atoms constituting the molecule in a three-dimensional space and the presence or absence of bonds between the atoms. Although this MOL file also contains information on the positions and bonds of hydrogen constituting molecules, this study does not consider hydrogen data for comparing molecular structures between metabolites. This is because the comparison of molecular structures of COMPLIG is a graph matching algorithm, and the time required for calculation increases as the number of atoms increase. From KNApSAcK Metabolite Activity database, we downloaded 9,809 secondary metabolite biological activity relationships involving in total 155 types of activity categories.

2.2. Activities of secondary metabolites

Secondary metabolites have various types of biological activity. Nakamura et al. classified 140 biological activities into two broad categories [18]. One is a biological activity type related to medicine and human health care classified into 92 activity categories (M01-M092 in Table 1) based on the KEGG DRUG DB(http:// www.genome.jp/kegg/drug/). The other is concerning chemical ecology, are classified into 48 categories (E01-E48 in Table 1). Later on, 15 other activities and related metabolites were included in KNApSAcK metabolite activity database. These activities (O1-O15) are mostly related to medicine and human healthcare. In Table 1, we show the names of the biological activities. and corresponding number of records included in KNApSAcK metabolite activity database.

Table 1. Statistics for the 155 activity categories

General description	Activity category	No. of records
Plant growth regulator (976)	Enhance germination (E01)	26
	Enhance stem growth (E02)	138
	Enhance root growth (E03)	46
	Enhance leaf growth (E04)	48
	Enhance flowering (E05)	42
	Enhance fruiting (E06)	44
	Enhance plant growth (E07)	136
	Inhibit germination (E08)	28
	Inhibit stem growth (E09)	51
	Inhibit root growth (E10)	45
	Inhibit leaf growth (E11)	29
	Inhibit flowering (E12)	12
	Inhibit fruiting (E13)	3
	Inhibit plant growth (E14)	92
	Allelopathic (E15)	71
	Phytoalexin (E16)	165
Attractant/repellent (288)	Feeding attractant (E17)	48
	Feeding deterrent (E18)	85
	Pollinator attractant (E19)	45
	Oviposition attractant (E20)	26
	Oviposition deterrent (E21)	3
	Sex attractant (E22)	22
	Attractant (E23)	39
	Repellent (E24)	20
Selective toxicity (166)	Phytotoxic (E25)	45
	Herbicidal (E26)	7
	Insecticidal (E27)	79
	Acaricidal (E28)	3
	Molluscicidal (E29)	12
	Piscicidal (E30)	15
	Nematocidal (E31)	5
Antimicrobial agent (2230)	Antibacterial (E32)	1109
	Antituberculotic (E33)	77
	Antileprotic (E34)	7

	Antifungal (E35)	670
	Inhibit spore germination (E36)	15
	Antimicrobial (E37)	352
Antiviral agent (261)	Antiviral (E38)	114
	Antihepatitic (E39)	4
	Anti-HIV (E40)	102
	Anti-HSV (E41)	41
Antiparasitic agent (340)	Anthelmintic (E42)	55
	Antiprotozoal (E43)	13
	Antiamebic (E44)	14
	Antimalarial (E45)	143
	Antileishmanial (E46)	37
	Antitrypanosomal (E47)	76
	Pediculicidal (E48)	2
Nervous system agent (464)	Antipyretic (M01)	33
	Analgesic (M02)	95
	Antiarthritic (M03)	8
	Anesthetic (M04)	24
	Sedative (M05)	69
	Antispasmodic (M06)	77
	Anticonvulsant (M07)	11
	Antidementic (M08)	1
	Antidepressant (M09)	15
	CNS stimulant (M10)	35
	Diaphoretic (M11)	6
	Emetic (M12)	5
	Antiemetic (M13)	9
	Antigout (M14)	6
	Antimigraine (M15)	1
	Antimyasthenic (M16)	8
	Antiparkinson (M17)	13
	Antipsychotic (M18)	19
	Muscle relaxant (M19)	29
Cardiovascular agent (611)	Antidiabetic (M20)	88
	Hemostatic (M21)	27
	Antithrombotic (M22)	54

	Cardiotonic (M23)	60
	Antiarrhythmic (M24)	25
	Diuretic (M25)	17
	Antihypertensive (M26)	215
	Antihyperlipidemic (M27)	39
	Antianemic (M28)	5
	Other cardiovascular agent (M29)	81
Respiratory tract ahent (122)	Antitussive (M30)	44
	Expectorant (M31)	13
	Antiasthmatic (M32)	55
	Other respiratory tract agent (M33)	10
Digestive organ agent (184)	Antidiarrheic (M34)	12
	Carminative (M35)	5
	Stomachic (M36)	3
	Laxative (M37)	38
	Choleretic (M38)	20
	Antihepatotoxic (M39)	58
	Other digestive organ agent (M40)	48
Genitourinary agent (55)	Oxytocic (M41)	21
	Antifertility (M42)	22
	Abortifacient (M43)	5
	Other genitourinary agent (M44)	7
Anticancer agent (1377)	Antioxidant (M45)	368
	Anticancer (M46)	771
	Antitumor (M47)	138
	Antineoplastic (M48)	88
	Antimutagenic (M49)	12
Anti-inflammatory agent (446)	Anti-inflammatory (M50)	279
	Antiallergic (M51)	99
	UV shield (M52)	7
	Antidermatitic (M53)	19
	Antiedemic (M54)	42
Immunological agent (32)	Immunosuppressant (M55)	11
	Immunostimulant (M56)	11
	Immunomodulative (M57)	10
Nutrient (78)	Nucleic acid (M58)	4

	Essential amino acid (M59)	15
	Nonessential amino acid (M60)	8
	Vitamin (M61)	15
	Nutrient (M62)	26
	Tonic (M63)	10
Nontherapeutic agent (336)	Solvent (M64)	4
	Flavor (M65)	131
	Odor (M66)	65
	Pigment (M67)	98
	Emulsifying agent (M68)	2
	Antiseptic (M69)	36
Other health agent (206)	Antiulcerogenic (M70)	33
	Depilatory (M71)	6
	Antidote (M72)	18
	Hormonal (M73)	20
	Dental (M74)	5
	Other health agent (M75)	124
Narcotic (13)	Narcotic (M76)	13
Toxic (1033)	Phototoxic (M77)	33
	Neurotoxic (M78)	52
	Pneumotoxic (M79)	33
	Hepatotoxic (M80)	42
	Cytotoxic (M81)	464
	Toxic (M82)	409
Tumorigenic (87)	Tumorigenic (M83)	30
	Mutagenic (M84)	33
	Genotoxic (M85)	5
	Teratogenic (M86)	19
Other disease-causing agent (104)	Psychotomimetic (M87)	11
	Hemolytic (M88)	8
	Allergenic (M89)	25
	Irritant (M90)	36
	Dermatitic (M91)	21
	Edematous (M92)	3
Others (400)	Amnestic (O1)	1
	Anti-anxiety (O2)	7

Anti-muscle rigidity (O3)	1
Anti-tremor (O4)	1
Anticholinergic (O5)	18
Anticholinesterase (O6)	97
Anxiogenic (O7)	1
Biomarker (O8)	136
Convulsant (O9)	31
Defense (O10)	44
Hallucinogenic (O11)	15
Hypnotic (O12)	4
Induce tremor (O13)	10
Inhibit CYP (O14)	30
Other nervous system agent (O15)	4

2.3. Flow of the proposed method

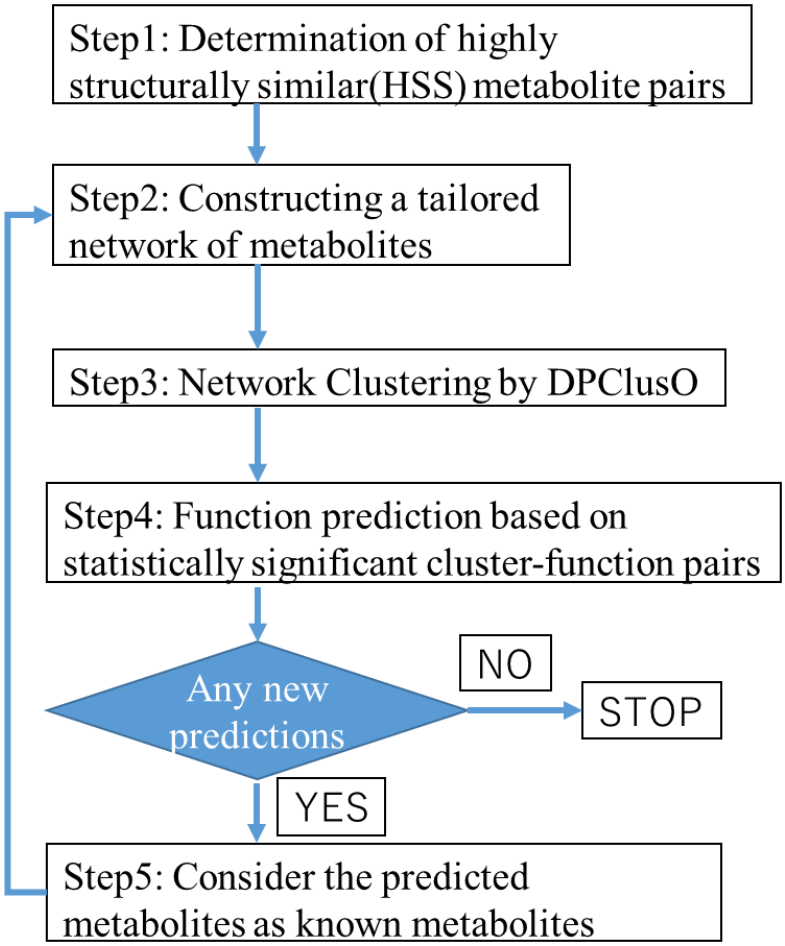


Figure 1. Flow of the proposed method

Figure 1 shows the flow of the proposed activity prediction method. First, we determined highly structurally similar metabolite pairs based on COMPLIG algorithm[17] and then constructed a restricted network of secondary metabolites. Next, we extracted clusters from the network by overlapping graph clustering algorithm DPClusO[19-21], and validated the relationship between cluster and bioactivity by calculating p-values using a hypergeometric distribution. Finally, we select statistically significant cluster-activity pairs based on false discovery rate (FDR) analysis and predict activities of some activity unknown metabolites. Below we discuss each step of the flow in detail.

2.4. Calculate molecular structural similarity between metabolites

We used fast heuristic graph match algorithm COMPLIG to calculate 3D-chemical similarity of secondary metabolites. This algorithm compares the 3-dimentional structures of small molecules. COMPLIG has been used for the systematic classification of PDB ligands [17]. Usually, a fingerprint method using a MACCS structure key or the like is used for comparing the molecular structures of small molecule ligands. In this method, the characteristics of a molecule (having an amino group, five-membered ring, etc.) are represented by a 1/0 bit string, and the ratio of bits shared between molecules is determined. Finger print method can be compared in a short time become the structure between two molecules is compared by comparing bit strings. However, it is not possible to obtain a correspondence between atoms such as which atom of the molecule B corresponds to a certain atom of molecule A.

COMPLIG expresses the molecular structure using a graph with atoms as nodes and chemical bonds as edges, and searches for a graph of nodes and edges between two molecules. As a result, atomic correspondence between molecules can be obtained, and it is considered that more accurate structural comparison can be performed. Therefore, in the present work we used COMPLIG algorithm for measuring structural similarity. Structure similarity is measured by COMPLIG based on three factors namely, topology-distance, bond-order, and rotatable-bonds not only type of atoms, presence or absence of bonds, and the type of bonds. The similarity score S(A, B) between two metabolites A and B is estimated using the following Equation:

$$\mathrm{S(A,B)} = \frac{M(A,B)}{\max(N(A),N(B)}$$

Where M(A, B) is the number of atoms and bonds matched between metabolites A and B, and N(A) and N(B) are the total numbers of atoms and bonds in metabolites A and B, respectively. Metabolite pairs for which similarity score is higher than a threshold (0.95) are considered as highly structurally similar metabolite pairs in this study.

2.5. Network construction based on molecular similarity

Our objective is to predict activities of metabolites by network clustering based on relationship between similarity of molecular structure and similarity of biological activity. The hypothesis is that if a certain percentage of metabolites in a cluster has a certain biological activity, the unknown metabolites in the cluster are considered to have the same biological activity, and that biological activity is assigned.

For the purpose of generating clusters rich with known metabolites we constructed a tailored network of metabolites i.e. we considered the metabolite pairs where the activity of both or at least one metabolite is known and both are highly similar in terms of structure. All such metabolite pairs are used as the edges of the tailored network of metabolites which is used for activity prediction.

2.6. Network Clustering by the DPClusO algorithm

We used a network-clustering algorithm DPClusO to cluster the metabolites into groups based on 3D-structure similarities among them. DPClusO generates clusters characterized by high density of network and identified by periphery in the networks [19,20,21]. This algorithm has been used in several big data analyses in molecular biology such as protein-protein interaction [22], metabolomics [14] and prediction of functional relations between genes using gene expression data [23]. DPClus/DPClusO has been previously utilized in versatile omics research e.g. protein complex prediction [19], disease gene prediction [24] etc. DPClusO generates overlapping clusters and ensures coverage. In case of the secondary metabolites addressed in this study, one metabolite may have more than one biological activity. In order to perform more diverse biological activity predictions, we used DPClusO, which can create overlapped clusters, for clustering. There are two important parameters in this algorithm, which are, density d_k and cluster property cp_{nk}. Density d_k of any cluster *k* is the ratio of the number of edges present in the cluster($|E|$) and the maximum possible number of edges in the cluster($|E|_{max}$). The cluster property of node *n* with respect to cluster *k* is represented by

$$cp_{nk} = \frac{|E_{nk}|}{d_k \times |N_k|}$$

N_k is the number of nodes in cluster k. E_{nk} is the total number of edges connecting the node n with node of cluster k.

2.7. Activity prediction based on statistically significant cluster-activity pairs

We extracted clusters of secondary metabolites from the structural similarity network using network clustering algorithm DPClusO in order to extract similar metabolites of the 3D-structure as a cluster. We validated the relationship between cluster and biological activity by calculating p-value of a cluster with respect to an activity according to hypergeometric distribution using the following equation.

$$p = 1 - \sum_{i=0}^{k-1} \frac{\binom{F}{i}\binom{N-F}{C-i}}{\binom{N}{C}}$$

Here, N represents total number of metabolites in the network, C is the number of metabolites in a cluster, F is the total number of metabolites of a certain activity category in the network and k is the number of metabolites of that activity category in the cluster. Thus we can calculate a p-value for each cluster-activity pair. Finally, we selected statistically significant cluster-activity pairs based on False Discovery Rate(FDR) analysis [25,26].

We assessed accuracy of the prediction by random cross validation, using a part of the activity known metabolites as test data. We randomly sampled 10% of the known metabolites as a test data set and predicted their activity to evaluate accuracy of our function prediction method. We repeated this process 500 times and computed the average accuracy.

2.8. Function prediction by iteration

The biological activity was predicted by the above method, and the biological activity was assigned to activity unknown metabolites. We considered the predicted activities of the activity unknown metabolites as true or known activities and go back to step 2 of Figure 1 and continued the iteration until a new prediction was made.

Chapter 3 Prediction of Metabolite activities

3.1. Links between structural and activity similarities

We have 50,037 metabolites in our primary data out of which 3,161 are activity known metabolites. To assess overall relation between structure and activity, initially we calculated structural similarity between all possible metabolite pairs involving the activity known metabolites, i.e. 50,037(50,037 - 1)/2=1,251,825,666 pairs, using the COMPLIG algorithm. Next, we examined the proportion of metabolite pairs having the same biological activity as the molecular structure similarity of the metabolite pairs. We sorted the metabolite pairs in order of descending molecular structural similarity and then selected subsets of the metabolite pairs above different minimum threshold similarities i.e. 0.90, 0.91, …, 0.99, and calculated the percentage of metabolite pairs with the same activity within each subset. Figure 2 implies that the higher the structural similarity between a metabolite pair the higher is the probability that their activities are similar. Based on this finding, we can hypothesize that a high structural similarity between two metabolites indicates a high probability of activity similarity between them.

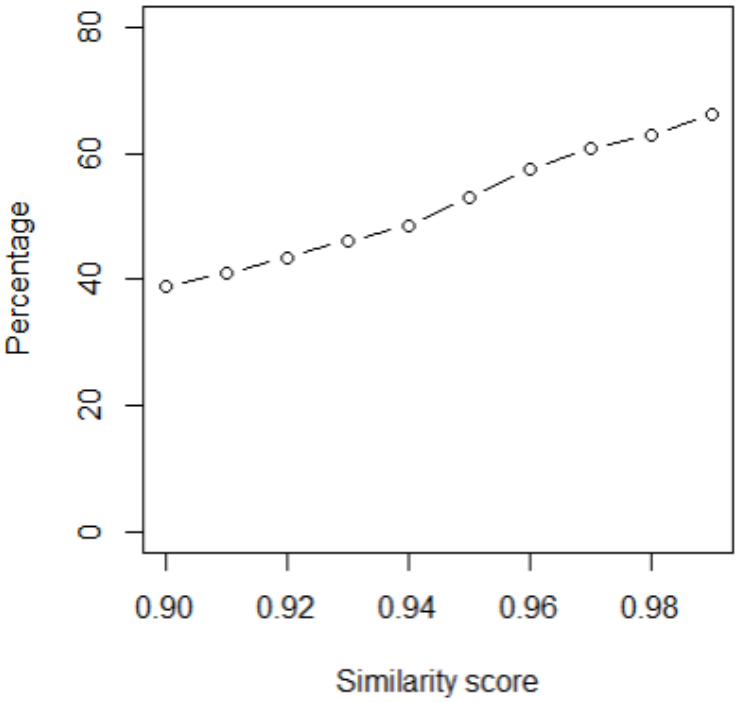


Figure 2. Percentage of the metabolite pairs with same activity (corresponding to activity known metabolites) is increasing with respect to increasing structural similarity score.

3.2. Structural similarity based network of metabolites

We constructed a simple network where nodes are the metabolites and edges represent COMPLIG similarity between them with score more than 0.95. In our data we have there are 3,161 activity known metabolites and 46,876 activity unknown metabolites. Similarity of molecular structure was calculated by COMPLIG for all possible combinations of the two metabolites in dataset and selected the pairs with similarity score higher than 0.95, which are represented by the edges of the network of metabolites. The network constructed by the selected metabolite pairs contained 37,119 nodes and 177,179 edges. Among them, 2,140 activity known metabolite nodes, 34,979 activity unknown metabolite nodes, 1,925 edges connecting activity known metabolite nodes, 20,557 edges connecting activity known metabolite node and activity unknown metabolite node, and 154,697 edges connecting activity unknown metabolite nodes. The activity known metabolite nodes in the network accounted for about 5.8% of the total nodes, and the edges connecting both ends or one side with activity known metabolite node were about 12.7% of the total edge. This network has a large proportion of nodes and edges of metabolites not related to biological activity, and it is considered that it is difficult to predict biological activity. Therefore, in order to create a network with more information on biological activity, select a metabolite pair in which both or one of the metabolites has a known biological activity from a metabolite pairs with a molecular structural similarity score of 0.95 or more and create network. The objective of selecting such pairs is to increase the possibility of getting more known metabolites with similar activity in high density clusters of the network. The resulted network consisted of 10,773 metabolite nodes, of which 2,140 activity known metabolite nodes and 8,633 activity unknown metabolite nodes. This network contains 22,482 edges. Also, this network is not a single large network but consists of 837 components. Figure 3 shows the largest connected component visualized by the Cytoscape software [27]. The Global properties of the network are as follows: Average degree: 4.174, Clustering coefficient: 0.305, average path length: 4.934, diameter: 18. The degree distribution of the network is shown in Figure 4 which indicates its compliance to the power law[28] i.e. the network is a scale free

network. A scale free network with clustering coefficient 0.305 is likely to contain high density clusters representing cohesive groups of metabolites.

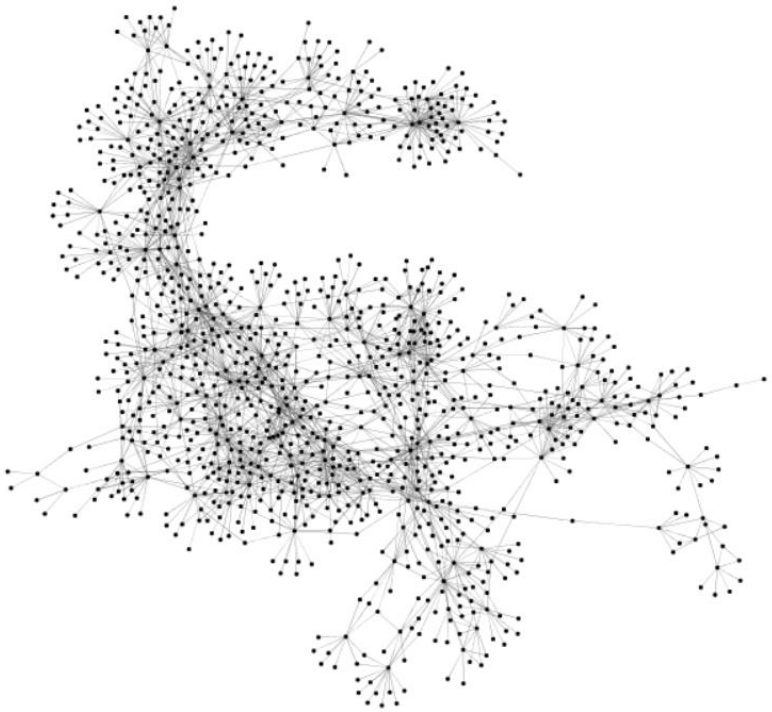


Figure 3. Visualization of the largest component of the initial structure similarity based network of metabolites.

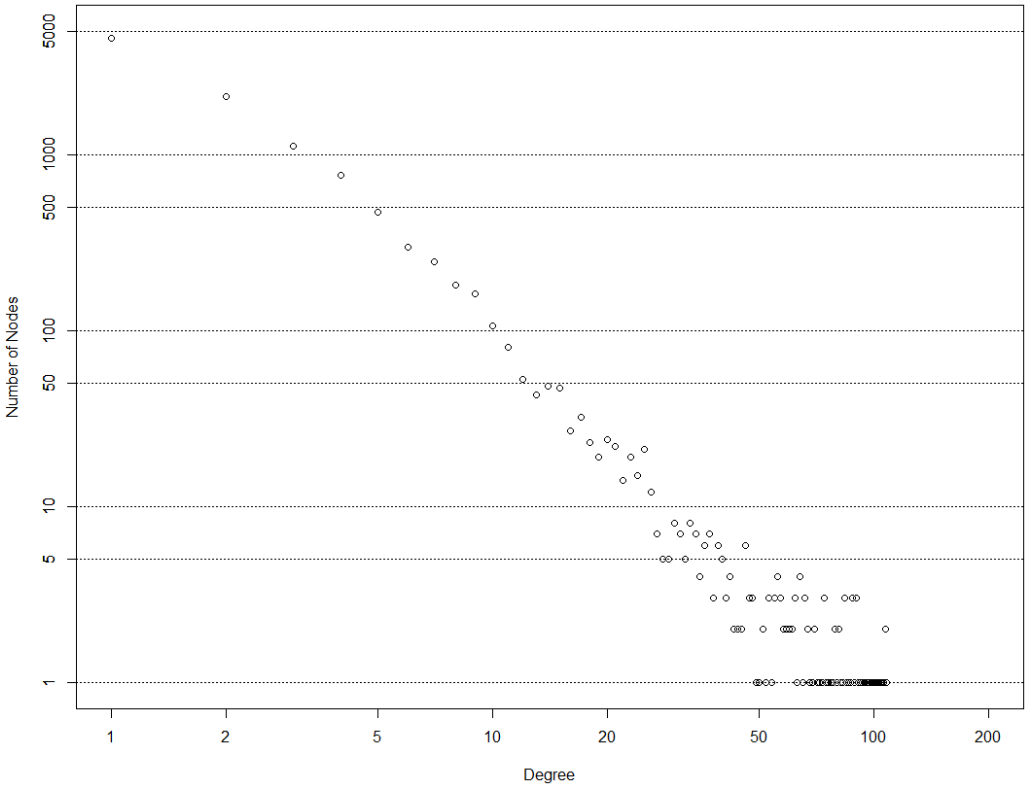
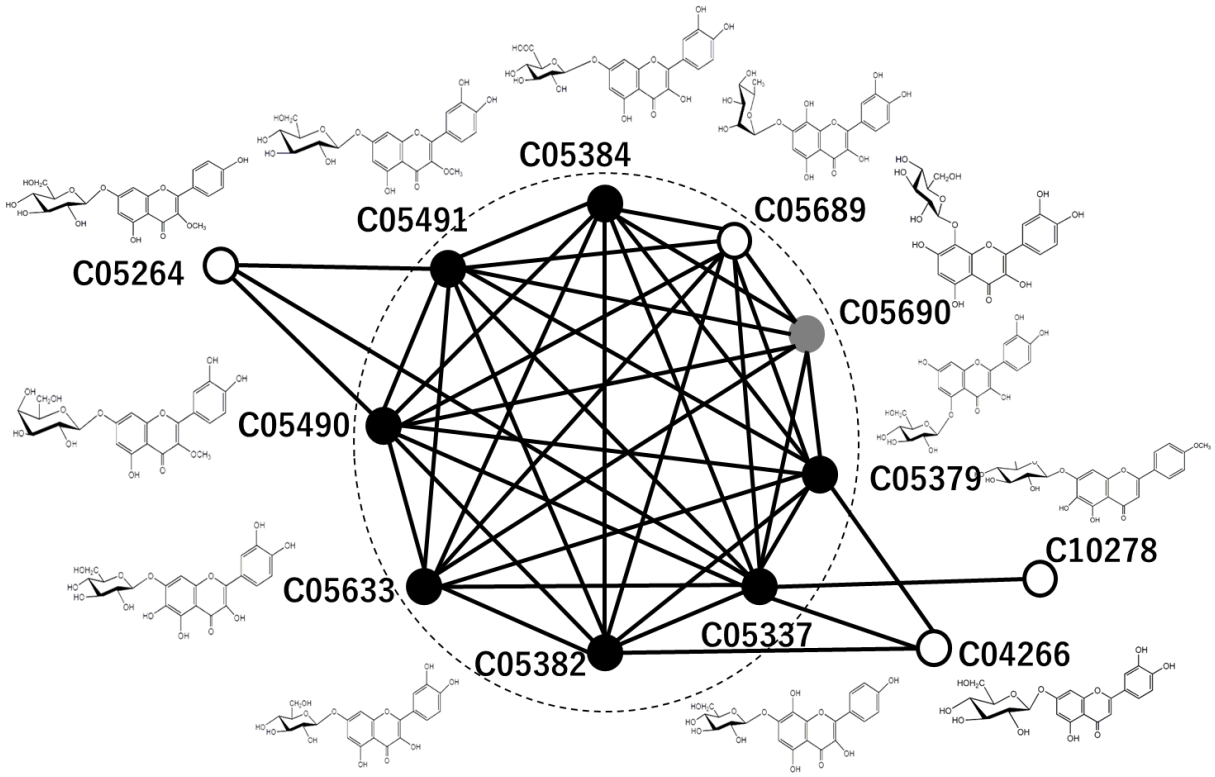


Figure 4. Degree distribution of the initial network

3.3. Clustering by the DPClusO

We then clustered the network using the DPClusO algorithm. The following values are used as the input parameters of the clustering using the DPClusO: cp=0.5, din=0.9. The DPClusO generates overlapping clusters we finally selected 2,739 clusters for subsequent analysis. The largest cluster consists of 12 metabolites and the smallest clusters consists of 3 metabolites and average size of the clusters is 7.2. An example of the obtained cluster is indicated by a dotted line in Figure 5. This cluster contains nine metabolites, seven of which have the Pollinator attractant activity (Substances that attract pollen-borne insects). This cluster and Pollinator attractant activity are related with p-value = 2.22e-16. This cluster includes activity unknown metabolite C05689(Gossypetin 7-rhamnoside). We can predict that C05689 is also related to Pollinator attractant activity (Figure 5).

Figure 5. The cluster consisting of the pollinator attractants



3.4. FDR analysis and activity prediction

We selected only the clusters that are statistically significantly related with biological activities for activity prediction purpose. To do so we first determined the p-values of the clusters based on hypergeometric distribution in the context of each of the 155 activity categories. We determined the p-value of a cluster with respect to category if there exists at least one metabolite of that category in the cluster. Thus, we obtained 27,345 cluster-activity pairs with p-values. We plotted the frequency of cluster-activity pairs with respect to p-values (Figure 6).

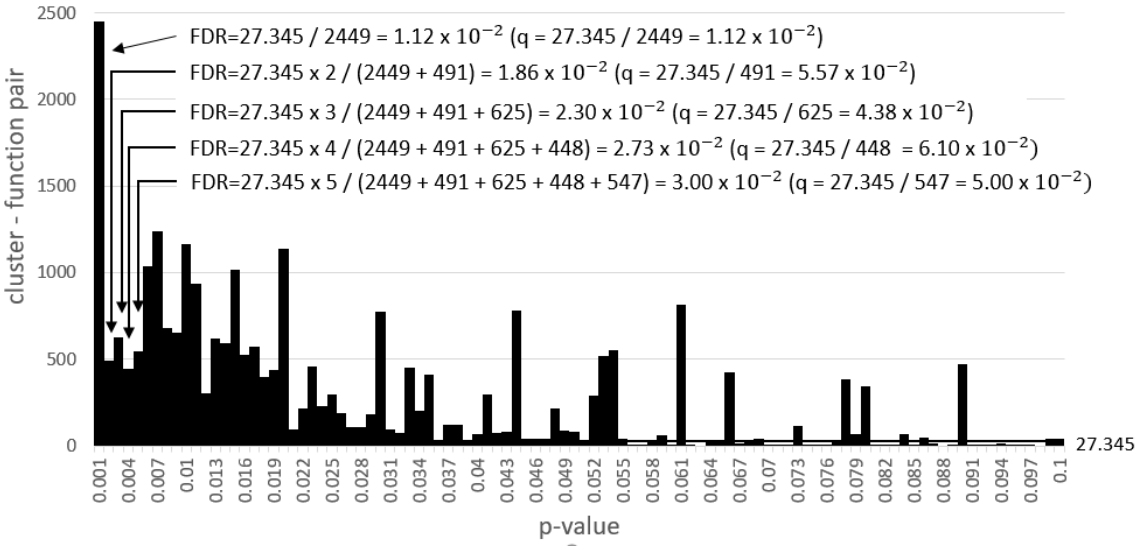


Figure 6. Frequencies of cluster-activity pairs with respect to p-value

The p-values ranges from 0 to 1 and this range was divided into 1,000 intervals. The abscissa in Figure 6 is also divided by the same range and interval for p-values. Therefore in each interval the theoretical count of random cluster-activity pairs is 27.345(27,345/1,000). In the interval from 0 to 0.001 in p-value, we obtained 2,449 cluster-activity pairs which correspond to $q = 1.12 \times 10^{-2}$ in FDR estimated by the Benjamini-Hochberg[25] because theoretical counts of false positive were estimated as 27.345(27,345×0.001). FDR thus become 0.0186 for $p < 0.002$, and 0.02319 for $p < 0.003$. We selected cluster-activity pairs with FDR < 0.05 and finally used those cluster-activity pairs for activity prediction.

3.5. Assessment of prediction accuracy

We assessed our prediction accuracy by considering part of the activity known metabolites as activity unknown metabolites. We selected randomly 10% of the activity known metabolites and consider selected metabolites as activity unknown metabolites. And then conducted the process from Step 3 to Step 5 of Figure 1 and verified whether they have been predicted correctly or not. We repeated this process 500 times. The distribution of the accuracy achieved by 500 iterations is shown in Figure 7.

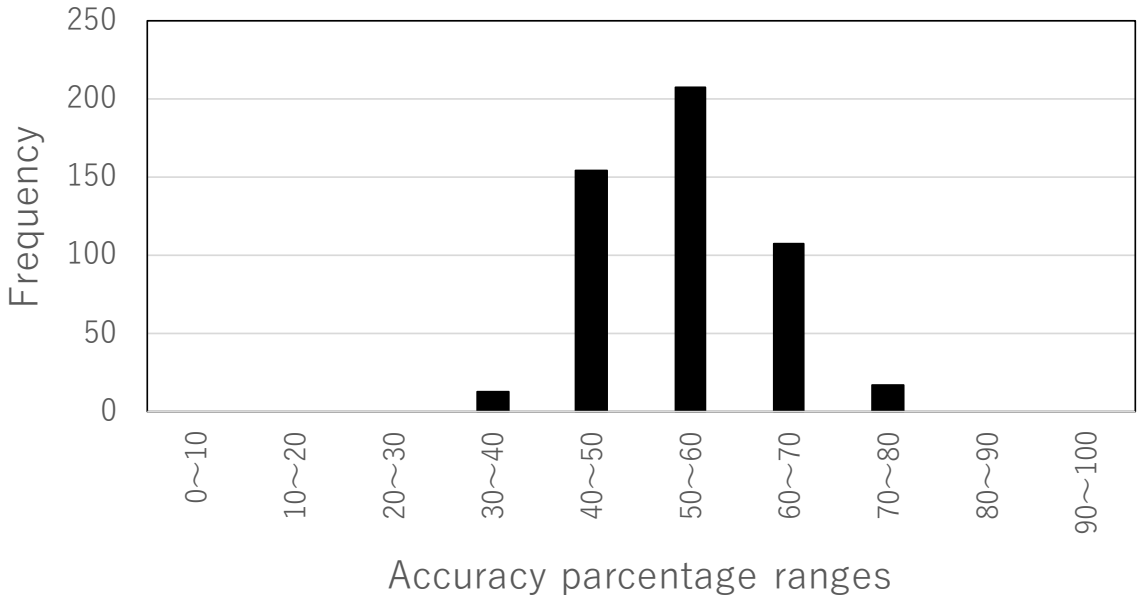


Figure 7. Frequencies of obtained accuracies among 500 experiments

In our dataset each activity known metabolite is associated to 1 to 7 activities. If metabolite has only one activity and if we randomly assign an activity to that metabolite, the probability of correct prediction is 1/155 = 0.00645 (155 different types of activities considered in the present work). Similarly, if a metabolite has 7 different activities and if we randomly assign an activity to that metabolite, the probability of correct prediction is 7/155 = 0.0451. This means by random assignment much less than 4.5% accuracy is achievable. Therefore, under the fact that we have 155 different types of activity category and unavailability of complete information of the activities of metabolites the average accuracy of 54.6% achieved by the proposed method is reasonably good.

Further, when the activity prediction is performed only based on the high similarity of the

molecular structure, the prediction accuracy is about 52.9% at the maximum, as shown Figure2.

3.6. Activity prediction of metabolites by iterations

After the first round of biological activity prediction, we considered the predicted metabolites as known metabolites and updated the datasets. Then, we repeated Step2 to Step4 of Figure 1. This makes it possible to further predict the biological activity after increasing the number of data. Such iteration of the activity prediction was performed until a new prediction result could not be obtained. This way repeated the activity prediction process 246 times. Finally predicted activities of 945 metabolites from the activity unknown metabolites. The list of these 945 metabolites and predicted activities are presented in Appendix A. In the method of this study, the activity prediction results for all metabolites of unknown biological activity could not be obtained. This is because a p-value that satisfies FDR < 0.05 was used as a threshold in order to reduce the possibility of false positives i.e. associating a biological activity with a metabolite not related to the activity. Since this prediction method considers low FDR, it is considered that there are few false positive results in the biological activity prediction obtained by the method of this study. Theoretically, 945×0.05 = 47.25 i.e. there might be about 47 false predictions out of these 945 predictions. Our future work is to investigate a new method for predicting activities of more metabolites.

Chapter 4 Relation between activities

4.1. Preliminaries

In our data set some individual metabolites have more than one activities. It is not surprising that similar structures may be associated to more than one activity resulting in some misclassifications in the prediction process. However, such misclassification affinities of some metabolites to certain activity categories can be utilized to assess activity-activity relations in the context of relevant structures.

4.2. Method

We utilized the results of the 500 iterative experiments we conducted for assessing the prediction accuracy to identify activity-activity relations. When the structural similarity of metabolites in two different activity categories are very close, the activities of the two categories may be also related to each other. To discover activity-activity relations we took into account the affinity of metabolites to be included in different activity categories

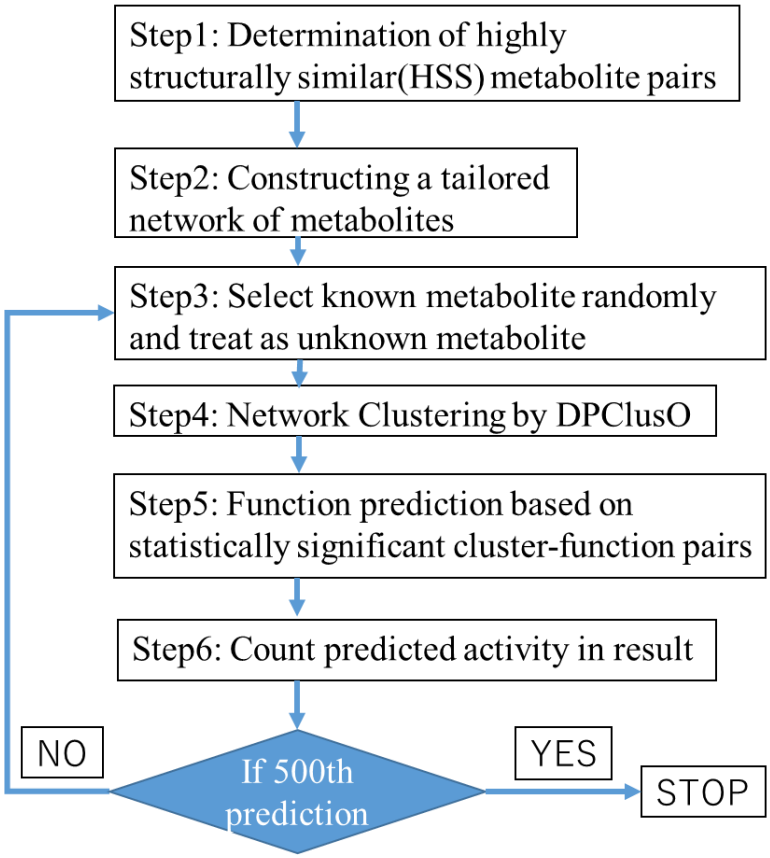


Figure 8. Flow of the method to count predicted activity in 500 predictions.

based on our experiments regarding prediction accuracy assessment. In total 500 iterations, we predicted activities of 1,340 metabolites. We have 155 activity categories in our data but activity prediction results were associated to only 79 categories. Therefore, we made a 1,340×79 matrix, say matrix M showing counts of which metabolites have been predicted as which category in 500 activity prediction iterations. Figure 8 shows flow to count each predicted activity in 500 prediction iteration.

We normalized this data using the following linear transformation and then conducted hierarchical clustering.

$$M'_{ij} = \frac{M_{ij} - \min(ith\ \ row)}{\max(ith\ \ row) - \min(ith\ \ row)}$$

And then applied Ward's method of hierarchical clustering using the aforementioned matrix and classified 79 activity categories.

4.3. Discussions

The dendrogram of the hierarchical clustering is shown in Figure 9. In order to divide into the optimal number of clusters by calculation, we tried to determine the number of clusters using GAP statistics. However, our data could not determine the optimal number of clusters. Therefore, based on the dendrogram we determined 9 activity clusters as shown in Figure 9. Figure 10 shows the result of GAP statistic calculation.

Before discussing the clusters, we provide in Table 2 short description of the 79 activities included in this part of the work i.e. activity-activity relation prediction.

Table 2. description of 79 biological activities

Activity	Description
Anticholinergic	An anticholinergic agent is a substance that blocks the neurotransmitter acetylcholine in the central and the peripheral nervous system. These agents inhibit parasympathetic nerve impulses by selectively blocking the binding of the neurotransmitter acetylcholine to its receptor in nerve cells.
Pneumotoxic	Having a toxic effect on the lungs.

Antiseptic	A chemical used for preventing infection in an injury, especially by killing bacteria.
Odor	A quality of something that stimulates the olfactory organ : scent.
Anthelmintic	Anthelmintics or antihelminthics are a group of antiparasitic drugs that expel parasitic worms (helminths) and other internal parasites from the body by either stunning or killing them and without causing significant damage to the host.
Antifungal	An antifungal medication, also known as an antimycotic medication, is a pharmaceutical fungicide or fungistatic used to treat and prevent mycosis such as athlete's foot, ringworm, candidiasis (thrush), serious systemic infections such as cryptococcal meningitis, and others.
Vitamin	A vitamin is an organic molecule (or related set of molecules) which is an essential micronutrient that an organism needs in small quantities for the proper functioning of its metabolism.
Expectorant	A medicine which promotes the secretion of sputum by the air passages, used to treat coughs.
Antiamebic	Destroying or suppressing the growth of amebas, or an agent that does this.
Emetic	(of a substance) Causing vomiting.
Antileishmanial	(pharmacology, immunology) Acting against Leishmania parasites.
Antitrypanosomal	(pharmacology) Countering trypanosomes.
Anesthetic	A substance that induces insensitivity to pain.
Narcotic	An addictive drug affecting mood or behaviour, especially an illegal one. The term narcotic originally referred medically to any psychoactive compound with sleep-inducing properties.
Allergenic	Having the capacity to induce allergy
Phytoalexin	A substance that is produced by plant tissues in response to contact with a parasite and specifically inhibits the growth of that parasite. Phytoalexins are antimicrobial and often antioxidative substances synthesized de novo by plants that accumulate rapidly at areas of pathogen infection.
Inhibit CYP	Inhibit CYP. Cytochromes are proteins containing heme as a cofactor. They are classified according to the type of heme and its mode of binding. Cytochrome function is linked to the reversible redox change from ferrous (Fe(II)) to the ferric (Fe(III)) oxidation

	state of the iron found in the heme core. In addition to the classification by the IUBMB into four cytochrome classes, several additional classifications such as cytochrome o and cytochrome P450 can be found in biochemical literature.
Pigment	the natural colouring matter of animal or plant tissue.
Allelopathic	A usually negative effect on the growth or development of an organism of one species, caused by a chemical released by an organism of another species. Usually used of plants.
Inhibit root growth	Inhibit root growth.
Anticonvulsant	(chiefly of a drug) Used to prevent or reduce the severity of epileptic fits or other convulsions.
Antipyretic	(chiefly of a drug) Used to prevent or reduce fever.
Attractant	A substance (such as a pheromone) that attracts specific animals (such as insects or individuals of the opposite sex).
Repellent	A substance that deters insects or other pests from approaching or settling.
CNS stimulant	A type of drug that increases the levels of certain chemicals in the brain and increases alertness, attention, energy, and physical activity. CNS stimulants also raise blood pressure and increase heart rate and breathing rate.
Inhibit leaf growth	Inhibit leaf growth
Antigout	Anti gout. Gout is a form of inflammatory arthritis that develops in some people who have high levels of uric acid in the blood.
Feeding deterrent	A secondary plant compound (or substance) that hinders a pest from subsequent food intake after an initial tasting.
Dermatitic	Pertaining to, or characteristic of dermatitis
Defense	For example, Defend against biotic stressors such as insects and pathogens, Indirect defense responses.
Antiparkinson	An antiparkinson medication is a type of drug which is intended to treat and relieve the symptoms of Parkinson's disease. Most of these agents act by either increasing dopamine activity or reducing acetylcholine activity in the central nervous system.
Antihypertensive	(of a drug) Used to lower high blood pressure.
Piscicidal	Acting as a piscicide; fatal to fish.
Inhibit stem growth	Inhibit stem growth

Convulsant	(chiefly of drugs) Producing sudden and involuntary muscle contractions.
Biomarker	A naturally occurring molecule, gene, or characteristic by which a particular pathological or physiological process, disease, etc. can be identified.
Antitussive	(especially of a drug) Used to prevent or relieve a cough.
Antispasmodic	(chiefly of a drug) Used to relieve spasm of involuntary muscle.
Antiprotozoal	Antiprotozoal drug, any agent that kills or inhibits the growth of organisms known as protozoans.
Antiarthritic	Of or relating to a medicine or therapy that relieves the symptoms of arthritis.
Antiarrhythmic	Antiarrhythmic agents, also known as cardiac dysrhythmia medications, are a group of pharmaceuticals that are used to suppress abnormal rhythms of the heart (cardiac arrhythmias), such as atrial fibrillation, atrial flutter, ventricular tachycardia, and ventricular fibrillation.
Antibacterial	Anything that destroys bacteria or suppresses their growth or their ability to reproduce.
Antihepatotoxic	That acts against hepatotoxicity
Phototoxic	Rendering the skin susceptible to damage (such as sunburn or blisters) upon exposure to light and especially ultraviolet light
Insecticidal	Destroying or controlling insects
Hormonal	Relating to or containing a hormone or hormones.
Choleretic	Choleretics are substances that increase the volume of secretion of bile from the liver as well as the amount of solids secreted.
Antimalarial	(of a drug) Used to prevent malaria.
Other health agent	Other health agent
Toxic	Poisonous substances
Antidote	A medicine taken or given to counteract a particular poison.
Antithrombotic	An antithrombotic agent is a drug that reduces the formation of blood clots (thrombi).Antithrombotics can be used therapeutically for prevention (primary prevention, secondary prevention) or treatment of a dangerous blood clot (acute thrombus).
Anticancer	Used against or tending to arrest or prevent cancer
Antiviral	(chiefly of a drug or treatment) Effective against viruses.

Laxative	(chiefly of a drug or medicine) Tending to stimulate or facilitate evacuation of the bowels.
Anti-HSV	Acting against a herpes virus or the symptoms caused by infection with a herpesvirus anti-herpes medication anti-herpes antibodies.
Antitumor	Inhibiting the growth of a tumour or tumours.
Pollinator attractant	Substances that attract pollinators(ex. bee)
Anti-inflammatory	(chiefly of a drug) Used to reduce inflammation.
Oviposition attractant	Attract the process of laying eggs(oviposition)
Antiallergic	Used to prevent an allergic response : tending to relieve or control allergic symptoms
Analgesic	(of a drug) Acting to relieve pain.
Antimicrobial	An antimicrobial is an agent that kills microorganisms or stops their growth. Antimicrobial medicines can be grouped according to the microorganisms they act primarily against.
Flavor	Flavor is the sensory impression of food or other substance, and is determined primarily by the chemical senses of taste and smell. The "trigeminal senses", which detect chemical irritants in the mouth and throat, as well as temperature and texture, are also important to the overall gestalt of flavor perception. The flavor of the food, as such, can be altered with natural or artificial flavorants which affect these senses.
Nutrient	A substance that provides nourishment essential for the maintenance of life and for growth.
Antidiabetic	A substance that helps a person with diabetes control their level of glucose (sugar) in the blood. Antidiabetic agents include insulin and the oral hypoglycemic agents.
Immunosuppressant	Immunosuppressive agents is drugs that inhibit or prevent activity of the immune system. Immunosuppressants are used to prevent rejection of a transplanted organ and to treat autoimmune diseases such as psoriasis, rheumatoid arthritis, and Crohn's disease. Some treatments for cancer act as immunosuppressants.
Mutagenic	Capable of inducing mutation or increasing its rate. In genetics, a mutagen is a physical or chemical agent that changes the genetic material, usually DNA, of an organism and thus increases the frequency of mutations above the natural background level.

Hepatotoxic	Damaging or destructive to liver cells.
Antioxidant	A substance that inhibits oxidation, especially one used to counteract the deterioration of stored food products. a substance such as vitamin C or E that removes potentially damaging oxidizing agents in a living organism.
Cytotoxic	Toxic to living cells.Toxic to cells. Any agent or process that kills cells. Chemotherapy and radiotherapy are forms of cytotoxic therapy. They kill cells.
Enhance plant growth	Enhance plant growth agent.(ex. stimulate cell divisionwith its stimulation of growth. ex2. intermediate in the shikimic acid pathway)
Enhance stem growth	Enhnce stem growth agent. (ex. induce leaf-sheath elongation. ex2. exhibit tuberization process)
Feeding attractant	Ex. feeding attractant on Polygonum species
Antihyperlipidemic	Acting to prevent or counteract the accumulation of lipids in the blood an antihyperlipidemic drug.
Antiasthmatic	Bronchodilators relax the smooth muscles that line the airway. This makes the airways open wider, letting more air pass through them.
Antituberculotic	A drug used to counter tuberculosis.
Hemostatic	Hemostasis is a process which causes bleeding to stop, meaning to keep blood within a damaged blood vessel. It is the first stage of wound healing. This involves coagulation, blood changing from a liquid to a gel.

Below we separately discuss on the 9 clusters indicated in the dendrogram of Figure 9.

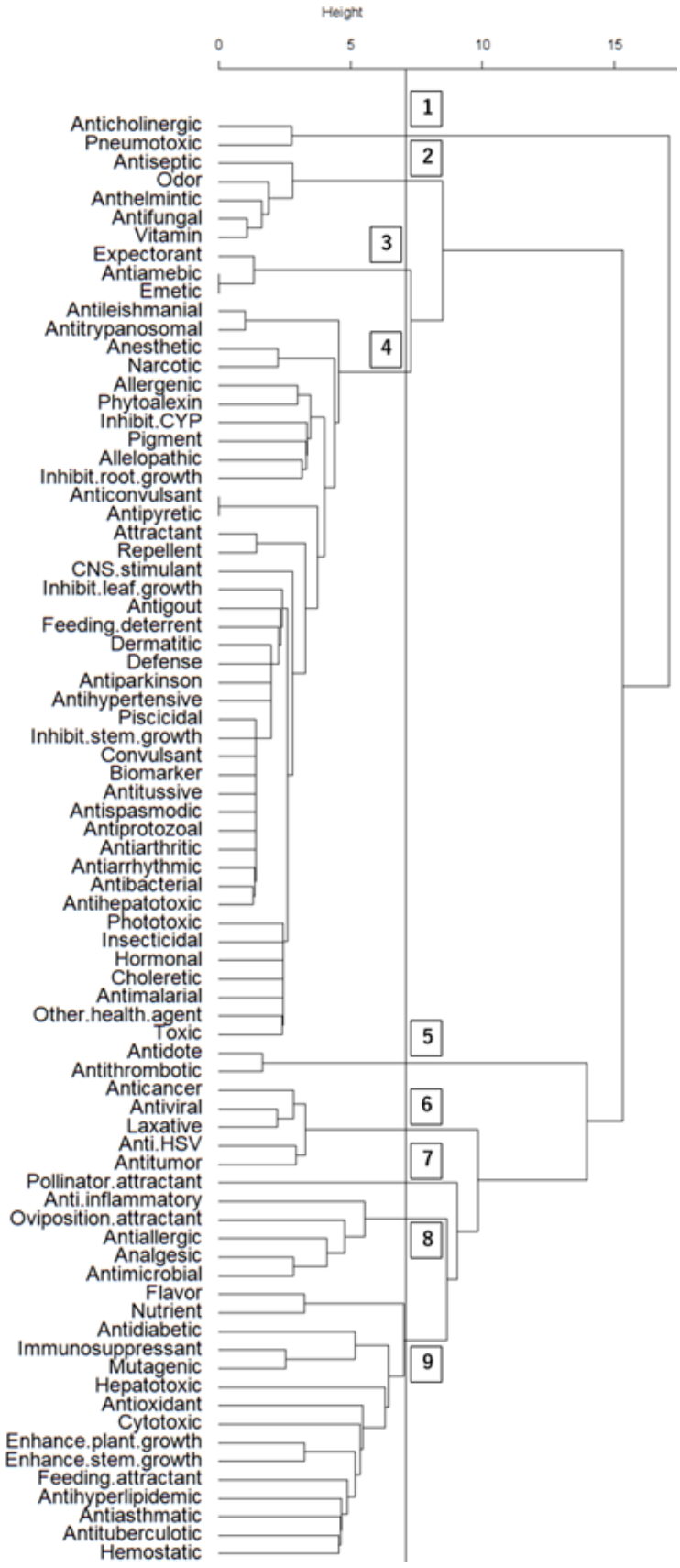


Figure 9. Dendrogram of activity –activity relations.

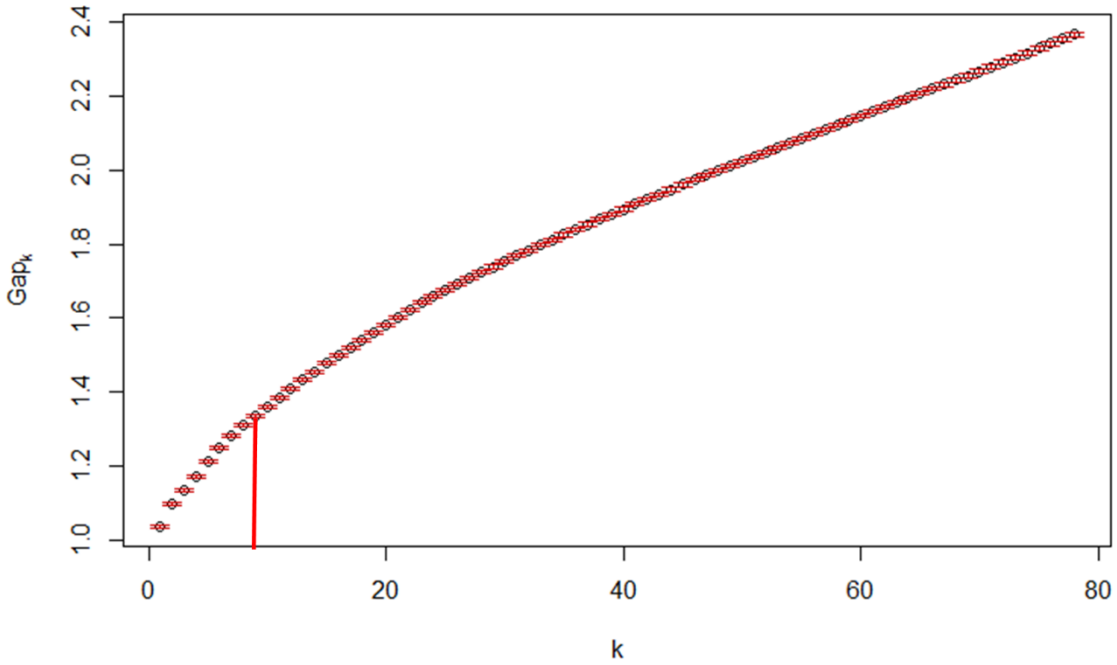


Figure 10. The result of GAP statistics calculations. k is number of clusters. Gap_k is GAP statistics based on k clusters.

Cluster 1

The activities included in cluster 1 are Anticholinergic and Pneumotoxic. Both are used as drugs for humans. Anticholinergic activity is an activity that inhibits neurotransmission by inhibiting the binding between acetylcholine and acetylcholine receptor. Pneumotoxic is toxicity to the lung. Both of these seem to be activities related to the effect and side effects of the drug. The molecular structures of activity known metabolites in the same cluster are similar. Therefore, metabolites with anticholinergic activities may also have pneumotoxic activity.

Cluster 2

This cluster has 5 activity categories which mostly act to kill parasites and microorganisms such as Antiseptic activity, Anthelmintic activity, Antifungal activity. Vitamin, which is a nutrient necessary for living things to live, is also included in this cluster. Another one is Odor which is volatile organic compound related to pleasant or an unpleasant scent/smell. In this cluster, Vitamin and Antifungal were the closest. There are

already drugs that contain vitamins A and E to enhance the effect of Antifungal activity.

Cluster 3

There are 3 activities in this cluster. Two of them Expectorant activity and Emetic activity are used as drugs acting on nerves. The other one is Antiamebic activity. Emetic and Antiamebic were the closest.

Cluster 4

This cluster is the largest cluster consisting of 40 activity categories. In this cluster, four elements are related to Antiparasitic agents, five are plant growth regulators, three are attractant/repellent, two are selective toxicity, four are antiparasitic agent, eight are nervous system agent, two are cardiovascular agent, two are digestive organ agent, two are other health agent, two are toxic and others are mostly health care related activities. Anticonvulsant and Antipyretic were the closest.

Cluster 5

This cluster has 2 activity categories both are related to human health. Those are Antidote and Antitumor. If the toxin causing tumors, removing the toxin may help the tumor to subside, so the two activities may be related.

Cluster 6

This cluster has 5 activity categories. Two of them are Anticancer and Antitumor. The other two are Antiviral agents. Another one is Laxative. Regarding the relationship between cancer and viruses, infection with hepatitis B or C virus may cause liver cancer. Therefore, by suppressing the growth of hepatitis virus with a drug having an antiviral effect, liver cancer can be eventually suppressed. Antiviral and Laxative activity is closest in this cluster. Laxative activity may also be related to the Antiviral activity in that it excretes the virus out of the body.

Cluster 7

Only Pollinator attractant activity belongs to cluster 7. The reason why they were classified into clusters that differed only on one activity is probably because their molecular structure were not so similar to the metabolites of other activity clusters.

Cluster 8

This cluster has 5 activity categories. Two of the activity categories are Anti-inflammatory and Antiallergic. Some metabolites with Antiallergic activity category have Anti-inflammatory activity. Two other related categories are Antimicrobial and Analgesic. Another less related category is Oviposition attractant. Analgesic and Antimicrobial were closest in this cluster.

Cluster 9

This cluster has 15 activity categories. Within this cluster Flavor and Nutrient activity categories are located close. Also Enhance plant growth and Enhance stem growth activity categories are located close. These two activity categories are involved in plant growth. Then Immunosuppressant and Mutagenic activity categories are located close. Some drugs that have Immunosuppressant activity also Mutagenic at certain concentrations. Two other similar activities Hepatotoxic and Cytotoxic are in this cluster. Also Antiasthmatic and Antitubeculotic are closely placed. The three bioactivity categories of Antidiabetic, Antihyperlipidemic and Hemostatic were cardiovascular agents. These three bioactivity categories were all cardiovascular agents, but were not located close to each other in the cluster. In particular, the biological activity category of Antidiabetic activity was located far from the other two.

Chapter 5 Conclusion

The secondary metabolites synthesized by plants, fungi and microorganisms are diverse, e.g. there are at least 30,000 terpenoids[11], 9,000 flavonoids, 1,600 isoflavonoids and 12,000 alkaloids[12]. These metabolites have been deposited in the species-metabolite relational database(DB) known as the KNApSAcK Core DB, which contains 111,199 species-metabolite relationships involving altogether 50899 metabolites and 22,350 species [13]. Out of 50,899 metabolites of the KNApSAcK database the activity of only 3,161 metabolites are recorded. Metabolome studies have been extended from model species such as crops and medicinal plants to health care and medicinal systems based on the interactions between organisms and humans. Secondary metabolites are important components in the drug discovery process, as such biodiversity produces a variety of small molecules and provides a source of new chemicals that are potential drug leads. As a whole secondary metabolites play important roles in agriculture, ecology and healthcare

In this thesis we proposed a network based computational method for predicting activities of metabolites. First we calculate structural similarity of metabolites and we constructed metabolite network based on high structural similarity score. Next, we constructed a tailored network consisting of activity known and activity unknown metabolites and then determined high density clusters in such network using DPClusO algorithm. Next, we selected statistically significant clusters overrepresented with known similar activity metabolites based on FDR Analysis and used those clusters for predicting functions of unknown metabolites. As a byproduct of our work we also explored relations between activity categories. Because of using strict FDR thresholds, we finally succeeded to predict only 945 metabolites. Still activities of huge number of metabolites are unknown.

Biological molecular usually proteins and metabolites have many functions/activities and new functions are coming out time to time. Therefore it in difficult for a method to predict all function correctly.

Our future work will be to develop a new method for predicting activities of more metabolites. For example, in our method, we can create larger clusters by changing the

parameter of DPClusO before clustering. Therefore, more prediction results may be obtained. In addition, the same result may be obtained by changing the FDR threshold. But in this case, the number of erroneous prediction results included in the obtained prediction results may increase. So a new method is needed to predict activities of more metabolites while at the same time the error rate remains low. As another method, a new prediction result can be obtained by focusing on a part of the molecular structure, particularly a partial structure important for expressing biological activity, and creating a network based on the similarity of the partial structures.

We utilized about FDR(0.05) thresholds in our prediction method, and very high density threshold(d=0.9) for clustering. These can limit incorrect predictions. Finally, we predicted activities as only 945 metabolites of around 47000 unknown metabolites. Also, in order to make predictions with higher accuracy, it is considered effective to use a method of semi-supervised learning such as the bootstraping method to select prediction results to be added to the data set in iterative prediction.

By our method the possibility of incorrect prediction is low became our prediction is based on structural similarity. Mismatched predictions may indicate new activities.

By applying the method of this study and predicting the activity of many metabolites, it can be expected to contribute to metabolome research, healthcare based on the interaction between humans and organisms, and medicinal systems.

References and Notes

[1] S. Martens, and A. Mithofe, *Phytochemistry* **66** (2005) 2399–2407.

[2] O. Lapcik, *Phytochemistry* **68** (2007) 2909-2916.

[3] R.J. Bino, R.D. Hall, O. Fiehn, , J. Kopka, K. Saito, J. Draper *Trends Plant Sci.* **9** (2004), 418-425.

[4] T. Tohge, A. Fernie, *Phytochemistry*, **70** (2009) 450-456.

[5] M. Macel, N.M. van Dam, J.B. Keurentjes. *Mol. Ecol. Resour.* **10** (2010) 583–593.

[6] K. Saito, F. Matsuda, Annu. *Rev. Plant Biol.* **61** (2010) 463–489.

[7] S. Ikeda, T. Abe, Y. Nakamura, N. Kibinge, A. Hirai-Morita, A. Nakatani, A. , M. Altaf-Ul-Amin, S. Kanaya, *Plant Cell Physiol.* **54** (2013) 711–727.

[8] R.A.Dixon, *Nature* **411** (2001) 843–847.

[9] B.A. Halkier, J. Gershenzon, *Annu. Rev. Plant Biol.* **57** (2006) 303–333.

[10] P. Bednarek, A. Osbourn. *Science* **324** (2009) 746–748.

[11] D.B. Kell, R. Goodacre, *Drug Discov. Today* **19** (2014) 171-182.

[12] J.D. Connolly, R.A. Hill *Dictionary of Terpenoids* Chapman and Hall, London (1991).

[13] J. Ziegler,P.J. Facchini, *Annu. Rev. Plant Biol.* **59** (2008) 735–769.

[14] Y. Ohtana, T. Shirai, S. Kanaya, Md. Altaf-Ul-Amin, T. Sugiura, N. Ono, T. Sato, Y. Nakamura, T. Katsuragi, *Mol. Inf.* **33** (2014) 790-801.

[15] A. Abdullah, Md. Altaf-Ul-Amin, N. Ono, T. Sato, T. Sugiura, A. H. Morita, T.Katsuragi, A. Muto, T. Nishioka, S. Kanaya, *BioMed research international* 2015 (2015) 139254.1-13.

[16] Y. Shinbo, Y. Nakamura, M. Altaf-Ul-Amin, H. Asahi, K. Kurokawa, M. Arita, K. Saito, D. Ohta, D. Shibata, S. Kanaya *Plant metabolomics.* Springer, Berlin, Heidelberg, (2006) 165-181.

[17] M. Saito, N. Takemura, T. Shirai, J. *Mol. Biol.* 2012, 424, 379–390

[18] Y. Nakamura, F. Mochamad Afendi, A. Kawsar Parvin, N. Ono, K. Tanaka, A. Hirai Morita, T. Sato, T. Sugiura, Md. Altaf-Ul-Amin , S. Kanaya, *Plant Cell Physiol.* 55(1): e7(1-9) (2014)

[19] Md. Altaf-Ul-Amin, Y. Shinbo, K. Mihara, K. Kurokawa, S. Kanaya *BMC bioinformatics* **7** (2006): 207.

[20] Md. Altaf-Ul-Amin, M. Wada, S. Kanaya. *ISRN Biomathematics* 2012 (2012). 726429.1-11.

[21] M.B. Karim, N. Wakamatsu, and Md Altaf-Ul-Amin. *Journal of Computer Aided Chemistry* **18** (2017): 76-93.

[22] Md. Altaf-Ul-Amin, H. Tsuji, K. Kurokawa, H. Asahi, Y. Shinbo, S. Kanaya, *J.Comp. Aided Chem.* **7** (2006) 150-156.

[23] Md. Altaf-Ul-Amin, T. Katsuragi, T. Sato, N. Ono, and S. Kanaya *BioMed research international* 2014 (2014) 154594.1-8.

[24] Eguchi, R., Karim, M.B., Hu, P., Sato, T., Ono, N., Kanaya, S., & Altaf-Ul-Amin, M.(2018). An integrative network-based approach to identify novel disease genes and pathways: a caase study in the context of inflammatory bowel disease. *BMC bioinformatics, 19*(1), 264.

[25] Y. Benjamini, Y. Hochberg, J. R. *Stat. Soc. Ser. B Methodol.* **57** (1995) 289-300.

[26] Storey, J. D., & Tibshirani, R.. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA*, **100** (2003) 9440-9445.

[27] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, *Genome Research* **13** (2003) 2498-504.

[28] M.E.J. Newman, *Contemporary physics* **46** (2005): 323-3

Acknowledgements

I am grateful to Professor Shigehiko Kanaya for his great help and guiding me to accurate direction.

I am grateful to Associate Professor Md. Altaf-Ul-Amin for his enthusiastic guidance, valuable discussions and technical helps.

I would like to thank to Associate Professor Naoaki Ono, Assistant Professor Ming Huang, Tetsuo Sato and Professor Takaaki Nishioka for all their valuable comments, suggestions and knowledge sharing.

I am grateful to Nagahama Institute of Bio-Science and Technology Professor Tsuyoshi Shirai for letting me use the COMPLIG algorithm and computers for huge calculation.

I thank Postdoctoral Researcher Ryohei Eguchi for knowledge sharing and helping data extraction from huge data. I thank Tetsuo Katsuragi for giving lots of advice for life of research in NAIST. I am grateful to Research fellow Aki Morita for providing metabolite activity data from KNApSAcK DB and teaching how to read data file.

I am grateful to Secretary Minako Ohashi for many official helps.

I thank my fellow lab mates in Computational Systems Biology Laboratory for making the laboratory such a good place. I am grateful to Kennosuke Wada and Yoshiko Wada for everything they do. I thank Yuki Ohtana for knowledge sharing about this research and teaching about life in NAIST.

I heartily thank my parents for encouraging me to higher studies and supporting my life and education in many ways.

Finally, I really appreciate for everyone and everything for supporting me.

Achievements

Peer review journal paper

1. Nobutaka Wakamatsu, Ming Huang, Naoaki Ono, Md. Altaf-Ul-Amin and Shigehiko Kanaya, Prediction of metabolite activities by repetitive clustering of the structural similarity based networks. Journal of Computer Aided Chemistry, 20, 76-83. (Chapter 1, 2, 3, 4)
2. Karim, M. B., Wakamatsu, N., & Altaf-Ul-Amin, M. (2017). [Dedicated to Prof. T. Okada and Prof. T. Nishioka: data science in chemistry] DPClusOST: A Software Tool for General Purpose Graph Clustering. Journal of Computer Aided Chemistry, 18, 76-93.(Chapter 2, 3)

Peer review international conference

1. Md. Altaf-Ul-Amin, Nobutaka Wakamatsu and Shigehiko Kanaya. "An Approach to Function Prediction of Metabolites by Clustering the 3D-Chemical Structural Similarity Based Network", Translational Bioinformatics Conference 2016 (TBC 2016)', Oct. 15-17, 2016 in Jeju Island, Korea.(Chapter 3)
2. Nobutaka Wakamatsu, Tsuyoshi Shirai, Ryohei Eguchi, Mohammad Bozlul Karim, Naoaki Ono, Shigehiko Kanaya and Md. Altaf-Ul-Amin, “A Network based Approach to Predict Functions of Metabolites”, International Conference on Metabolomics, 2018, Bangkok. (Chapter 2,3)

Appendices

A. Prediction results

CID	Metabolite name	Predicted activity
C00000345	omega-Cycloheptylundecanoic acid	Pollinator attractant
C00000346	Dihydromalvalic acid	Pollinator attractant
C00000347	cis-11-Octadecanoic acid; cis-vaccenic acid	Anti-inflammatory
C00000403	13-Hydroxyoctadecadienoic acid; 13-HOD; 13-Hydroxy-9Z,11E-octadecadienoic acid	Antiarthritic
C00000408	Eicosapentaenoic acid; EPA	Anti-HSV, Antiarrhythmic, Antihyperlipidemic
C00000429	11R-Hydroxy-9Z,12Z-octadecadienoic acid	CNS stimulant
C00000432	11-Keto-9Z,12Z-octadecadienoic acid	Narcotic
C00000436	6Z,8E,10E,12Z-Octadecatetraenoic acid	Pollinator attractant
C00000556	Parasperone A; Hydroxymethylnaphthopyrone	Inhibit leaf growth, Inhibit stem growth
C00000569	Emodin anthrone; Frangula emodin anthrone	Feeding attractant
C00000621	Chavicol; p-Hydroxyallylbenzene	Feeding deterrent, Pollinator attractant
C00000625	(-)-Lirioresinol A; (-)-Episyringaresinol	Dermatitic
C00000627	(-)-Lirioresinol B; (-)-Syringaresinol	Dermatitic
C00000698	(+)-Epieudesmin	Pollinator attractant
C00000711	Retrojusticidin B	Anticholinergic, Hepatotoxic
C00000923	Demethylspheroidene	Narcotic
C00000925	beta-Echinenone; Myxoxanthine	Antileprotic, Tonic
C00000926	gamma-Carotene	Anti-HSV, Antitumor, Antiviral
C00000927	Hydroxyneurosporene; Chloroxanthin	Hemostatic
C00000964	Garbanzol	Antiprotozoal
C00000975	Kazinol A	Pollinator attractant
C00001043	Genkwanin	Nutrient
C00001112	beta-D-Apiose	Antiasthmatic
C00001113	beta-L-Arabinopyranose	Antidiabetic, Cytotoxic, Enhance plant growth, Enhance stem growth, Pollinator attractant
C00001130	beta-L-Ribopyranose	Anticancer
C00001171	Ribitol	Antibacterial, Antioxidant
C00001227	Malvalic acid	Phytoalexin
C00001235	Petroselinic acid	Pneumotoxic
C00001237	Ricinoleic acid	Inhibit CYP
C00001260	Nonacosane	Enhance plant growth
C00001262	Ginnone; Nonacosan-10-one	Antiulcerogenic

C00001278	Crepenynic acid	Antimalarial
C00002046	(+)-Hygrine; Hygrine; D-(+)-Hygrine	Anti-HSV, Antiarrhythmic, Antihyperlipidemic
C00002047	(-)-Hygroline	Antigout, Inhibit leaf growth, Inhibit stem growth
C00002157	Evoxanthidine	Anthelmintic, Antileprotic, Tonic
C00002167	Glycophymoline	Phytoalexin
C00002183	Maculine	Antimalarial
C00002229	Nuttalline; 4-beta-Hydroxylupanine; 7-Hydroxylupanine	Pneumotoxic
C00002234	(-)-Rhombifoline; Rhombifoline	Enhance stem growth
C00002286	Cochlearine	Immunosuppressant, Phytoalexin
C00002862	Rubiadin; 1,3-Dihydroxy-2-methyl-9,10-anthraquinone	Oviposition attractant
C00003037	(R)-Citronellal; (R)-3,7-Dimethyl-6-octenal	Anticholinergic
C00003205	Achillin	Phototoxic
C00003213	(+)-Arbusculin A; Arbusculin A; [3aS-(3aalpha,5abeta,9alpha,9aalpha,9bbeta)]-Decahydro-9-hydroxy-5a,9-dimethyl-3-methylenenaphtho[1,2-b]furan-2(3H)-one	Pollinator attractant
C00003218	Aromaticin	Anesthetic, Anticholinergic
C00003224	Baileyin	Antiamebic, Emetic, Expectorant
C00003237	Confertin	Pollinator attractant
C00003239	Coronopilin	Analgesic, Anti-inflammatory, Antiallergic
C00003244	Damsin	Antiallergic
C00003248	Desacetoxymatricarin; Leucodin	Anticancer
C00003286	Florilenalin	Enhance plant growth, Enhance stem growth
C00003296	Granilin	Flavor
C00003297	Grosshemin	Allelopathic, Antigout, Antihyperlipidemic
C00003298	Helenalin	Antidiabetic
C00003303	Hymenolin	Antidiabetic, Pollinator attractant
C00003310	Ivalin; [3aR-(3aalpha,4aalpha,7alpha,8abeta,9aalpha)]-Decahydro-7-hydroxy-8a-methyl-3,5-bis(methylene)-naphtho[2,3-b]furan-2(3H)-one	Pollinator attractant
C00003344	Parthenin	Anticancer
C00003351	Pleniradin	Analgesic, Antiallergic, Pollinator attractant
C00003352	Plenolin; Dihydrohelenalin; 11,13-Dihydrohelenalin	Antihepatotoxic, Phytoalexin
C00003393	Tauremizin; Vulgarin; Judaicin (sesquiterpene); Barrelin	Antiallergic
C00003593	Tigogenin	Anti-HSV, Antiarrhythmic, Antihyperlipidemic

C00003658	Lichesterol; (24S)24-Methylcholesta-5,8,22-trien-3beta-ol; Ergosta-5,8,22-trien-3beta-ol	Pneumotoxic
C00003659	Lophenol	Antigout
C00003766	delta-Carotene	Antiallergic, Antibacterial, Antidiabetic, Antioxidant
C00003785	Rubixanthin	Antileishmanial
C00003786	Torulene; 3',4'-Didehydro-beta,psi-carotene	Pollinator attractant
C00003788	Primuletin	Antiallergic, Antibacterial, Antidiabetic, Antioxidant
C00003790	7-Methoxyflavone	Anti-HSV, Antiarrhythmic, Antihyperlipidemic
C00003791	2'-Hydroxyflavone	Hepatotoxic
C00003792	5-Hydroxy-6-methoxyflavone	Feeding attractant
C00003796	5,2'-Dihydroxyflavone	Anticancer, Pollinator attractant
C00003802	8,2'-Dihydroxyflavone	Anti-HSV, Antiarrhythmic, Antihyperlipidemic
C00003803	2',5'-Dihydroxyflavone	Pollinator attractant
C00003815	5,7,2'-Trihydroxyflavone	Antigout, Piscicidal
C00003816	Echioidinin	Inhibit root growth
C00003823	5,2',5'-Trihydroxyflavone	Pneumotoxic, Toxic
C00003940	Agehoustin G	Insecticidal
C00003951	5,6,8,3',4',5'-Hexamethoxyflavone	Cytotoxic
C00003962	Bannamurpanisin	Cytotoxic
C00003963	5,6,2',3',4',6'-Hexamethoxyflavone; 5,6-Dimethoxy-2-(2,3,4,6-tetramethoxyphenyl)-4H-1-benzopyran-4-one	Phytoalexin
C00003967	Agehoustin F	Pollinator attractant
C00003970	Gardenin E; 3',5,5'-Trihydroxy-4',6,7,8-tetramethoxyflavone; 2-(3,5-Dihydroxy-4-methoxyphenyl)-5-hydroxy-6,7,8-trimethoxy-4H-1-benzopyran-4-one	Antioxidant
C00003971	Scaposin	Hepatotoxic
C00003972	Gardenin C; 3',5-Dihydroxy-4',5',6,7,8-pentamethoxyflavon; 5-Hydroxy-2-(3-hydroxy-4,5-dimethoxyphenyl)-6,7,8-trimethoxy-4H-1-benzopyran-4-one	Flavor, Hepatotoxic, Piscicidal
C00003974	3'-Hydroxy-5,6,7,8,4',5'-hexamethoxyflavone	Enhance root growth
C00003976	Gardenin A; 5-Hydroxy-3',4',5',6,7,8-hexamethoxyflavone; 5-Hydroxy-6,7,8-trimethoxy-2-(3,4,5-trimethoxyphenyl)-4H-1-benzopyran-4-one	Inhibit leaf growth, Inhibit stem growth
C00003978	Psiadiarabicin	Anticancer

C00004005	Linderoflavone B; Lucidin dimethyl ether; 5,6,7,8-Tetramethoxy-3',4'-(methylenedioxy)flavone; 2-(1,3-Benzodioxol-5-yl)-5,6,7,8-tetramethoxy-4H-1-benzopyran-4-one	Enhance stem growth
C00004006	5,6,7,3'-Tetramethoxy-4',5'-methylenedioxyflavone	Antidiabetic, Pollinator attractant
C00004007	5,3',4',5'-Tetramethoxy-6,7-methylenedioxyflavone	Anti-inflammatory, Pollinator attractant
C00004009	7-Hydroxy-5,6,8,3'-tetramethoxy-4',5'-methylenedioxyflavone	Antiprotozoal
C00004017	Gancaonin Q	Oviposition attractant
C00004082	5,6,7,3',4',5'-Hexamethoxyflavone; 5,6,7-Trimethoxy-2-(3,4,5-trimethoxyphenyl)-4H-1-benzopyran-4-one	Anti-inflammatory
C00004088	Licoflavone B; Prenyllicoflavone A; 7-Hydroxy-2-[4-hydroxy-3-(3-methyl-2-butenyl)phenyl]-6-(3-methyl-2-butenyl)-4H-1-benzopyran-4-one	Analgesic, Anti-inflammatory, Antiallergic
C00004093	Laxifolin; 5-Hydroxy-2-(4-hydroxyphenyl)-8,8-dimethyl-6-(3-methyl-2-butenyl)-4H,8H-benzo[1,2-b:3,4-b']dipyran-4-one	Phototoxic
C00004110	Aequinetin	Nutrient
C00004111	Chrysin 7-O-beta-galactopyranoside; Chrysin 7-galactoside	Teratogenic
C00004113	Chrysin 7-rutinoside	Nutrient
C00004120	Baicalein 6-glucoside	Inhibit root growth
C00004135	7,3',4'-Trihydroxyflavone 7-rutinoside	Sex attractant
C00004140	Apigenin 5-galactoside	Hallucinogenic
C00004142	Apigenin 7-rhamnoside	Nutrient
C00004154	Apigenin 7-xylosyl-(1->6)-glucoside	Muscle relaxant
C00004157	Rhoifolin; (-)-Rhoifolin	Antiamebic, Emetic, Expectorant
C00004167	Apigenin 7-neohesperidoside-4'-glucoside	Muscle relaxant
C00004205	Linarin; Acacetin 7-O-rutinoside; Acacetin 7-rutinoside; Buddleoflavonoloside	Hallucinogenic
C00004217	Scutellarein 6-xyloside	Sex attractant
C00004227	Scutellarein 6-xyloside-7-rhamnoside	Hallucinogenic
C00004417	8-Hydroxyluteolin 7-xyloside	Sex attractant
C00004474	Isoetin 7-glucoside-2'-xyloside	Inhibit plant growth
C00004531	3,7-Dihydroxyflavone	Antimalarial
C00004534	Galangin 3-methyl ether; 5,7-Dihydroxy-3-methoxyflavone; 3-Methylgalangin; 5,7-Dihydroxy-3-methoxy-2-phenyl-4H-1-benzopyran-4-one	Inhibit plant growth
C00004539	Zuccagin	Sex attractant
C00004582	Fisetin 4'-methyl ether	Nutrient
C00004665	Emmaosunin	Phytoalexin

C00004673	5-Hydroxyauranetin; 5-Hydroxy-3,6,7,8,4'-pentamethoxyflavone; 5-Hydroxy-3,6,7,8-tetramethoxy-2-(4-methoxyphenyl)-4H-1-Benzopyran-4-one	Enhance plant growth, Enhance stem growth
C00004674	3,5,6,7,8,4'-Hexamethoxyflavone; 3,5,6,7,8-Pentamethoxy-2-(4-methoxyphenyl)-4H-1-benzopyran-4-one	Antiasthmatic, Anticancer
C00004676	Grantiodin	Anticholinergic, Hepatotoxic, Pneumotoxic
C00004711	Quercetagetin 3,5,6,3',4'-pentamethyl ether; 7-Hydroxy-3,5,6,3',4'-pentamethoxyflavone; 2-(3,4-Dimethoxyphenyl)-7-hydroxy-3,5,6-trimethoxy-4H-1-benzopyran-4-one	Insecticidal
C00004713	Quercetagetin hexamethyl ether; 3,5,6,7,3',4'-Hexamethoxyflavone; Hexamethylquercetagetin; Oxyayanin B trimethyl ether; 2-(3,4-Dimethoxyphenyl)-3,5,6,7-tetramethoxy-4H-1-benzopyran-4-one	Phytoalexin
C00004718	5-Hydroxy-3,7,8,2',4'-pentamethoxyflavone	Anticancer
C00004742	Gossypetin 3,5,7,8,3'-pentamethyl ether	Antiallergic
C00004743	Gossypetin 3,5,8,3',4'-pentamethyl ether	Analgesic, Toxic
C00004744	Gossypetin 3,7,8,3',4'-pentamethyl ether; 5-Hydroxy-3,7,8,3',4'-pentamethoxyflavone; 2-(3,4-Dimethoxyphenyl)-5-hydroxy-3,7,8-trimethoxy-4H-1-benzopyran-4-one	Anesthetic, Pneumotoxic
C00004745	Gossypetin hexamethyl ether; 3,5,7,8,3',4'-Hexamethoxyflavone; 2-(3,4-Dimethoxyphenyl)-3,5,7,8-tetramethoxy-4H-1-benzopyran-4-one	Anticancer, Antitumor
C00004755	5'-hydroxy-3,5,7,2',4'-pentamethoxyflavone	Anticholinergic, Hepatotoxic, Pneumotoxic
C00004756	2'-Hydroxy-3,5,7,4',5'-pentamethoxyflavone	Anticholinergic, Hepatotoxic, Pneumotoxic
C00004757	5-Hydroxy-3,7,2',4',5'-pentamethoxyflavone	Anticholinergic, Hepatotoxic, Pneumotoxic
C00004758	3,5,7,2',4',5'-Hexamethoxyflavone	Antigout
C00004777	Combretol; 5-Hydroxy-3,3',4',5',7-pentamethoxyflavone; 5-Hydroxy-3,7-dimethoxy-2-(3,4,5-trimethoxyphenyl)-4H-1-benzopyran-4-one	Anticancer
C00004778	Myricetin hexamethyl ether; 3,5,7,3',4',5'-Hexamethoxyflavone; Hexa-O-methylmyricitin; Hexamethylmyricetin; 3,5,7-Trimethoxy-2-(3,4,5-trimethoxyphenyl)-4H-1-benzopyran-4-one	Hepatotoxic
C00004781	5,2',4'-Trihydroxy-3,6,7,8-tetramethoxyflavone	Pigment
C00004782	5,4'-Dihydroxy-3,6,7,8,2'-pentamethoxyflavone	Hepatotoxic
C00004783	5,2'-Dihydroxy-3,6,7,8,4'-pentamethoxyflavone	Anesthetic
C00004793	5,3',4'-Trihydroxy-3,6,7,8-tetramethoxyflavone; 5,4',5'-Trihydroxy-3,6,7,8-tetramethoxyflavone; 2-(3,4-Dihydroxyphenyl)-5-hydroxy-3,6,7,8-tetramethoxy-4H-1-benzopyran-4-one	Anticancer

C00004794	5,7,4'-Trihydroxy-3,6,8,3'-tetramethoxyflavone; 5,7-Dihydroxy-2-(4-hydroxy-3-methoxyphenyl)-3,6,8-trimethoxy-4H-1-benzopyran-4-one	Anticholinergic, Hepatotoxic, Pneumotoxic
C00004795	5,7,3'-Trihydroxy-3,6,8,4'-tetramethoxyflavone	Hepatotoxic
C00004796	5,6,3'-Trihydroxy-3,7,8,4'-tetramethoxyflavone	Pollinator attractant
C00004797	3,5,4'-Trihydroxy-6,7,8,3'-tetramethoxyflavone; 3,5-Dihydroxy-2-(3-hydroxy-4-methoxyphenyl)-6,7,8-trimethoxy-4H-1-benzopyran-4-one	Anticancer
C00004799	5,4'-Dihydroxy-3,6,7,8,3'-pentamethoxyflavone; 3'-Methoxycalycopterin; 5-Hydroxy-2-(4-hydroxy-3-methoxyphenyl)-3,6,7,8-tetramethoxy-4H-1-benzopyran-4-one	Anesthetic
C00004800	5,3'-Dihydroxy-3,6,7,8,4'-pentamethoxyflavone; 5-Hydroxy-2-(3-hydroxy-4-methoxyphenyl)-3,6,7,8-tetramethoxy-4H-1-benzopyran-4-one	Pollinator attractant
C00004801	3,5-Dihydroxy-6,7,8,3',4'-pentamethoxyflavone; 2-(3,4-Dimethoxyphenyl)-3,5-dihydroxy-6,7,8-trimethoxy-4H-1-benzopyran-4-one	Anticancer, Antitumor
C00004802	5-Hydroxy-3,6,7,8,3'4'-hexamethoxyflavone; 2-(3,4-Dimethoxyphenyl)-5-hydroxy-3,6,7,8-tetramethoxy-4H-1-benzopyran-4-one	Anticholinergic
C00004812	5,7,2'-Trihydroxy-3,6,4',5'-tetramethoxyflavone	Immunomodulative
C00004813	5,6,5'-Trihydroxy-3,7,2',4'-tetramethoxyflavone	Allelopathic
C00004814	Apulein	Anticholinergic, Hepatotoxic, Pneumotoxic
C00004815	6,5'-Dihydroxy-3,5,7,2',4'-pentamethoxyflavone	Anticholinergic, Hepatotoxic, Pneumotoxic
C00004816	6,2'-Dihydroxy-3,5,7,4',5'-pentamethoxyflavone	Anticancer
C00004817	5,5'-Dihidroxy-3,6,7,2',4'-pentamethoxyflavone	Allelopathic
C00004818	Brickellin	Antidiabetic
C00004819	Distemonatin	Anticholinergic, Hepatotoxic, Pneumotoxic
C00004820	Benthamitin	Phototoxic
C00004828	5,7,4'-Trihydroxy-3,6,3',5'-tetramethoxyflavone	Antiparkinson
C00004829	5,7,3'-Trihydroxy-3,6,4',5'-tetramethoxyflavone	Hepatotoxic
C00004830	Apuleitrin	Allelopathic, Antihyperlipidemic
C00004831	3,5,3'-Trihydroxy-6,7,4',5'-tetramethoxyflavone	Allelopathic
C00004832	Apuleirin	Enhance root growth
C00004833	Murrayanol	Allelopathic
C00004834	5,3'-Dihydroxy-3,6,7,4',5'-pentamethoxyflavone	Anticholinergic, Hepatotoxic, Pneumotoxic, Pollinator attractant
C00004835	4'-Hydroxy-3,5,6,7,3',5'-hexamethoxyflavone	Molluscicidal
C00004838	5,8-Dihydroxy-3,7,2',3',4'-pentamethoxyflavone	Antioxidant

C00004844	Hibiscetin 3,7,8,4'-tetramethyl ether	Antigout
C00004845	Hibiscetin 3,8,3',4'-tetramethyl ether	Hepatotoxic
C00004846	Conyzatin	Antiasthmatic
C00004849	5,7,4',5'-Tetrahydroxy-3,6,8,2'-tetramethoxyflavone	Pigment
C00004850	5,7,2',5'-Tetrahydroxy-3,6,8,4'-tetramethoxyflavone	Teratogenic
C00004851	5,7,2',4'-Tetrahydroxy-3,6,8,5'-tetramethoxyflavone	Allelopathic
C00004859	5,7,3',5'-Tetrahydroxy-3,6,8,4'-tetramethoxyflavone	Flavor
C00004860	5,7,3',4'-Tetrahydroxy-3,6,8,5'-tetramethoxyflavone	Allelopathic, Antihyperlipidemic
C00004861	5,6,3',5'-Tetrahydroxy-3,7,8,4'-tetramethoxyflavone	Anticancer
C00004862	5,3',5'-Trihydroxy-3,6,7,8,4'-pentamethoxyflavone	Molluscicidal
C00004863	5,7,4'-Trihydroxy-3,6,8,3',5'-pentamethoxyflavone	Molluscicidal
C00004937	Quercetagetin 5,6,7,3',4'-pentamethyl ether; Marionol; 3',4',5,6,7-Pentamethylquercetagetin; 2-(3,4-Dimethoxyphenyl)-3-hydroxy-5,6,7-trimethoxy-4H-1-benzopyran-4-one	Enhance root growth
C00005029	Broussoflavonol F; 3,5,7-Trihydroxy-2-[4-hydroxy-3-(3-methyl-2-butenyl)phenyl]-8-(3-methyl-2-butenyl)-4H-1-benzopyran-4-one	Sex attractant
C00005061	Meliternin	Antiallergic, Antibacterial, Antidiabetic, Antioxidant
C00005065	5-Hydroxy-3,6,7,8-tetramethoxy-3',4'-methylenedioxyflavone; 5-Demethylmelibentin; 2-(1,3-Benzodioxol-5-yl)-5-hydroxy-3,6,7,8-tetramethoxy-4H-1-benzopyran-4-one	Antihypertensive
C00005066	Melibentin; 3,5,6,7,8-Pentamethoxy-3',4'-methylenedioxyflavone; 2-(1,3-Benzodioxol-5-yl)-3,5,6,7,8-pentamethoxy-4H-1-benzopyran-4-one	Muscle relaxant
C00005067	3,5,8-Trimethoxy-6,7:3',4'-bis(methylenedioxy)flavone; 6-(1,3-Benzodioxol-5-yl)-4,7,9-trimethoxy-8H-1,3-dioxolo[4,5-g][1]benzopyran-8-one	Narcotic
C00005122	Galangin 7-glucoside	Nutrient
C00005133	Kaempferol 3-alpha-L-arabinopyranoside	Antiprotozoal
C00005134	Kaempferol 3-xyloside	Sex attractant
C00005156	Kaempferol 3-lathyroside	Hallucinogenic
C00005162	Kaempferol 3-vicianoside	Muscle relaxant
C00005163	Leucoside; Kaempferol 3-sambubioside; 5,7-Dihydroxy-2-(4-hydroxyphenyl)-3-[(2-O-beta-D-xylopyranosyl-beta-D-glucopyranosyl)oxy]-4H-1-benzopyran-4-one	Antileprotic, Tonic
C00005175	Kaempferol 3-xyloside-7-glucoside	Anti-HSV, Hemostatic
C00005176	Lepidoside	Nutrient
C00005179	Kaempferol 3-glucoside-7-alpha-L-arabinopyranoside	Antiamebic, Emetic, Expectorant
C00005188	Kaempferol 3-rhamnoside-7-glucoside	Neurotoxic

C00005196	Kaempferol 7-glucosyl-(1->4)-xyloside	Pneumotoxic
C00005207	Primflasine	Molluscicidal
C00005295	Kaempferide 3,7-dirhamnoside	Teratogenic
C00005368	Guaijaverin; Quercetin 3-O-alpha-L-arabinopyranoside	Hallucinogenic
C00005369	Quercetin 3-beta-L-arabinopyranoside	Anti-HIV
C00005389	Quercetin 3-xylosyl-(1->2)-alpha-L-arabinofuranoside	Antipsychotic
C00005390	Quercetin 3-rhamnosyl-(1->2)-alpha-L-arabinopyranoside	Pneumotoxic
C00005415	Quercetin 3-xylosyl-(1->2)-rhamnoside	Anticholinergic, Pneumotoxic
C00005422	Quercetin 3-xyloside-7-rhamnoside	Inhibit root growth
C00005871	Kaempferol 7-(6''-p-succinylglucoside)	Teratogenic
C00005972	Quercetin 3-(6'''-sinapylglucosyl)(1->2)-galactoside	Antitumor
C00006083	Bayin; 8-beta-D-Glucopyranosyl-4',7-dihydroxyflavone	Anti-HSV, Antiarrhythmic, Antihyperlipidemic
C00006085	3'-Deoxyderhamnosylmaysin	Anti-HSV, Antiarrhythmic, Antihyperlipidemic
C00006097	8-C-beta-D-Galactopyranosylapigenin	Other digestive organ agent
C00006167	Orientin 2''-O-beta-L-arabinofuranoside	Hallucinogenic
C00006169	Isoorientin 6''-O-alpha-L-arabinoside	Molluscicidal
C00006192	6-C-Xylosylluteolin 2''-O-rhamnoside	Antileprotic, Tonic
C00006197	6-C-Xylopyranosyl-8-C-galactopyranosylapigenin	Hallucinogenic
C00006209	Isoorientin 6''-O-rhamnoside	Anthelmintic, Antileprotic, Tonic
C00006226	Isoswertisin 2''-O-xyloside	Anti-HSV, Antiarrhythmic, Antihyperlipidemic
C00006380	Vicenin 3	Hallucinogenic
C00006381	Isoschaftoside; Apigenin 6-C-alpha-L-arabinopyranoside-8-C-beta-D-glucopyranoside; Isoshaftoside	Hallucinogenic
C00006396	6,8-Di-C-rhamnosylapigenin	Molluscicidal
C00006405	Polygonatiin; Vitexin 2''-O-sophoroside	Laxative
C00006543	Ochnaflavone 4'-methyl ether	Antiamebic, Emetic, Expectorant
C00006549	Lophirone H	Antihypertensive
C00006611	Luteolinidin	Anti-HSV, Antiarrhythmic, Antihyperlipidemic
C00006633	Pelargonidin 3-lathyroside	Teratogenic
C00006634	Pelargonidin 3-robinobioside	Phototoxic
C00006639	Pelargonidin 3-rutinoside	Feeding attractant
C00006641	Pelargonidin 3-rhamnoside-5-glucoside	Antileprotic, Tonic
C00006653	Cyanidin 3-rhamnoside	Teratogenic
C00006655	Cyanidin 3-lathyroside	Hallucinogenic

C00006657	Sambicyanin; Cyanidin 3-xyloglucoside; Cyanidine 3-sambubioside	Mutagenic, Pneumotoxic
C00006824	Malonylshisonin	Antitussive
C00006922	Echinatin	Sex attractant
C00006936	Flavokawin B	Antileprotic, Tonic
C00007072	Kanzonol C; 2',4',4-Trihydroxy-3'-prenylchalcone	Allelopathic
C00007096	3,2',4'-Trihydroxy-6'',6''-dimethyl-3'-prenylpyrano[2'',3'':4,5]chalcone	Attractant
C00007138	Ammothamnidin	Emetic
C00007183	4,2'-Dihydroxychalcone 4-glucoside	Emetic
C00007261	5-Dehydroepisterol; 24-Methylcholesta-5,7,24(28)-trienol; Ergosta-5,7,24(28)-trien-3beta-ol	Antiulcerogenic
C00007273	6-Oxocampestanol	Teratogenic
C00007290	Xylose; D-(+)-Xylose	Antifungal
C00007321	24-Ethylidenelophenol	Nutrient
C00007327	Phenylacetaldoxime	Antiallergic, Antibacterial, Antidiabetic, Immunosuppressant, Phytoalexin
C00007382	Anthranilic acid	Antituberculotic, Phytoalexin
C00007444	Adenosine	Antifungal
C00007517	6alpha-Hydroxycampestanol	Inhibit leaf growth, Inhibit stem growth
C00007546	Methylbenzoate	Pollinator attractant
C00007685	13-Hydroxysparteine	Allelopathic
C00007694	alpha-Isolupanine	Enhance flowering, Enhance fruiting, Enhance germination, Enhance leaf growth
C00007696	5,6-Dehydro-alpha-isolupanine	Antitumor, Antiviral
C00007697	17-Oxolupanine	Antiprotozoal
C00007699	3beta-Hydroxylupanine	Inhibit root growth
C00007700	5alpha-Hydroxy-7-17-dehydroisolupanine	Pneumotoxic
C00007703	12alpha-Hydroxylupanine	Antifungal
C00007717	4beta,13alpha-Dihydroxyupanine	Hallucinogenic
C00007722	(-)-Epibaptifoline; Epibaptifoline	Feeding attractant
C00007734	11-Oxotetrahydrorhombifoline	Enhance plant growth, Enhance stem growth
C00007741	N-Formylangustifoline	Phototoxic
C00007776	Mamanine	Antiprotozoal
C00007926	Davidigenin	Anticancer, Cytotoxic
C00007928	2',4',6'-Trihydroxydihydrochalcone; 3-Phenyl-1-(2,4,6-trihydroxyphenyl)-1-propanone	Pollinator attractant
C00007934	4,2',4'-Trihydroxy-3-methoxydihydrochalcone	Hallucinogenic

C00007963	Anguvetin	Cytotoxic
C00007993	Sieboldin	Other digestive organ agent
C00008024	Hispidol	Anticholinergic, Hepatotoxic, Pneumotoxic
C00008025	4,6,4'-Trihydroxyaurone	Molluscicidal
C00008043	Sulfurein	Sex attractant
C00008068	2,6,3',4'-Tetrahydroxy-2-benzylcoumaranone	Inhibit root growth
C00008087	Kuwanol E	Teratogenic
C00008137	Pinocembrin 5-O-glucoside	Pneumotoxic
C00008146	Dihydrobaicalein	Enhance root growth
C00008188	5,7-Dihydroxy-6,8-di-C-prenylflavanone	Feeding attractant
C00008198	Methyl-liquiritigenin	Attractant
C00008223	Didymin; Isosakuranetin-7-rutinoside	Antiamebic, Emetic, Expectorant
C00008232	Isocarthamidin	Sex attractant
C00008266	Euchrestaflavanone A	Molluscicidal
C00008270	Kushenol A	Feeding attractant
C00008278	Isomonospermoside	Molluscicidal
C00008280	Plathymenin	Antipsychotic
C00008334	Kushenol F	Sex attractant
C00008367	Norartocarpanone; Steppogenin; 5,7,2',4'-Tetrahydroxyflavanone	Nutrient
C00008407	2beta-5,7-Trihydroxyflavanone	Sex attractant
C00008419	Strobopinin 7-glucoside	Sex attractant
C00008444	5,8-Dihydroxy-7-methoxyflavanone	Inhibit plant growth
C00008467	Lespedezaflavanone D; Euchrestaflavanone B	Molluscicidal
C00008546	3,7-dihydroxyflavanone	Cytotoxic, Pollinator attractant
C00008669	Aromadendrin 3-beta-L-arabinopyranoside	Feeding attractant
C00008696	(2S,3R)-Epitaxifolin 3-xyloside	Inhibit root growth
C00008704	Isoastibin	Neurotoxic, Teratogenic
C00008717	Taxifolin 3,5-dirhamnoside	Nutrient
C00008750	7-Hydroxy-5-methoxyflavan	Inhibit CYP
C00008752	4'-Hydroxy-7-methoxyflavan	Defense
C00008753	Apigeniflavan	Analgesic, Toxic
C00008804	Epidistenin	Anticancer
C00008834	Afzelechin 3-O-alpha-L-rhamnopyranoside; Afzelechin-3-O-alpha-L-rhamnoside	Feeding attractant
C00008836	Afzelechin 7-O-beta-D-apiofuranoside	Feeding attractant
C00008994	Epiguibourtinidol-4alpha-ol	Antiamebic, Emetic
C00008995	Epiguibourtinidol-4beta-ol	Antiamebic, Emetic
C00009357	Guibourtinidol; (2R,3S)-4',7-Dihydroxyflavan-3-ol	Pollinator attractant

C00009379	7-Hydroxy-2-methylisoflavone	Cytotoxic
C00009382	Isoformononetin	Feeding attractant
C00009383	2'-Hydroxydaidzein; 7,2',4'-Trihydroxyisoflavone	Feeding attractant
C00009384	3'-Hydroxydaidzein; 7,3',4'-Trihydroxyisoflavone	Anti-HSV, Antiarrhythmic, Antihyperlipidemic
C00009450	5-Hydroxy-7-methoxyisoflavone	Anticholinergic
C00009491	5,6,7,8-Tetramethoxy-3',4'-methylenedioxyisoflavone	Pollinator attractant
C00009585	Neobanone; 12a-Methoxyerosone	Antidiabetic
C00009587	Millettosin; 12a-Hydroxymillettone	Teratogenic
C00009610	Demethylmedicarpin; 3,9-Dihydroxypterocarpan	Antiprotozoal
C00009667	Lespedezin; 10-Geranyl-3,9-dihydroxypterocarpan	Antihypertensive
C00009709	Isovestitol; 7,4'-Dihydroxy-2'-methoxyisoflavan	Antileishmanial
C00009748	Haginin D; 7,2',4'-Trihydroxyisoflavene	Antihypertensive
C00009750	Sepiol; 7,2',3'-Trihydroxy-4'-methoxyisoflavene	Hallucinogenic
C00009799	6,2'4'-Trihydroxy-2-phenylbenzofuran; 2-(2,4-Dihydroxyphenyl)-6-hydroxybenzofuran	Allelopathic
C00009801	6-Demethylvignafuran; 6-Hydroxy-2-(4-hydroxy-2-methoxyphenyl)benzofuran	Antitumor
C00009814	5,7-Dihydroxyisoflavone	Allelopathic, Antihyperlipidemic, Hormonal
C00009845	8,4'-Dihydroxyisoflavone	Oviposition attractant
C00009935	Ulexone A; 5,7-Dihydroxy-8-prenyl-6'',6''-dimethylpyrano[2'',3'':4',3']isoflavone	Nutrient
C00009940	8-Prenylluteone; 5,7,2',4'-Tetrahydroxy-6,8-diprenylisoflavone	Sex attractant
C00009948	Euchrenone b9	Antigout
C00009974	3-O-Demethylamorphigenin	Enhance root growth
C00009977	6-Hydroxyrotenone	Enhance root growth
C00009992	Nissicarpin; 3,7-Dihydroxy-9-methoxypterocarpan	Anthelmintic, Antileprotic, Laxative, Tonic
C00010108	Genistein 4'-O-neohesperidoside; Sophorabioside	Antiamebic, Emetic, Expectorant
C00010194	Isodalbergin; 7-Hydroxy-6-methoxy-4-phenylcoumarin	Phototoxic
C00010195	Nordalbergin; 6,7-Dihydroxy-4-phenylcoumarin	Antidiabetic
C00010196	Serratin; 5,7-Dihydroxy-4-phenylcoumarin	Anti-HSV, Antiarrhythmic, Antihyperlipidemic
C00010250	5,3',4'-Trihydroxy-7-methoxy-4-phenylcoumarin 5-O-xylosyl-(1->6)-glucoside	Other digestive organ agent
C00010277	Scutellarein 7-methyl ether 6-rhamnosyl-(1->4)-xyloside	Sex attractant
C00010475	Unedide	Antihypertensive
C00010555	10-Dehydrogardenoside	Allergenic
C00010589	Lamalbid; Lamalbide; Lamiridoside	Cytotoxic
C00010644	Paederosidic acid	Antiamebic, Emetic, Expectorant

C00010667	Pulchelloside I	Antithrombotic
C00010687	Shanzhiside methyl ester	Cytotoxic
C00010778	Ligustaloside A	Expectorant
C00010860	(R)-(-)-Cryptone	Pollinator attractant
C00010873	(-)-beta-Phellandrene	Pollinator attractant
C00010912	(+)-Carvomenthene; (+)-p-Menth-1-ene	Anti-HSV, Antitumor, Antiviral
C00010913	(-)-Carvomenthene; (-)-p-Mentha-1-ene; p-1-Menthene	Pollinator attractant
C00010914	p-Menth-4(8)-ene	Pollinator attractant
C00010935	p-Menth-4(8)-en-1-ol	Anthelmintic, Antileprotic, Tonic
C00010969	m-Mentha-1,3(8)-diene; Sylveterpinolene	Antigout
C00010975	(R)-o-Mentha-1(7),5,8-triene	Feeding attractant
C00011140	Stipulin	Teratogenic
C00011207	6,7-seco-Agroclavine	Insecticidal
C00011212	Isochanoclavine-I	Antihypertensive
C00011214	N-Demethylchanoclavine-II; Norchanoclavine II	Toxic
C00011342	Peonidin 3-(4'-arabinosylglucoside)	Hepatotoxic
C00011540	Nigellic acid	Flavor, Nutrient
C00011541	5-[1-Hydroxy-6-(hydroxymethyl)-2,6-dimethyl-4-oxo-2-cyclohexen-1-yl]-3-methyl-2,4-pentadienoic acid	Antidiabetic, Pollinator attractant
C00011758	(+)-Herbolide F; Herbolide F	Cytotoxic
C00011767	Dihydroridentin; 1,3-Dihydroxy-4,10(14)-germacradien-12,6-olide	Antihepatotoxic
C00011770	Pulverulide	Antitrypanosomal
C00011863	Eupalinin D	Pneumotoxic, Toxic
C00011974	(+)-Melitensin	Allelopathic, Antihyperlipidemic
C00011975	(+)-Dihydromelitensin	Antigout
C00011995	Epoxyzinamultifloride	Immunomodulative
C00012099	Melampomagnolide A	Immunomodulative
C00012100	8alpha-Hydroxyparthenolide; Stizolin; Stizoline	Immunomodulative
C00012107	9alpha-Hydroxydihydroparthenolide	Immunomodulative
C00012109	Melampomagnolide B	Antiarthritic
C00012111	2alpha-Hydroxydihydroparthenolide	Immunomodulative
C00012112	3beta-Hydroxy-11alpha,13-dihydroparthenolide	Antiarthritic
C00012120	Anhydroverlotorin-4alpha,5beta-epoxide	Enhance stem growth
C00012138	Rotundin	Feeding attractant
C00012331	1-epi-Tatridin B; Tanachin; 1alpha-Hydroxy-1-desoxotamirin	Antiarthritic
C00012332	Tamirin	Pigment
C00012334	Argentiolide A	Pigment
C00012344	Quadrangolide	Flavor

C00012361	Simsiolide	Antiamebic, Emetic, Expectorant
C00012362	Speciformin; Spiciformin	Hemostatic
C00012363	Purpurascenolide	Pigment
C00012694	Arbusculin E; [1S-(1alpha,2beta,4abeta,8alpha,8aalpha)]-Decahydro-1,8-dihydroxy-4a,8-dimethyl-a-methylene-2-naphthaleneacetic acid	Cytotoxic
C00012865	Ivasperine; Ivasperin; [3aR-(3aalpha,4aalpha,7alpha,8beta,8abeta,9aalpha)]-Decahydro-7,8-dihydroxy-8a-methyl-3,5-bis(methylene)-naphtho[2,3-b]furan-2(3H)-one	Antitrypanosomal
C00012871	Pulchellin C; [3aR-(3aalpha,4aalpha,6beta,7alpha,8abeta,9aalpha)]-Decahydro-6,7-dihydroxy-8a-methyl-3,5-bis(methylene)-naphtho[2,3-b]furan-2(3H)-one	Allelopathic, Antihyperlipidemic
C00012875	[3aR-(3aalpha,4aalpha,5alpha,8alpha,8abeta,9aalpha)]-Decahydro-5,8-dihydroxy-5,8a-dimethyl-3-methylene-naphtho[2,3-b]furan-2(3H)-one	Anticholinergic, Immunomodulative
C00012882	[3aR-(3aalpha,6alpha,8abeta,9aalpha)]-3a,4,6,7,8,8a,9,9a-Octahydro-6-hydroxy-5-(hydroxymethyl)-8a-methyl-3-methylenenaphtho[2,3-b]furan-2(3H)-one; Dihydroxyalloalantolactone	Antitrypanosomal
C00012913	(+)-Isotelekin; Isotelekin	Inhibit root growth
C00012917	[3aR-(3aalpha,4aalpha,8abeta,9aalpha)]-Decahydro-4a-hydroperoxy-8a-methyl-3,5-bis(methylene)-naphtho[2,3-b]furan-2(3H)-one; 5alpha-Hydroperoxy-4(15),11(13)-eudesmadien-12,8beta-olide	Pigment
C00012918	5-Epitelekin; [3aR-(3aalpha,4abeta,8abeta,9aalpha)]-Decahydro-4a-hydroxy-8a-methyl-3,5-bis(methylene)-naphtho[2,3-b]furan-2(3H)-one	Inhibit root growth
C00012919	5-Deoxy-5-hydroperoxy-5-epitelekin; [3aR-(3aalpha,4abeta,8abeta,9aalpha)]-Decahydro-4a-hydroperoxy-8a-methyl-3,5-bis(methylene)-naphtho[2,3-b]furan-2(3H)-one	Enhance stem growth, Inhibit root growth
C00012928	[3aR-(3aalpha,8abeta,9aalpha)]-3a,4,6,7,8,8a,9,9a-Octahydro-5-(hydroxymethyl)-8a-methyl-3-methylenenaphtho[2,3-b]furan-2(3H)-one	Inhibit root growth
C00012931	[3R-(3alpha,3aalpha,4aalpha,5beta,6alpha,8abeta,9aalpha)]-Decahydro-6-hydroxy-3,5,8a-trimethylnaphtho[2,3-b]furan-2(3H)-one; 3-Hydroxy-12,8-eudesmanolide	Inhibit leaf growth

C00012963	Ludovicin B; [3aS-(3aalpha,5abeta,6alpha,8alpha,9aalpha,9bbeta)]-Decahydro-6,8-dihydroxy-5a-methyl-3,9-bis(methylene)-naphtho[1,2-b]furan-2(3H)-one	Enhance stem growth
C00012964	Erivanin; [3S-(3alpha,3aalpha,5abeta,6alpha,8alpha,9aalpha,9bbeta)]-Decahydro-6,8-dihydroxy-3,5a-dimethyl-9-methylenenaphtho[1,2-b]furan-2(3H)-one	Enhance flowering, Enhance fruiting, Enhance germination, Enhance leaf growth
C00012965	1-Epierivanin; (11S)-1beta,3alpha,6alpha-trihydroxyeudesm-4(14)-en-12-oic acid gamma-lactone	Antiarthritic
C00012966	(+)-Ridentin B; Ridentin B	Enhance stem growth
C00012967	Tanacetin; [3aS-(3aalpha,5abeta,6beta,9aalpha,9bbeta)]-Decahydro-6,9a-dihydroxy-5a-methyl-3,9-bis(methylene)-naphtho[1,2-b]furan-2(3H)-one	Feeding attractant
C00012968	Arsubin; [3R-(3alpha,3abeta,5aalpha,6alpha,9abeta,9balpha)]-Decahydro-6,9a-dihydroxy-3,5a-dimethyl-9-methylenenaphtho[1,2-b]furan-2(3H)-one	Anticholinesterase
C00012969	Artemine; 11S-Artemin; Artemin (sesquiterpene); [3S-(3alpha,3aalpha,5abeta,6beta,9aalpha,9bbeta)]-Decahydro-6,9a-dihydroxy-3,5a-dimethyl-9-methylenenaphtho[1,2-b]furan-2(3H)-one	Pigment
C00012970	(-)-Isogallicadiol; Isogallicadiol; [3S-(3alpha,3aalpha,5abeta,6beta,9abeta,9bbeta)]-Decahydro-6,9a-dihydroxy-3,5a-dimethyl-9-methylenenaphtho[1,2-b]furan-2(3H)-one	Mutagenic
C00012971	1-Epiartemin; [3S-(3alpha,3aalpha,5abeta,6alpha,9aalpha,9bbeta)]-Decahydro-6,9a-dihydroxy-3,5a-dimethyl-9-methylenenaphtho[1,2-b]furan-2(3H)-one	Pollinator attractant
C00012980	[3S-(3alpha,3aalpha,4alpha,5abeta,6beta,9bbeta)]-3a,4,5,5a,6,7,8,9b-Octahydro-4,6-dihydroxy-3,5a,9-trimethylnaphtho[1,2-b]furan-2(3H)-one	Pigment
C00012981	[3aR-(3aalpha,4alpha,5abeta,6alpha,9aalpha,9bbeta)]-Decahydro-4,6-dihydroxy-5a-methyl-3,9-bis(methylene)-naphtho[1,2-b]furan-2(3H)-one	Allelopathic, Antihyperlipidemic
C00012989	[3aS-(3aalpha,5beta,5abeta,6beta,9aalpha,9bbeta)]-Decahydro-5,6-dihydroxy-5a-methyl-3,9-bis(methylene)-naphtho[1,2-b]furan-2(3H)-one	Anesthetic, Narcotic

C00012996	Rothin B; [3aR-(3aalpha,4alpha,5abeta,9aalpha,9bbeta)]-Decahydro-4,9a-dihydroxy-5a-methyl-3,9-bis(methylene)naphtho[1,2-b]furan-2(3H)-one	Antihypertensive
C00012998	[3S-(3alpha,3aalpha,5abeta,6beta,9alpha,9aalpha,9bbeta)]-Decahydro-6,9-dihydroxy-3,5a,9-trimethylnaphtho[1,2-b]furan-2(3H)-one	Teratogenic
C00012999	[3S-(3alpha,3aalpha,5abeta,6beta,7alpha,9aalpha,9bbeta)]-Decahydro-6,7-dihydroxy-3,5a-dimethyl-9-methylenenaphtho[1,2-b]furan-2(3H)-one	Antitumor, Enhance stem growth
C00013000	Alhanol; Isoerivanin; [3S-(3alpha,3aalpha,5abeta,6alpha,8alpha,9bbeta)]-3a,4,5,5a,6,7,8,9b-Octahydro-6,8-dihydroxy-3,5a,9-trimethylnaphtho[1,2-b]furan-2(3H)-one	Antitumor
C00013001	[3S-(3alpha,3aalpha,5abeta,6beta,8alpha,9bbeta)]-3a,4,5,5a,6,7,8,9b-Octahydro-6,8-dihydroxy-3,5a,9-trimethylnaphtho[1,2-b]furan-2(3H)-one; 1-Epialkhanol	Immunomodulative
C00013003	1beta-Hydroxyarbusculin A; [3aS-(3aalpha,5abeta,6beta,9alpha,9aalpha,9bbeta)]-Decahydro-6,9-dihydroxy-5a,9-dimethyl-3-methylenenaphtho[1,2-b]furan-2(3H)-one	Phytoalexin
C00013004	Dendroserin; [3S-(3alpha,3aalpha,5abeta,6beta,9beta,9aalpha,9bbeta)]-Decahydro-6,9-dihydroxy-3,5a,9-trimethylnaphtho[1,2-b]furan-2(3H)-one	Antioxidant
C00013006	Artapshin; [3S-(3alpha,3aalpha,4alpha,5abeta,6beta,9aalpha,9bbeta)]-Decahydro-4,6-dihydroxy-3,5a-dimethyl-9-methylenenaphtho[1,2-b]furan-2(3H)-one	Analgesic
C00013012	[3aS-(3aalpha,5abeta,6beta,9alpha,9aalpha,9bbeta)]-Decahydro-6-hydroxy-9-(hydroxymethyl)-5a-methyl-3-methylenenaphtho[1,2-b]furan-2(3H)-one	Inhibit leaf growth
C00013016	Sonchucarpolide; [3aS-(3aalpha,5abeta,6beta,9alpha,9aalpha,9bbeta)]-Dodecahydro-6-hydroxy-5a-methyl-3-methylene-2-oxonaphtho[1,2-b]furan-9-carboxaldehyde	Anticholinesterase
C00013065	Arsantin; [3S-(3alpha,3aalpha,5abeta,6alpha,9alpha,9aalpha,9bbeta)]-Octahydro-6-hydroxy-3,5a,9-trimethylnaphtho[1,2-b]furan-2,8(3H,4H)-dione	Cytotoxic

C00013066	(11S)-Arsanin; Arsanin; Taraxacolide; [3S-(3alpha,3aalpha,5abeta,6beta,9alpha,9aalpha,9bbeta)]-Octahydro-6-hydroxy-3,5a,9-trimethylnaphtho[1,2-b]furan-2,8(3H,4H)-dione	Enhance flowering, Enhance fruiting, Enhance germination, Enhance leaf growth
C00013067	(+)-Artecalin; [3aS-(3aalpha,5abeta,6beta,9alpha,9aalpha,9bbeta)]-Octahydro-6-hydroxy-5a,9-dimethyl-3-methylenenaphtho[1,2-b]furan-2,8(3H,4H)-dione	Enhance stem growth
C00013075	Pseudosantonin; [3S-(3alpha,3aalpha,4alpha,5abeta,9balpha)]-3a,5,5a,7,8,9b-Hexahydro-4-hydroxy-3,5a,9-trimethylnaphtho[1,2-b]furan-2,6(3H,4H)-dione	Antihypertensive
C00013197	Ivangulinic acid; [3aR-[3aalpha,5(S*),7aalpha]]-2,3,3a,4,7,7a-Hexahydro-gamma,6-dimethyl-3-methylene-2-oxo-5-benzofuranbutanoic acid	Pollinator attractant
C00013211	(+)-Umbellifolide; Umbellifolide; [3aR-(3aalpha,6alpha,7aalpha)]-Tetrahydro-6-methyl-3-methylene-6-(4-oxopentyl)-2,5(3H,4H)-benzofurandione	Pigment
C00013245	Tupichinol C; (R)-3,4-Dihydro-2-(4-hydroxyphenyl)-2H-1-benzopyran-7-ol	Allelopathic
C00013251	Anadanthoside; Fisetinidol 3-beta-D-xylopyranoside	Feeding attractant
C00013291	7-Hydroxy-6-methoxyflavone; Trigraecum	Enhance root growth
C00013342	5,6,2',3',5',6'-Hexamethoxyflavone; 5,6-Dimethoxy-2-(2,3,5,6-tetramethoxyphenyl)-4H-1-benzopyran-4-one	Anticancer
C00013347	5,4'-Dihydroxy-6,7,8,3',5'-pentamethoxyflavone; 5-Hydroxy-2-(4-hydroxy-3,5-dimethoxyphenyl)-6,7,8-trimethoxy-4H-1-benzopyran-4-one	Antihypertensive
C00013373	3-Hydroxy-5,7,3',4',5'-pentamethoxyflavone; 3-Hydroxy-5,7-dimethoxy-2-(3,4,5-trimethoxyphenyl)-4H-1-benzopyran-4-one	Antigout
C00013375	2'-Hydroxy-3,7,8,4',5'-pentamethoxyflavone; 2-(2-Hydroxy-4,5-dimethoxyphenyl)-3,7,8-trimethoxy-4H-1-benzopyran-4-one	Anticholinergic, Hepatotoxic, Pneumotoxic
C00013376	3,5,8-Trihydroxy-6,7,3',4'-tetramethoxyflavone; 2-(3,4-Dimethoxyphenyl)-3,5,8-trihydroxy-6,7-dimethoxy-4H-1-benzopyran-4-one	Antigout
C00013377	5,7-Dihydroxy-3,6,8,3',4'-pentamethoxyflavone; 2-(3,4-Dimethoxyphenyl)-5,7-dihydroxy-3,6,8-trimethoxy-4H-1-benzopyran-4-one	Anesthetic
C00013378	7-Hydroxy-3,5,6,8,3',4'-hexamethoxyflavone; 2-(3,4-Dimethoxyphenyl)-7-hydroxy-3,5,6,8-tetramethoxy-4H-1-benzopyran-4-one	Molluscicidal

C00013420	Epimedokoreanin B; 2-[3,4-Dihydroxy-5-(3-methyl-2-butenyl)phenyl]-5,7-dihydroxy-8-(3-methyl-2-butenyl)-4H-1-benzopyran-4-one	Antitussive
C00013515	2-(1,3-Benzodioxol-5-yl)-7-hydroxy-3,5,6,8-tetramethoxy-4H-1-benzopyran-4-one	Inhibit leaf growth, Inhibit stem growth
C00013602	5,7,8-Trihydroxyflavone 7-galactoside	Attractant
C00013604	5,7,2'-Trihydroxy 7-glucoside	Antipsychotic
C00013625	Acacetin 7-glucosyl-(1->4)-xyloside; Apigenin 4'-methyl ether 7-glucosyl-(1->4)-xyloside; 7-[(4-O-beta-D-Glucopyranosyl-alpha-D-xylopyranosyl)oxy]-5-hydroxy-2-(4-methoxyphenyl)-4H-1-benzopyran-4-one	Antiulcerogenic
C00013626	Acacetin 7-apiosyl (1->6)-glucopyranoside; Apigenin 4'-methyl ether 7-apiosyl (1->6)-glucopyranoside; 5-Hydroxy-2-(4-methoxyphenyl)-7-[(6-O-D-apio-beta-D-furanosyl-beta-D-glucopyranosyl)oxy]-4H-1-benzopyran-4-one	Anti-HSV, Antiarrhythmic, Antihyperlipidemic
C00013641	8-Hydroxyapigenin 7-glucosyl-(1->2)-xyloside; 7-[(2-O-beta-D-Glucopyranosyl-beta-D-xylopyranosyl)oxy]-5,8-dihydroxy-2-(4-hydroxyphenyl)-4H-1-benzopyran-4-one	Anti-HSV, Antiarrhythmic, Antihyperlipidemic
C00013658	Luteolin 7-glucosyl(1->4)-alpha-L-arabinopyranoside; 2-(3,4-Dihydroxyphenyl)-7-[(4-O-beta-D-glucopyranosyl-alpha-L-arabinopyranosyl)oxy]-5-hydroxy-4H-1-benzopyran-4-one	Antiamebic, Emetic
C00013661	Luteolin 7-rhamnosyl(1->6)galactoside; Luteolin 7-robinobioside; 7-[[6-O-(6-Deoxy-alpha-L-mannopyranosyl)-beta-D-galactopyranosyl]oxy]-2-(3,4-dihydroxyphenyl)-5-hydroxy-4H-1-benzopyran-4-one	Phototoxic
C00013736	Kaempferol 3-alpha-D-arabinopyranoside; 3-(alpha-D-Arabinopyranosyloxy)-5,7-dihydroxy-2-(4-hydroxyphenyl)-4H-1-benzopyran-4-one	Inhibit leaf growth, Inhibit stem growth
C00013754	Kaempferol 3-robinobioside-7-alpha-L-arabinofuranoside	Other digestive organ agent
C00013781	Kaempferol 3-(2''-(E)-feruloylgalactosyl-(1->4)-glucoside); 5,7-Dihydroxy-3-[[4-O-[2-O-[(2E)-3-(4-hydroxy-3-methoxyphenyl)-1-oxo-2-propenyl]-beta-D-galactopyranosyl]-beta-D-glucopyranosyl]oxy]-2-(4-hydroxyphenyl)-4H-1-benzopyran-4-one	Pneumotoxic
C00014011	Chrysin 8-C-glucopyranoside; Chrysin 8-C-beta-D-glucopyranoside; 8-beta-D-Glucopyranosyl-5,7-dihydroxy-2-phenyl-4H-1-benzopyran-4-one	Hallucinogenic
C00014013	Kaplanin; 8-beta-D-Glucopyranosyl-5-hydroxy-7-methoxy-2-phenyl-4H-1-benzopyran-4-one	Inhibit plant growth

C00014029	6-C-alpha-L-Arabinopyranosyl-8-C-beta-L-arabinopyranosylapigenin	Antiulcerogenic
C00014038	Isovitexin 4'-O-rhamnopyranoside; Isovitexin 4'-rhamnoside	Hallucinogenic
C00014041	Isovitexin 6''-O-rhamnoside; Dulcinoside; 6-[6-O-(6-Deoxy-alpha-L-mannopyranosyl)-beta-D-glucopyranosyl]-5,7-dihydroxy-2-(4-hydroxyphenyl)-4H-1-benzopyran-4-one	Antileishmanial
C00014052	Precatorin III; 6-(2-O-D-Apio-beta-D-furanosyl-beta-D-glucopyranosyl)-5-hydroxy-2-(4-hydroxyphenyl)-7-methoxy-4H-1-benzopyran-4-one	Antiprotozoal
C00014057	Isocytisoside 6''-O-beta-D-apiofuranoside	Inhibit root growth
C00014064	2''-O-alpha-L-Rhamnosyl-6-C-quinovopyranosyl-luteolin	Antipsychotic
C00014141	5,6,7,3',4',5'-Hexamethoxyflavanone	Anticancer
C00014185	Prostratol F; (2S)-7,4'-Dihydroxy-8-geranylflavanone	Antiamebic, Emetic, Expectorant
C00014222	Epidorsmanin G	Cardiotonic
C00014315	(2S)-5,6,7-Trihydroxyflavanone 7-glucoside	Mutagenic, Pneumotoxic
C00014455	Paratocarpin E; 3-Prenyl-3'-(2-hydroxy-3-methylbutyl-3-enyl)-4,2',4'-trihydroxychalcone	Antiulcerogenic
C00014471	3'-Geranylchalconaringenin; 3'-Geranyl-4,2',4',6'-tetrahydroxychalcone	Pneumotoxic
C00014528	Flavumone A	Molluscicidal
C00014591	2'-Hydroxy-4'-methoxydihydrochalcone; 2'-Hydroxy-4'-methoxy-3-phenylpropiophenone	Enhance stem growth
C00014595	4-Hydroxy-2',4'-dimethoxydihydrochalcone	Molluscicidal
C00014611	(-)-Linderatin; (1S-trans)-3-Phenyl-1-[2,4,6-trihydroxy-3-[3-methyl-6-(1-methylethyl)-2-cyclohexen-1-yl]phenyl]-propanone	Antiparkinson
C00014635	2',4',3,4,alpha-Pentahydroxydihydrochalcone 3'-C-xyloside	Feeding attractant
C00015262	Coelonin; 2,7-Dihydroxy-4-methoxy-9,10-dihydrophenanthrene	Phytoalexin
C00015310	3,4-Dihydroxy-3',5'-dimethoxybibenzyl; Combretastatin B4	Antidote
C00015316	3,4'-Dihydroxy-2-stilbenecarboxylic acid; Hydrangeic acid	Pollinator attractant
C00015524	3'-Methoxy-3,4-methylenedioxybibenzyl	Antihepatotoxic
C00015568	5,3',4'-Trihydroxybibenzyl 2-O-beta-D-glucopyranoside	Antimalarial
C00015584	5,4'-Dihydroxybibenzyl 2-O-beta-D-glucopyranoside	Pollinator attractant
C00015605	5-Hydroxy-3,4-dimethoxy-9,10-dihydrophenanthrene	Hemostatic
C00015768	Flavidinin	Antiprotozoal
C00017058	(Z,E,E,E,E)-5-(2,4,6,12-Tetradecatetraene-8,10-diynylidene)-2(5H)-furanone; Xerulin	Antiparkinson, Flavor
C00017113	Pannorin	Antiallergic
C00017305	1beta,10alpha-Dihydroxy-11(13)-eremophilen-12,8beta-olide	Immunomodulative
C00018071	4',5,7,8-Tetrahydroxyisoflavone; 8-Hydroxygenistein	Inhibit root growth

C00018099	3,4-Dihydroxybenzaldehyde; Protocatechualdehyde; Catechaldehyde	Antitumor, Antiviral
C00018685	Purine riboside; 9-beta-D-Ribofuranosyl-9H-purine; Nebularine; Purine-9-beta-D-ribofuranoside; Purinosine; Ribosylpurine	Allelopathic
C00018862	7,4'-Dihydroxy-3'-methoxyisoflavan	Oviposition attractant
C00018924	Eryzerin C; (3R)-7,2',4'-Trihydroxy-6,8-diprenylisoflavan	Antihypertensive
C00018969	Sigmoidin K; 3,9-Dihydroxy-2,10-diprenylcoumestan	Sex attractant
C00018976	Tetrapterol I; 7,4'-Dihydroxy-6,3'-diprenylisoflavanone	Anticancer
C00019003	Soyasaponin V; Soyasaponin Ba	Enhance flowering, Enhance fruiting, Enhance germination, Enhance leaf growth
C00019050	(+)-12a-Epimillettosin; (6aS,12aS)-12a-Hydroxy-2,3-methylenedioxy-6',6'-dimethoxypyr[2',3':9,8]rotenone	Inhibit leaf growth, Inhibit stem growth
C00019226	3-Epitaraxerol	Anti-HIV
C00019240	Oxyresveratrol 2-O-beta-glucopyranoside	Hepatotoxic
C00019301	4'-O-Methylgenistein 8-C-glucoside; 5,7-Dihydroxy-4'-methoxyisoflavone 8-C-glucoside	Sex attractant
C00019325	Mulberrofuran V; 6,3',5'-Trihydroxy-2',6'-diprenyl-2-arylbenzofuran	Pollinator attractant
C00019375	Daidzein G 3; 7,4'-Dihydroxyisoflavone 7-O-(2-O-methylrhamnoside)	Molluscicidal
C00019467	Kanzonol X; Tenuifolin B (isoflavan); (3R)-7,2',4'-Trihydroxy-8,3'-diprenyloxyisoflavan	Hepatotoxic
C00019493	6,7,8,3',4',5'-Hexamethoxyisoflavone	Enhance plant growth
C00019506	2',4'-Dihydroxy-7-methoxyisoflavone	Nutrient
C00019634	L-Ribulose; L-erythro-Pentulose	Antihepatotoxic
C00019648	Striatin; (6aR,11aR)-3,9-Dihydroxy-2-(1,1-dimethyl-2-propenyl)-10-prenylpterocarpan	Narcotic
C00019684	D-Xylulose; D-threo-Pentulose	Defense
C00019919	(2R)-2'-Hydroxymarmesin 2'-O-beta-D-glucopyranoside	Antiparkinson
C00020097	Heritol	Allelopathic, Phytoalexin
C00020212	7,3',4'-Trihydroxyflavone 6,8-diglucoside	Teratogenic
C00020218	7-Hydroxyflavone	Antigout
C00020344	7,4'-Dimethoxyflavone	Mutagenic, Pneumotoxic
C00020469	Isopyrethroidinin	Antihypertensive
C00020473	Vestenolide	Antihypertensive
C00020474	Magnograndiolide	Anti-HSV, Antiarrhythmic, Antihyperlipidemic
C00020477	Cumambrin B	Pigment

C00020498	[3aR-(3aalpha,4beta,6aalpha,7beta,9aalpha,9balpha)]-3a,4,5,6,6a,7,9a,9b-Octahydro-4,7-dihydroxy-9-methyl-3,6-bis(methylene)-azuleno[4,5-b]furan-2(3H)-one	Pigment
C00020504	8-Hydroxyzaluzanin C; Deacylaguerin A; Eleganin (guaiane)	Inhibit root growth
C00020514	Integrifolin	Inhibit root growth
C00020519	9alpha-Hydroxyzaluzalin C	Allelopathic
C00020541	Grosheiminol	Analgesic, Anti-inflammatory, Antiallergic
C00020542	9beta-Hydroxy-4beta,11beta,13,15-tetrahydrozaluzanin C	Hormonal
C00020543	9beta-Hydroxy-11beta,13-dihydrozaluzanin	Inhibit root growth
C00020545	9beta-Hydroxy-4beta,11alpha,13,15-tetrahydrozaluzanin C	Antiparkinson
C00020546	[3R-(3alpha,3abeta,5alpha,6abeta,8alpha,9abeta,9balpha)]-Decahydro-5,8-dihydroxy-3-methyl-6,9-bis(methylene)-azuleno[4,5-b]furan-2(3H)-one	Allelopathic, Analgesic, Anti-inflammatory, Antiallergic, Antihyperlipidemic
C00020547	[3R-(3alpha,3abeta,5alpha,6abeta,9beta,9abeta,9balpha)]-Octahydro-5-hydroxy-3,9-dimethyl-6-methylene-azuleno[4,5-b]furan-2,8(3H,4H)-dione	Antitumor, Enhance stem growth
C00020616	[3S-(3alpha,3aR*,4aalpha,6aalpha,9abeta,9balpha)]-5,6,6a,7,9a,9b-Hexahydro-3-hydroxy-1,4a-dimethyl-7-methylene-3H-oxireno[8,8a]azuleno[4,5-b]furan-8(4aH)-one	Phototoxic
C00020688	(+)-Mucronulatol; (R)-Mucronulatol	Anti-HIV
C00020731	1alpha-Hydroxy-4alpha-hydroperoxy-11betaH-guaia-2,10(14)-dien-6alpha,12-olide	Antihepatotoxic
C00020825	[3aS-(3aalpha,5beta,6aalpha,9aalpha,9bbeta)]-Octahydro-5-hydroxy-3,6,9-tris(methylene)-azuleno[4,5-b]furan-2,8(3H,4H)-dione	Antihypertensive
C00020839	Dihydroestafiatone	Antitrypanosomal
C00020883	[3aR-(3aalpha,5alpha,7alpha,7aalpha,8alpha,9abeta)]-3a,5,6,7,7a,8,9,9a-Octahydro-5,7-dihydroxy-5,8-dimethyl-3-methylene-azuleno[6,5-b]furan-2(3H)-one	Piscicidal
C00020885	Hymenoratin G	Antihepatotoxic
C00020886	Dihydroflorilenalin	Phototoxic
C00020888	[3R-(3alpha,3aalpha,4aalpha,6beta,7aalpha,9abeta)]-Decahydro-4a,6-dihydroxy-3-methyl-5,8-bis(methylene)-azuleno[6,5-b]furan-2(3H)-one	Inhibit root growth
C00020892	Carolenalin	Inhibit root growth
C00020921	4alpha,5alpha-Epoxyinuviscolide	Analgesic
C00020928	[3aR-(3aalpha,4aalpha,5alpha,7abeta,9abeta)]-3a,4,4a,5,6,7,7a,9a-Octahydro-5-hydroxy-5,8-dimethyl-3-methylene-azuleno[6,5-b]furan-2(3H)-one	Antihepatotoxic

C00020929	9beta,10beta-Epoxy-4alpha-hydroxy-11(13)-guaien-12,8alpha-olide	Narcotic
C00020933	2-Deoxypleniradin	Analgesic, Anti-inflammatory, Antiallergic
C00020944	Isogeigerin	Enhance plant growth, Enhance stem growth
C00020960	11,13-Dehydrocarolenalin	Immunomodulative
C00021031	Hymenoratin	Anticholinesterase
C00021032	Neohymenoratin; Geigerinine	Antioxidant
C00021057	Ambrosic acid	Antiallergic
C00021065	Pulchellin A	Antiasthmatic
C00021067	4-Epipulchellin	Antihepatotoxic
C00021068	Neopulchellin	Anti-HSV, Antiarrhythmic, Antihyperlipidemic, Pollinator attractant
C00021073	Ambrosiol	Immunomodulative
C00021074	Cumanin	Enhance stem growth
C00021077	Peruvinine	Antiallergic
C00021080	Carpesiolin; 6alpha-Hydroxy-2,3-dihydroaromaticin	Enhance flowering, Enhance fruiting, Enhance germination, Enhance leaf growth
C00021081	4beta-Hydroxy-3-oxo-1alpha,10betaH-pseudoguaia-11(13)-en-12,8beta-olide	Inhibit leaf growth
C00021093	Rudmollin	Antidiabetic
C00021106	10alpha-Chloro-1beta-hydroxyeremophil-11(13)-en-12,8beta-olide	Enhance plant growth, Enhance stem growth
C00021117	8-epi-Helenalin	Antioxidant
C00021121	Villosin A	Inhibit root growth
C00021122	Franserin	Antiallergic
C00021123	Peruvin	Antihypertensive
C00021124	Ivoxanthin	Pollinator attractant
C00021125	Bipinnatin	Antihypertensive
C00021132	Deacetylisotenulin	Analgesic, Antiallergic
C00021133	Deacetylconfertiflorin	Antihepatotoxic
C00021134	Mexicanin A	Piscicidal
C00021139	10alpha-Hydroxy-8-epi-confertin	Antidiabetic
C00021186	Hymenoratin C	Anticancer
C00021198	Psilostachyin C	Pollinator attractant
C00021366	2-Hydroxydendrobine	Inhibit CYP
C00021367	6-Hydroxydendrobine	Inhibit CYP
C00021688	9alpha-Hydroxy-11,13alpha-dihydrozaluzanin C	Enhance stem growth
C00021689	Hymenin; Epiparthenin	Pollinator attractant

C00021696	Asperilin	Pollinator attractant
C00021704	4beta,5alpha-Epoxy-4,5-cis-inunolide	Antiamebic, Emetic, Expectorant
C00021707	4-epi-Isoinuviscolide	Antihepatotoxic
C00021912	(-)-Deoxocrispolide	Allergenic
C00021918	Ferreyrantholide	Antihyperlipidemic
C00022195	Isotrixagol	Molluscicidal
C00022589	3Z-Hexenyl-cis-ferulate; 3Z-Hexenylferulate	Anticancer, Pollinator attractant
C00022620	(E,Z)-Matricarianol; (E,Z)-2,8-Decadiene-4,6-diyn-1-ol	Feeding attractant
C00022625	2alpha-Hydroxyartemorin	Antihepatotoxic
C00022655	11beta,13-Dihydro-epi-ligustrin	Pigment
C00022680	5beta-Hydroxyasperilin	Antitumor
C00022759	8(17),12E-Labdadien-3alpha-ol	Phototoxic
C00022807	beta-Zeacarotene; 7',8'-Dihydro-beta,psi-carotene	Enhance plant growth
C00022811	3'4'-Didehydro-gamma,psi-carotene	Antifungal, Phytoalexin
C00022812	gamma,psi-Carotene; 5,18-Didehydro-5,6-dihydro-beta,psi-carotene	Inhibit CYP
C00022825	1,2,7,8-Tetrahydro-phi,phi-carotene	Anticancer, Antitumor
C00022827	1',2',7',8',11',12'-Hexahydrolycopene	Antidiabetic
C00022829	1',2',7,7',8,8'-Hexahydrolycopene	Anti-HSV, Antiarrhythmic, Antihyperlipidemic
C00022835	(2R,6'R)-beta,epsilon-Caroten-2-ol	Dermatitic
C00022836	(3R,6'R)-7,8-Didehydro-beta,epsilon-caroten-3-ol	Dermatitic
C00022860	(2R,3R)-beta,beta-Carotene-2,3-diol	Sex attractant
C00022867	7,8-Dihydroparasiloxanthin; 7,7',8,8'-Tetrahydro-beta,beat-carotene-3,3'-diol	Antileprotic, Tonic
C00022868	Isozeaxanthin; beta,beta-Carotene-4,4'-diol	Antipsychotic
C00022908	(3S,4S,3'R)-4-Hydroxyalloxanthin; (3S,3'R,4S)-7,7',8,8'-Tetradehydro-beta,beta-carotene-3,3',4-triol	Nutrient
C00022935	1',2'-Epoxy-1',2'-dihydro-beta,psi-carotene	Pneumotoxic
C00022940	(2R,5S,6R)-5,6-Epoxy-5,6-dihydro-beta,beta-caroten-2-ol	Antiamebic, Emetic, Expectorant
C00022976	(2R)-2-Hydroxy-beta,beta-caroten-4-one	Antileprotic, Tonic
C00022980	3'-Hydroxyechinenone; 3'-Hydroxy-beta,beta-caroten-4-one	Muscle relaxant
C00022984	Cryptocapsin; (3'S,5'R)-3'-Hydroxy-beta,kappa-caroten-6'-one	Antileprotic, Tonic
C00023072	Torularhodin	Muscle relaxant
C00023168	Semi-beta-carotenone	Phototoxic
C00023215	7,7',8,8'-Tetrahydro-beta,beta-carotene	Allelopathic
C00023307	(+)-Sophoranol N-oxide	Muscle relaxant
C00023663	Alternariol 3,4',5-Trihydroxy-6'-methyl-dibenzo[a]pyrone; Alternariol; 3,4',5-Trihydroxy-6'-methyl-dibenzo[a]pyrone	Pollinator attractant

C00023665	Altenuisol; Altertenuol 3,5,5'-Trihydroxy-4'-methoxydibenzo[a]pyrone	Antiamebic, Emetic, Expectorant
C00023744	Lathosterol; Cholest-7-en-3beta-ol	Antihypertensive
C00023983	Variecoxanthone C	Antitumor
C00024126	Abrussaponin I	Pneumotoxic
C00024127	Abrussaponin II	Antigout
C00024150	Ephemeranthol C	Anticonvulsant, Antithrombotic
C00024151	Ephemeranthoquinone	Pollinator attractant
C00024163	Fragransin B1	Phytoalexin
C00024164	Fragransin B2	Anticholinesterase
C00024165	Fragransin B3	Antihypertensive
C00024200	Gerontoxanthone I	Pollinator attractant
C00024273	Xanthoxoline	Nutrient
C00024635	Chelilutine	Antiamebic, Emetic, Expectorant
C00024642	Dihydrochelerythrine; 12,13-Dihydrochelerythrine	Repellent
C00025306	Fangchinoline; dl-Fangchinoline; 7-O-Demethyltetrandrine	Antigout
C00025338	O-Methyltriphyophylline	Pneumotoxic
C00025379	Deacetylfawcettiine; beta-Dihydroannofoline	Pneumotoxic
C00025381	Deacetylpaniculine	Antioxidant
C00025382	Dihydrolycopodine; beta-Dihydrolycopodine; 5-Deoxo-5beta-hydroxylycopodine	Antihypertensive
C00025451	Glaucamine	Hepatotoxic
C00025456	Oreogenine	Antihypertensive
C00025478	Buxamine C	Feeding attractant
C00025486	Cyclobuxargentine G	Cytotoxic
C00025490	Cyclopapilosine D	Antiamebic, Emetic
C00025552	3alpha-Benzoyloxytropane; Benzoyltropine; O-Benzoyltropine	Hepatotoxic
C00025571	Schizanthine X; (-)-Schizanthine X	Anesthetic, Narcotic
C00025573	Deoxyharringtonine	Antihyperlipidemic
C00025579	Anhydroharringtonine	Antituberculotic
C00025582	5'-Des-O-methylharringtonine	Pneumotoxic
C00025741	2'-Nortetrandrine; Cycleanorin; Cycleanorine	Antidiabetic
C00025742	2'-N-Oxyisotetrandrine	Antidiabetic
C00025873	Fenfangjine A; Tetrandrine 2beta-N-oxide; Tetrandrine 2beta-oxide	Antidiabetic
C00025874	Fenfangjine B	Antiallergic, Antibacterial, Antidiabetic, Immunosuppressant, Phytoalexin
C00026030	Secocepharanthine	Antimalarial

C00026245	Glycosmicine; 1,2,3,4-Tetrahydro-1-methyl-2,4-dioxoquinazoline; 1-Methyl-2,4(1H,3H)-quinazolinedione; 1-Methyl-2,4-quinazolinedione	Anticholinesterase
C00026263	(-)-14beta-Hydroxymatrine; 14b-Hydroxymatrine	Inhibit leaf growth
C00026269	(-)-7-Hydroxy-beta-isosparteine; 7-Hydroxypusilline	Allelopathic
C00026275	(-)-N-(3-Oxobutyl)cytisine; N-(3-Oxobutyl)cytisine	Enhance plant growth, Enhance stem growth
C00026280	(-)-Pohakuline; Pohakuline	Antigout
C00026284	(+)-14alpha-Hydroxymatrine	Teratogenic
C00026291	(+)-Darvasoline	Inhibit leaf growth
C00026308	13-Hydroxy-alpha-isolupanine; 13alpha-Hydroxyisolupanine	Oviposition attractant
C00026318	5,6-Dehydroisolupanine	Pneumotoxic
C00026339	Methyl aphyllate	Anticonvulsant, Antithrombotic
C00026415	Delbine; Delbine (Monnieria)	Analgesic, Antiallergic, Antimicrobial
C00026440	Heliparvifoline; 7-Hydroxy-6-methoxydictamnine	Phototoxic
C00026447	Isomaculosidine; 6,8-dimethoxydictamnine	Antituberculotic
C00026449	Isoskimmianine	Anti-HSV, Antiarrhythmic, Antihyperlipidemic
C00026452	Kokusagine; 7,8-Methylenedioxydictamnine	Analgesic
C00027079	Alstovine	Antihypertensive
C00027303	beta-Hydrastine	Inhibit CYP
C00027329	Corlumine; (+)-Carlumine; (+)-Corlumine; NSC 32983	Flavor
C00027354	Egenine	Antituberculotic
C00027377	Homodeoxyharringtonine; Deoxyhomoharringtonine	Pollinator attractant
C00027512	Adlumidine; (+)-Adlumidine	Antifungal
C00027537	Dehydroisocorydione	Pollinator attractant
C00027540	Fumarofine	Emetic
C00027672	12-Hydroxychelirubine	Antihypertensive
C00027723	2-Amino-3-cyclopropylbutanoic acid; (-)-2-Amino-3-cyclopropylbutanoic acid	Phytoalexin
C00027728	2'-Norisotetrandrine	Hemostatic
C00027743	Deoxyaconitine; 3-Deoxyaconitine	Phytoalexin
C00027864	Anhydrolycodoline	Inhibit CYP
C00028113	Cycleanine N-oxide; Cycleanine, 2-oxide	Enhance plant growth
C00028131	Daechuine S8-1	Cytotoxic
C00028140	Dehydrolycopecurine; 5-Ketone-lycopecurine	Pollinator attractant
C00028235	Ebinone; Ebeiedinone; Eduardine	Anti-HSV, Antiarrhythmic, Antihyperlipidemic

C00028236	Ebinone; Ebeiedinone; Eduardine	Anti-HSV, Antiarrhythmic, Antihyperlipidemic
C00028342	Herveline B	Enhance root growth
C00028352	Humosine A; (+)-Corytensine	Antimalarial
C00028393	Isoanguivine	Anti-HSV, Antiarrhythmic, Antihyperlipidemic
C00028660	N-Demethylpuqietinone	Anti-HSV, Antiarrhythmic, Antihyperlipidemic
C00028679	Neoverataline A	Antioxidant
C00028697	N-Methylcrinasiadine	Antigout
C00028746	Nupharpumilamine A	Flavor
C00028747	Nupharpumilamine B	Antitumor
C00028748	Nupharpumilamine C	Anticancer
C00028769	O-Methylrepandine	Anticonvulsant, Antidote, Antithrombotic
C00029012	Sibiricine; (+)-Sibiricine	Enhance stem growth
C00029053	Stenophylline B	Anthelmintic, Antileprotic, Tonic
C00029095	Tetrahydroveralkamine	Anti-HSV, Antiarrhythmic, Antihyperlipidemic
C00029102	Thalidezine; Thalidesine	Anti-HSV, Antiarrhythmic, Antihyperlipidemic
C00029108	Thaligosine	Antiamebic, Emetic, Expectorant
C00029109	Thaligosinine	Enhance stem growth
C00029137	Tomatidenol	Anti-HSV, Antiarrhythmic, Antihyperlipidemic
C00029194	Veramiline	Hemostatic
C00029196	Veramine	Nutrient
C00029245	Yenhusomidine	Antituberculotic
C00029255	Yunaconitine	Antimalarial
C00029279	(-)-Aschantin	Feeding attractant
C00029284	(-)-Medioresinol	Anticholinesterase
C00029352	(Z)-beta-Terpineol; cis-beta-Terpineol	Hallucinogenic
C00029358	1,3,5-Trihydroxy-6-methoxyxanthone	Phototoxic
C00029360	1,5,8-Trihydroxy-3-methyl-anthraquinone	Hemostatic
C00029361	1,5-Dihydroxy-3-methylanthraquinone; Ziganein	Analgesic, Anti-inflammatory, Antiallergic
C00029363	1,6-Dihydroxy-3-methoxy-8-methylxanthone; Griseoxanthone C	Antipsychotic
C00029515	2-(4-Hydroxyphenyl)ethanol; 4-(2-Hydroxyethyl)phenol; p-Hydroxyphenethyl alcohol; p-Tyrosol	Anti-HSV, Anticancer, Antitumor, Antiviral, Laxative
C00029518	4',5,7-Trihydroxyflavanone; 5,7,4'-Trihydroxyflavanone	Antiulcerogenic
C00029519	4,5-Dihydroxy-2-methoxy-9,10-dihydrophenanthrene	Antiallergic, Antibacterial, Antidiabetic

C00029532	4-Hydroxy-benzeneacetonitrile; p-Hydroxyphenylacetonitrile; 4-Hydroxybenzyl cyanide	Antiallergic
C00029640	Achyranthoside G	Pneumotoxic
C00029700	Apigenin 7-O-beta-D-glucopyranosyl (1->6)-beta-D-glucopyranoside	Neurotoxic
C00029807	Bellissaponin BS6; Asterbatanoside G	Enhance root growth
C00029813	Bernardioside B2	Phototoxic
C00029829	Binankadsurin A	Antiprotozoal
C00029997	(E,Z)-2,4-Decadienal; trans-2,cis-4-Decadienal	Phototoxic
C00029998	Codonopsinol; (+)-Codonopsinol	Anesthetic
C00030075	Cyclocommunin	Phototoxic
C00030135	Dioscoreside A; (-)-Dioscoreside A	Antipsychotic
C00030166	Eleutherinol	Pollinator attractant
C00030203	Erucasaponin A	Toxic
C00030288	Foliachinenoside C; (-)-Foliachinenoside C	Inhibit plant growth
C00030536	Isomitraphylline N-oxide	Cytotoxic
C00030537	Isomitraphyllinol	Anti-HSV, Antiarrhythmic, Antihyperlipidemic
C00030541	Isophytol	Antiulcerogenic
C00030578	Kadsuphilin F	Piscicidal
C00030699	Lyoniside; (+)-Lyoniside	Hemostatic
C00030707	Magnaldehyde B	Attractant
C00030906	Pachybasin	Pollinator attractant
C00031016	Phyllanthusmin A	Pollinator attractant
C00031031	Pinoresinol O-beta-D-glucopyranoside; (+)-Pinoresinol-beta-D-glucoside	Anti-HSV, Antiarrhythmic, Antihyperlipidemic
C00031115	Reinutrin; Quercetin 3-O-beta-D-xylopyranoside	Antipsychotic
C00031258	4(10)-Thujene; Sabinene	Pneumotoxic, Toxic
C00031394	Stigmast 4-ene 3,6-dione; (-)-Stigmast 4-ene 3,6-dione	Hallucinogenic
C00031484	1,3,5,6-Tetrahydroxyxanthone	Antiallergic
C00031526	1-Hydroxy-2-methyl-9,10-anthraquinone; 1-Hydroxy-2-methylanthraquinone	Immunosuppressant, Phytoalexin
C00031527	1-Hydroxy-3-methoxyanthraquinone; Xanthopurpurin 3-methyl ether	Antitumor
C00031530	1-O-Methylemodin	Laxative
C00031533	2,3-bis[(4-hydroxy-3,5-dimethoxyphenyl)-methyl]-1,4-butanediol	Antipsychotic
C00031536	2',4'-Dihydroxy-6'-methoxychalcone	Anticholinergic
C00031543	2-Acetylfuro-1,4-naphthoquinone	Analgesic, Anti-inflammatory, Antiallergic

C00031591	Alizarin 1-methyl ether	Choleretic, Phytoalexin, Pollinator attractant
C00031809	Genameside A; (-)-Genameside A	Dermatitic
C00031810	Genameside B	Hepatotoxic
C00032284	Tanegoside	Antiulcerogenic
C00032368	Toxicarol isoflavone	Anesthetic, Narcotic
C00032375	(E)-Cinnamic alcohol; trans-Cinnamyl alcohol	Analgesic, Anti-inflammatory, Antiallergic
C00032428	Tupichilignan A	Analgesic
C00032518	Xanthopurpurin	Anticancer, Antitumor
C00032520	Xanthoxyletin	Anti-inflammatory
C00032586	1,3,6-Trihydroxy-5-methoxyxanthone	Inhibit CYP
C00032637	3,5-Dihydroxy-1-methoxyxanthone	Antiparkinson, Pollinator attractant
C00032701	Alpinoside; (-)-Alpinoside	Antihypertensive
C00032750	Avicequinone C	Pollinator attractant
C00032795	Calendasaponin A; (+)-Calendasaponin A	Enhance root growth
C00032973	Fargoside D; (+)-Fargoside D	Antihypertensive
C00033049	Hymenoside W; (-)-Hymenoside W	Antiasthmatic
C00033129	Lucidumoside D	Antiamebic, Emetic, Expectorant
C00033204	Montanacin E	Antipsychotic
C00033386	Squarroside IV; (-)-Squarroside IV	Pneumotoxic
C00033446	Trojanoside A; (+)-Trojanoside A	Mutagenic
C00033513	1,10-Epoxyparthenolide	Sex attractant
C00033576	6alpha-Hydroxyxerantholide	Allelopathic, Antihyperlipidemic
C00033599	8-Desoxyartelin	Enhance plant growth, Enhance stem growth
C00033666	Bannaxanthone A	Molluscicidal
C00033711	Cauloside G; (+)-Cauloside G	Phototoxic
C00033715	Chinensiolide A; (+)-Chinensiolide A	Antihypertensive
C00033718	Chinensiolide D; (-)-Chinensiolide D	Enhance flowering, Enhance fruiting, Enhance germination, Enhance leaf growth
C00033739	Cyanidin 3-O-(2-O-beta-xylopyranosyl)-beta-galactopyranoside	Hallucinogenic
C00033742	Cyanidin 3-O-(2-O-beta-xylopyranosyl-6-O-acetyl)-beta-galactopyranoside	Hepatotoxic
C00033758	delta7 Stigmastenol; (3beta)-Stigmast-7-en-3-ol	Molluscicidal
C00033762	Dentalactone	Allelopathic, Antihyperlipidemic
C00033823	Eryngioside A	Antiparkinson, Mutagenic
C00033960	Isospiciformin	Immunomodulative
C00034198	Ribose; D-Ribose	Phytoalexin
C00034344	Xyloccensin K	Insecticidal

C00034361	11,13-Dehydrodesacetylmatricarin	Antihyperlipidemic
C00034373	1beta-Hydroxy-11-epi-colartrin	Enhance flowering, Enhance fruiting, Enhance germination, Enhance leaf growth
C00034672	Sanjoinine A	Phototoxic
C00034725	Torularhodinaldehyde	Inhibit leaf growth
C00034760	2-Decanone	Teratogenic
C00034991	1,3-Dihydroxy-2-methoxyxanthone	Antihepatotoxic
C00034993	1,7-Dihydroxy-6-methoxyxanthone	Allergenic
C00034996	16,17-Dihydro-17beta-hydroxymitraphylline	Anticholinesterase
C00035094	Eicosan-1-ol; 1-Eicosanol	Enhance root growth
C00035132	m-Mentha-6,8-diene	Sex attractant
C00035173	4,2',4'-Trihydroxychalcone	Analgesic, Anti-inflammatory, Antiallergic, Flavor
C00035187	(E)-Cinnamaldehyde	Antiparkinson
C00035196	10S,13aR-antofine N-oxide; (-)-10S,13aR-antofine N-oxide	Antimalarial
C00035199	10S,13aR-Tylophorine N-oxide	Antileprotic, Tonic
C00035244	Antofine; (-)-Antofine	Enhance plant growth, Enhance stem growth
C00035262	Ciwujianoside C2	Antihypertensive
C00035263	Ciwujianoside C3; (-)-Ciwujianoside C3	Attractant
C00035289	Dehydrotylophorine	Feeding attractant
C00035309	Ficuformodiol B	Anesthetic, Anticholinergic
C00035311	Ficuseptine B; (-)-Ficuseptine B	Feeding attractant
C00035392	Spatheliachromene	Phototoxic
C00035482	2'-Methoxyflavone	Antiviral
C00035492	3'-Methoxyflavone	Pollinator attractant
C00035687	Methyl [6]-paradol	Molluscicidal
C00035739	Salsoloside D	Inhibit root growth
C00035744	Sodoponin	Antiprotozoal
C00035754	Stilbostemin E	Sex attractant
C00035786	1S,5S-(-)-alpha-Pinene	Antiarrhythmic, Anticholinergic
C00035803	Amphoricarpolide; (+)-Amphoricarpolide	Enhance flowering, Enhance fruiting, Enhance germination, Enhance leaf growth
C00035805	Ancistroheynine A	Enhance root growth
C00035894	Verbenene	Inhibit root growth
C00035974	7-Hydroxy-8-methoxy-2-methylanthraquinone	Sex attractant
C00035996	10-Deacetylcephalomannine	Antigout
C00036065	Buddlejasaponin I	Pneumotoxic
C00036066	Buddlejasaponin II	Antigout

C00036091	Chaihunaphthone	Molluscicidal
C00036092	Chinensinaphthol	Anti-HSV, Antiarrhythmic, Antihyperlipidemic
C00036146	Mimengoside B	Pneumotoxic
C00036171	Neoboutonin	Sex attractant
C00036188	Palicoside	Inhibit leaf growth, Inhibit stem growth
C00036318	(+)-Physoperuvine	Antiasthmatic
C00036346	(3R)-Hydroxymethylbutane-1,2,3,4-tetrol; (+)-(3R)-Hydroxymethylbutane-1,2,3,4-tetrol	Antifungal
C00036369	1,2,3-Trihydroxy-5-methoxyxanthone	Inhibit root growth
C00036378	Haploxanthone; 1,3,7-Trihydroxy-8-methoxyxanthone	Antipsychotic
C00036386	1,4-Dimethoxy benzene	Phytoalexin
C00036394	1,7-Dihydroxy-4-methoxyxanthone	Pollinator attractant
C00036455	2,4,5-Trihydroxy-1-methoxyxanthone	Antipsychotic
C00036505	2-Phenylethyl O-alpha-arabinopyranosyl-(1->6)-beta-glucopyranoside; 2-Phenylethyl O-??-L-arabinopyranosyl-(1->6)-??-glucopyranoside	Antihypertensive
C00036539	3beta,4alpha-Dihydroxyguaia-11(13),10(14)-dien-12,6alpha-olide	Antidiabetic
C00036553	3-Methoxy-3',4'-methylenedioxy-trans-stilbene	Antiallergic
C00036664	8-Epiisolippidiol	Enhance flowering, Enhance fruiting, Enhance germination, Enhance leaf growth
C00036669	9,12-Octadecadienoic acid	Antidiabetic
C00036736	Araliasaponin V; Congmunoside V	Teratogenic
C00036801	beta-Methylcubebin; beta-O-Methylcubebin	Antihypertensive
C00037090	Eleganoside C	Antiprotozoal
C00037096	Enanderinanin H	Sex attractant
C00037256	Helixoside B	Enhance stem growth
C00037423	Lintetralin; (+)-Lintetralin	Antiparkinson
C00037510	Montroumarin; (+)-Montroumarin	Feeding attractant
C00037511	Montrouxanthone	Pollinator attractant
C00037517	Morusignin G; (-)-Morusignin G	Antihypertensive
C00037568	O-Acetylgeniposidic acid; 10-O-Acetylgeniposidic acid	Hormonal
C00037600	Passiguatemalin; (-)-Passiguatemalin	Inhibit CYP
C00037702	Pyramidatin C	Antileprotic, Tonic
C00037707	Pyramidatin H	Molluscicidal
C00037781	Salvianolic acid I	Pneumotoxic
C00037871	Swertinin; 1,3-Dihydroxy-7,8-dimethoxyxanthone	Sex attractant
C00037977	Urinatetralin	Antiasthmatic
C00037991	Virgatusin	Antileprotic, Tonic

C00037993	Virolongin A	Mutagenic
C00038020	Xerophilusin A; (-)-Xerophilusin A	Enhance root growth
C00038060	(-)-Pinocembrin	Anti-inflammatory
C00038109	1,2,8-Trihydroxyxanthone	Analgesic
C00038114	1,5-Dihydroxy-6-methoxyxanthone	Flavor, Insecticidal
C00038190	1-Hydroxy-7-methoxyxanthone	Antimalarial
C00038228	2-Hydroxy-3-(hydroxymethyl)anthraquinone	Pollinator attractant
C00038256	3-Hydroxymethylphenol	Anesthetic, Narcotic
C00038332	8-Hydroxycudraxanthone G	Feeding attractant
C00038409	Alnifoliol	Anthelmintic, Antileprotic, Laxative, Tonic
C00038732	Chaetoxanthone C; (+)-Chaetoxanthone C	Antihypertensive
C00038834	Copteroside H	Emetic
C00038889	cyclo-(L-Phe-L-Pro); (-)-cyclo-(L-Phe-L-Pro); Maculosine 2	Pollinator attractant
C00038907	Cynarasaponin J	Piscicidal
C00038966	Dendroflorin	Anti-HSV, Antiarrhythmic, Antihyperlipidemic
C00038990	Deoxyjesaconitine	Anticholinesterase, Antimicrobial
C00039168	Esculentoside H	Molluscicidal
C00039169	Esculentoside L	Pneumotoxic
C00039460	Itoside A; (-)-Itoside A	Enhance plant growth
C00039549	Kadsuralignan H; (+)-Kadsuralignan H	Pollinator attractant
C00039639	Longikaurin C	Pneumotoxic
C00039745	Melitric acid A	Enhance root growth
C00040042	Poliothyrsoside	Enhance stem growth
C00040098	Pseurata C	Teratogenic
C00040141	Rabdoternin E	Expectorant
C00040240	Scaberoside B3	Pneumotoxic
C00040241	Scaberoside B4	Pneumotoxic
C00040277	Seneciovaltrate	Teratogenic
C00040332	Soranjidiol	Teratogenic
C00040346	Spicatolide H; (+)-Spicatolide H	Antihypertensive
C00040362	Squadiolin C; (+)-Squadiolin C	Hormonal
C00040370	Stachysoside B	Muscle relaxant
C00040607	Uvariamicin II	Muscle relaxant
C00040739	(-)-Epiafzelechin 7-O-beta-D-glucopyranoside; Symposide	Antiulcerogenic
C00040821	3'-Hydroxyacetophenone	Anthelmintic, Antifungal, Antiseptic, Odor, Vitamin
C00040822	3-Hydroxybenzoic acid	Pollinator attractant
C00040852	8-Epihastatoside	Cytotoxic

C00040875	Annonacin A	Sex attractant
C00041050	Mallotusinin	Pollinator attractant
C00041131	Taiwanin H	Analgesic, Anti-inflammatory, Antiallergic, Antimicrobial
C00041222	(-)-Cudraflavone A	Antitussive
C00041349	Arabinose	Antidiabetic
C00041469	Cyclokuraridin; (-)-Cyclokuraridin	Teratogenic
C00041514	Effusanin B	Enhance root growth
C00041562	Gomisin G	Antidote
C00041563	Gomisin K3; Schisanhenol	Pollinator attractant
C00041647	Mairetolide F; (-)-Mairetolide F	Enhance fruiting
C00041730	Parvifoline I; (-)-Parvifoline I	Molluscicidal
C00041757	Phomallenic acid C; (-)-Phomallenic acid C	Insecticidal
C00041775	Polygalasaponin F; (+)-Polygalasaponin F	Pigment
C00041802	Rubrisandrin A	Pneumotoxic
C00041865	Siebolside B	Pneumotoxic
C00041875	Sophoraflavanone L	Antipsychotic
C00041896	Stilbostemin H 3'-beta-D-glucopyranoside; (-)-Stilbostemin H 3'-beta-D-glucopyranoside	Expectorant
C00041963	Xanthoangelol I; (+)-Xanthoangelol I	Allergenic
C00041986	(-)-Dihydroclusin	Anticonvulsant, Antithrombotic
C00042089	2-Amino-5-chloro-5-hexenoic acid	Pollinator attractant
C00042431	Cyclo(phenylalanyl-prolyl)	Anesthetic, Narcotic
C00042487	Eminensin B; (+)-Eminensin B	Antiarthritic
C00042748	Morifoline	Phototoxic
C00042776	N-Methyl-beta-phenethylamine	Allelopathic
C00042852	p-Hydroxyamphetamine	Antiulcerogenic
C00043010	Snatzkein D; (-)-Snatzkein D	Antidote
C00043109	Vittarin A	Antihypertensive
C00043571	Hernanol; (-)-Hernanol	Pollinator attractant
C00043843	Praderin	Antihypertensive
C00043898	Scabiosaponin E; (-)-Scabiosaponin E	Emetic
C00044062	4-Ethoxyphenol	Hepatotoxic, Inhibit leaf growth
C00044110	Braylin	Antiarthritic
C00044112	Bullacin	Antiamebic, Emetic, Expectorant
C00044133	Corossolin; (+)-Corossolin	Hallucinogenic
C00044174	Gigantetrocin A	Hemostatic
C00044175	Gigantetrocin B	Hemostatic
C00044202	Isomolvizarin 1	Nutrient

C00044259	Murisolin	Antileprotic, Tonic
C00044462	5-Hydroxyflavanone	Hepatotoxic
C00044476	6-Octadecynoic acid	Anticholinergic, Hepatotoxic, Pneumotoxic
C00044665	Cratoxyarborenone F	Pollinator attractant
C00044694	Deacetyl-beta-cyclopyrethrosin	Pigment
C00044741	Eleganolactone A	Pigment
C00044785	Gnetofuran B	Antiamebic, Emetic, Expectorant
C00044816	Inophyllum D; (+)-Inophyllum D	Enhance stem growth
C00044822	iso-Caracurine V	Immunomodulative
C00044966	Nigrolineaxanthone R	Teratogenic
C00045487	(2S)-7,4'-Dihydroxyflavan	Pollinator attractant
C00045573	5,7,2',4-Tetrahydroxy-3-geranylflavone	Anthelmintic, Antileprotic, Tonic
C00045652	Aristelegin A; (-)-Aristelegin A	Narcotic
C00045683	Bidentatoside I	Emetic
C00045705	Broussonin A	Antiallergic
C00045706	Broussonin B	Repellent
C00045728	Casearlucin J; (+)-Casearlucin J	Antioxidant
C00046323	Polysyphorin	Antileishmanial
C00046544	1,3-O-di-trans-p-Coumaroylglycerol	Piscicidal
C00046545	1,8-Dihydroxy-4-methylanthraquinone	Pollinator attractant
C00046654	Caruifolin B	Pollinator attractant
C00046749	Garcidepsidone D	Hormonal
C00046805	Macluraxanthone C	Piscicidal
C00046895	Sandrosaponin X; (+)-Sandrosaponin X	Antihypertensive
C00046982	(+)-Allosedridine	Enhance plant growth, Enhance stem growth
C00047138	17-Octadecen-9-ynoic acid	Insecticidal
C00047145	2alpha-Hydroxyneoanisatin; (-)-2alpha-Hydroxyneoanisatin	Antileprotic, Tonic
C00047196	Buxippine K	Hemostatic
C00047212	Cyclomicrobuxinine	Phytoalexin
C00047265	Lamenallenic acid	Hepatotoxic
C00047281	Lespecyrtin F2	Phototoxic
C00047362	23-Hydroxyursolic acid 28-O-alpha-L-rhamnopyranosyl-(1->4)-beta-D-glucopyranosyl-(1->6)-beta-D-glucopyranosyl ester	Pneumotoxic
C00047388	Arboreaside D; (+)-Arboreaside D	Antiviral
C00047725	Arisanschinin F	Enhance plant growth
C00047726	Arisanschinin G	Enhance plant growth
C00047730	Arisanschinin K	Narcotic
C00047811	Clinopodic acid D; (-)-Clinopodic acid D	Anti-HSV

C00048239	(+)-Naringenin	Phototoxic
C00048249	1,3-Dihydroxy-5-methoxy-xanthone	Antihepatotoxic
C00048262	1-p-Menthene	Analgesic, Anti-inflammatory, Antiallergic, Antimicrobial, Antioxidant, Oviposition attractant
C00048365	Cryptone	Cytotoxic
C00048480	m-Hydroxybenzoic acid	Anti-inflammatory
C00048495	Nilotinin M1	Piscicidal
C00048539	Sinuolatin B; (-)-Sinuolatin B	Antileprotic, Tonic
C00048540	Sinuolatin C; (-)-Sinuolatin C	Enhance root growth
C00049055	(+)-Buxamine F	Feeding attractant
C00049064	10-Hydroxyligstroside; (-)-10-Hydroxyligstroside	Cytotoxic
C00049237	Nipponoside C; (-)-Nipponoside C	Teratogenic
C00049271	Polyscioside A; (+)-Polyscioside A	Pneumotoxic, Toxic
C00049273	Polyscioside C; (+)-Polyscioside C	Pneumotoxic
C00049275	Polyscioside E; (+)-Polyscioside E	Feeding attractant
C00049571	Cilinaphthalide A	Antitumor
C00049574	Chinensinaphthol methyl ether	Dermatitic
C00049724	Illudin F	Anticancer
C00049725	Illudin G	Phytoalexin
C00049726	Illudin H	Antipsychotic
C00049873	FLabellin	Flavor
C00049914	(Z)-9,10-Epoxynonacosane	Hormonal
C00049915	1,3,5,6-Tetrahydroxy-8-methylxanthone	Antileprotic, Tonic
C00050110	Glaucarubol	Hepatotoxic
C00050129	Jaslanceoside E; (-)-Jaslanceoside E	Teratogenic
C00050143	Lariciresinol acetate; (+)-Lariciresinol acetate	Antiallergic
C00050212	Samaderine Z; (+)-Samaderine Z	Dermatitic
C00050267	8beta-Hydroxyparthenolide	Antiarthritic
C00050315	(+)-Cyclobuxamidine	Mutagenic

B. Source codes

List of source codes in metabolite activity prediction. This source codes below were used in activity prediction. Lines starting with # in program file are comments.

Python : Step1_Get_Network_input_from_COMPLIG_output.py

```
# -*- coding: utf-8 -*-
import csv

# Description : Generates the metabolite network file from COMPLIG output file by 3D-structural
# similarity threshold 0.95.
# Input : COMPLIG_output.txt file is the result file of COMPLIG.
# Output : Network_input_95.tsv file is metabolite network file. This file is list of high similarity
# metabolites.

class CLASS():
    def __init__(self):
        self.filename = "COMPLIG_output.txt"
        self.filename_out = "Step1_Network_input_95.tsv"
        self.threshold = 0.95

    def make_Network_input_95(self):
        with open(self.filename,"r") as f:
            reader = csv.reader(f,delimiter = "¥t")
            next(reader)
            with open(self.filename_out,"w") as f_o:
                for row in reader:
                    if self.threshold <= float(row[8]):
                        f_o.write(row[2] + "¥t" + row[3] + "¥n")

cl = CLASS()
cl.make_Network_input_95()
```

Python : Step2_Make_tailored_network.py

```
# -*- coding: utf-8 -*-
import csv

# Description : Generates tailored network as no unknown-unknown metabolite pair.
# Input : MetaboliteActivity_List.csv file is the list of metabolite-activity pair.
#         Network_input_95.tsv file is metabolite network file.
# Output : Tailored_network.tsv file is tailored network includes activity known-known metabolite
# pair or activity known-unknown metabolite pair.

class CLASS():
    def __init__(self):
        self.filename_Known_Metabolite_List = "MetaboliteActivity_List.csv"
        self.filename_Network_input_95 = "Step1_Network_input_95.txt"

        self.out_filename = "Step2_Tailored_network.tsv"

    def make_tailored_network(self):
        Known_Metabolite_set = set()
        with open(self.filename_Known_Metabolite_List,mode="r") as f:
            reader =csv.reader(f,delimiter=",")
            for row in reader:
                Known_Metabolite_set.add(row[0])

        with open(self.filename_Network_input_95,mode="r") as f:
            reader =csv.reader(f,delimiter = "¥t")
            with open(self.out_filename,mode = "w") as f_o:
                for row in reader:
                    if row[0] in Known_Metabolite_set or row[1] in Known_Metabolite_set:
                        f_o.write(row[0] + "¥t" + row[1] + "¥n")

c = CLASS()
c.make_tailored_network()
```

Python : Step4_make_Cluster_with_Bioactivity.py

```
# -*- coding: utf-8 -*-
import csv

# Description : Add activity data to the list of each cluster and metabolite.
# Input : Fout_Step2_tailored_network.tsv file is result of DPClusO.
#         MetaboliteActivity_List.csv file is the list of metabolite-activity pair.
# Output : Step4_Cluster_with_Bioactivity.txt file is arranged cluster-metabolite-activity file.

class CLASS():
    def __init__(self):
        self.filename_DPClusO_result = "Fout_Step2_tailored_network.tsv"
        self.filename_KM_list = "MetaboliteActivity_List.csv"

        self.filename_out = "Step4_Cluster_with_Bioactivity.txt"

    def make_Metabolite_list_in_Cluster(self):
        KM_set = set()
        with open(self.filename_KM_list,mode="r") as f:
            reader = csv.reader(f,delimiter=",")
            for row in reader:
                KM_set.add(row[0])

        KM_Function_list = []
        with open(self.filename_KM_list,mode="r") as f:
            reader = csv.reader(f,delimiter=",")
            for row in reader:
                KM_Function_list.append(row)

        Cluster_No = 0
        with open(self.filename_out,mode="w") as f_o:
            with open(self.filename_DPClusO_result,mode="r") as f:
                reader = csv.reader(f,delimiter = "¥t")
                for row in reader:
                    if len(row) <= 3: break
                    Cluster_No += 1
```

```
            f_o.write("CLUSTER " + str(Cluster_No) + " " + str(len(row) - 1) + " metabolites are
here." + "¥n")
            for metabolite in row[0:-1]:
                if metabolite in KM_set:
                    for KM in KM_Function_list:
                        if KM[0] == metabolite:
                            f_o.write("¥t".join(KM) + "¥n")
                            break
                else:
                    f_o.write(metabolite + "¥n")
            f_o.write("¥n")


c = CLASS()
c.make_Metabolite_list_in_Cluster()
```

Python : Step4_make_datatable_CID-Function_and_cluster-Function.py

```
# -*- coding: utf-8 -*-
import csv

# Description : Generate datatable of CID-Bioactivity and cluster-Bioactivity.
# Input : Step4_Cluster_with_Bioactivity.txt file is list of cluster-metabolite-activity.
#         MetaboliteActivity_List.csv file is the list of metabolite-activity pair.
# Output : Step4_datatable_C_ID-Bioactivity.csv file is datatable of C ID of KNApSAcK DB and
# Bioactivity.
#            Step4_datatable_Cluster-Bioactivity.csv file is datatable of cluster and Bioactivity.

class CLASS():
    def __init__(self):
        self.filename = "Step4_Cluster_with_Bioactivity.txt"
        self.filename_activityList = "MetaboliteActivity_List.csv"

        self.filename_out_1 = "Step4_datatable_C_ID-Bioactivity.csv"
        self.filename_out_2 = "Step4_datatable_Cluster-Bioactivity.csv"

    def make_datatable_Cluster_and_Func(self):
        activity_set = set()
        with open(self.filename_activityList,mode="r") as f:
            reader = csv.reader(f,delimiter=",")
            for row in reader:
                for i in row[1:]:
                    activity_set.add(i)
            activity_list = list(activity_set)
            activity_list.sort()

            activity_01 = [0] * len(activity_list)

            with open(self.filename_out_2,mode="w") as f_o:
                f_o.write(",Cluster size," + "," .join(activity_list) + "¥n")
                cluster_name = ""
                cluster_size = 0
                with open(self.filename,mode="r") as f:
```

```
            reader = csv.reader(f,delimiter="¥t")
            for row in reader:
                if len(row) != 0:
                    if row[0][0:7] == "CLUSTER":
                        cluster_size = row[0].split(" ")[2]
                        cluster_name = row[0].split(" ")[0] + " " + str(row[0].split(" ")[1])
                    elif len(row) != 1:
                        for activity in row[1:]:
                            activity_01[activity_list.index(activity)] += 1
                else:
                    f_o.write(cluster_name + "," + cluster_size + "," + ","join(map(str,activity_01)) +
"¥n")
                    activity_01 = [0] * len(activity_list)

    def make_datatable_CID_and_Func(self):
        activity_set = set()
        with open(self.filename_activityList,mode="r") as f:
            reader = csv.reader(f,delimiter=",")
            for row in reader:
                for i in row[1:]:
                    activity_set.add(i)
        activity_list = list(activity_set)
        activity_list.sort()

        C_ID_set = set()
        with open(self.filename_out_1,mode="w") as f_o:
            f_o.write("," + ","join(activity_list) + "¥n")
            activity_bit = [0] * len(activity_list)
            with open(self.filename,mode="r") as f:
                reader = csv.reader(f,delimiter="¥t")
                for row in reader:
                    if len(row) != 0 and row[0][0:4] == "C000" and row[0] not in C_ID_set:
                        C_ID_set.add(row[0])
                        activity_bit = [0] * len(activity_list)
                        for activity in row[1:]:
                            for i in range(len(activity_bit)):
```

```
                        if activity_list[i] == activity:
                            activity_bit[i] = 1
                    f_o.write(str(row[0]) + ″,″ + ″,″.join(map(str,activity_bit)) + ″¥n″)


c = CLASS()
c.make_datatable_CID_and_Func()
c.make_datatable_Cluster_and_Func()
```

Python : Step4_prepare_material_for_R.py

```
# coding: utf-8
import csv

# Description : Generate the files needed to calculate the p-value and FDR.
# Input : MetaboliteActivity_List.csv file is the list of metabolite-activity pair.
#         Tailored_network.tsv file is tailored network includes activity known-known metabolite
#         Step4_datatable_Cluster-Bioactivity.csv file is datatable of cluster and Bioactivity.
# Output : Step4_datatable_Cluster-Bioactivity_without_ClusterSize.csv file is datatable of cluster-
bioactivity without ClusterSize column.
#         Step4_ClusterSizeList.csv file is list of cluster size.
#         Step4_NetworkSize.csv file is number of metabolite in tailored network.
#        Step4_FunctionSize_in_Network.csv file is list of number of metabolite has each activity in
# network.

class CLASS:
  def __init__(self):
    self.filename_Function = "MetaboliteActivity_List.csv"
    self.filename_Network = "Step2_Tailored_network.tsv"
    self.filename_datatable = "Step4_datatable_Cluster-Bioactivity.csv"
    self.filename_out_1 = "Step4_datatable_Cluster-Bioactivity_without_ClusterSize.csv"
    self.filename_out_2 = "Step4_ClusterSizeList.csv"
    self.filename_out_3 = "Step4_NetworkSize.csv"
    self.filename_out_4 = "Step4_FunctionSize_in_Network.csv"

  def main(self):
    self.remake_datatable()
    self.make_cluster_size_file()
    self.count_metabolite_in_network()
    self.make_FunctionSize_in_Network()

  def make_FunctionSize_in_Network(self):
    function_set = set()
    with open(self.filename_datatable,mode="r") as f:
      reader = csv.reader(f,delimiter=",")
      for row in reader:
```

```
            for i in row[2:]:
                function_set.add(i)
            break
    function_list = list(function_set)
    function_list.sort()

    function_num_in_Network = [0] * len(function_list)

    metabolite_in_network_set = set()
    with open(self.filename_Network,mode="r") as f:
        reader = csv.reader(f,delimiter="¥t")
        for row in reader:
            metabolite_in_network_set.add(row[0])
            metabolite_in_network_set.add(row[1])

    with open(self.filename_Function) as f:
        reader = csv.reader(f,delimiter=",")
        for row in reader:
            if row[0] in metabolite_in_network_set:
                for func in row[1:]:
                    function_num_in_Network[function_list.index(func)] += 1

    with open(self.filename_out_4,mode="w") as f_o:
        f_o.write(",".join(map(str,function_num_in_Network)) + "¥n")

def remake_datatable(self):
    with open(self.filename_out_1,mode="w") as f_o:
        with open(self.filename_datatable,mode="r") as f:
            reader = csv.reader(f,delimiter=",")
            for row in reader:
                f_o.write(row[0] + "," + ",".join(row[2:]) + "¥n")

def make_cluster_size_file(self):
    cluster_sizes = ""
    with open(self.filename_datatable,mode="r") as f:
        reader = csv.reader(f,delimiter=",")
```

```
            next(reader)
            for row in reader:
                cluster_sizes += str(row[1]) + ","

        with open(self.filename_out_2,mode="w") as f_o:
            f_o.write(cluster_sizes[:-1] + "¥n")

    def count_metabolite_in_network(self):
        metabolite_in_Network = set()
        with open(self.filename_Network,mode="r") as f:
            reader = csv.reader(f,delimiter = "¥t")
            for row in reader:
                    metabolite_in_Network.add(row[0])
                    metabolite_in_Network.add(row[1])
        with open(self.filename_out_3,mode="w") as f_o:
            f_o.write(str(len(metabolite_in_Network)) + "¥n")

c = CLASS()
c.main()
```

Rscript : Step4_Calculate_p_value_flow.R

```
# Description : Calculate p-values and False discovery rate and generate result files.
# Input : Step4_datatable_Cluster-Bioactivity_without_ClusterSize.csv file is datatable of cluster-
# bioactivity without ClusterSize column.
#         Step4_ClusterSizeList.csv file is list of cluster size.
#         Step4_NetworkSize.csv file is number of metabolite in tailored network.
#         Step4_FunctionSize_in_Network.csv file is list of number of metabolite has each activity in
# network.
# Output : Step4_p_value_matrix.csv file is matrix of p-values.
#          Step4_FDR.csv file is list of FDR corresponded each p-values.

k_mat <- read.csv("Step4_datatable_Cluster-Bioactivity_without_ClusterSize.csv",header=T,
row.names=1)
C_List <- read.csv("Step4_ClusterSizeList.csv",header=F)
C_mat <- matrix(0,nrow=1,ncol=ncol(C_List))
for(i in 1:ncol(C_List))
{
  C_mat[i] <- C_List[1,i]
}
N <- read.csv("Step4_NetworkSize.csv",header=F)
N <- as.numeric(N)
F_List <- read.csv("Step4_FunctionSize_in_Network.csv",header=F)
F_mat <- matrix(0,nrow=1,ncol=ncol(F_List))
for(i in 1:ncol(F_List))
{
  F_mat[i] <- F_List[1,i]
}

p_value_mat <- matrix(0,nrow = nrow(k_mat),ncol = ncol(k_mat))

for(i in 1:ncol(C_mat))
{
  for(j in 1:ncol(F_mat))
  {
      p_value_mat[i,j] <- (1 - phyper((k_mat[i,j] - 1),F_mat[j],(N - F_mat[j]),C_mat[i]))
  }
```

```
}


write.csv(p_value_mat,"Step4_p_value_matrix.csv",quote=FALSE,row.names=FALSE)


p_value_vector <- as.vector(p_value_mat)
p_value_vector <- sort(p_value_vector)
FDR <- p.adjust(p_value_vector,method="BH",length(p_value_vector))
p_FDR <- c(p_value_vector,FDR)
p_FDR <- matrix(p_FDR,nrow = length(p_FDR)/2,ncol = 2)
colnames(p_FDR) <- c("p-value","FDR")
p_FDR <- unique(p_FDR)
write.table(p_FDR,"Step4_FDR.csv",sep=",",row.names = F,quote=F)
```

Python : Step4_remake_p_matrix.py

```
# coding : utf-8
import csv

# Description : Generate labeled p-value matrix of cluster and activity.
# Input : Step4_p_value_matrix.csv file is matrix of p-values.
#         Step4_ClusterSizeList.csv file is list of cluster size.
#         Step4_datatable_Cluster-Bioactivity_without_ClusterSize.csv file is datatable of cluster-
# bioactivity without ClusterSize column.
# Output : Step4_p_value_matrix_mod.csv file is ;abe;ed p-value matrix of cluster and activity.

class CLASS:
    def __init__(self):
        self.filename = "Step4_p_value_matrix.csv"
        self.filename_ClusterSizeList = "Step4_ClusterSizeList.csv"
        self.filename_FunctionNameList = "Step4_datatable_Cluster-
Bioactivity_without_ClusterSize.csv"
        self.filename_out = "Step4_p_value_matrix_mod.csv"

    def remake_p_matrix(self):
        cluster_size_list = []
        with open(self.filename_ClusterSizeList,mode="r") as f:
            reader = csv.reader(f,delimiter=",")
            for row in reader:
                for cluster_size in row:
                    cluster_size_list.append(cluster_size)
        function_list = []
        with open(self.filename_FunctionNameList,mode="r") as f:
            reader = csv.reader(f,delimiter=",")
            for row in reader:
                for func in row:
                    if func != "":
                        function_list.append(func)
                break

        with open(self.filename_out,mode="w") as f_o:
```

```
        f_o.write("CLUSTER ID,Cluster size," + ",".join(function_list) + "¥n")
        with open(self.filename,mode="r") as f:
            reader = csv.reader(f,delimiter   = ",")
            cluster_count = 0
            next(reader)
            for row in reader:
                cluster_count += 1
                f_o.write("CLUSTER " + str(cluster_count) + "," + cluster_size_list[cluster_count - 1] +
",")
                f_o.write(",".join(row) + "¥n")

c = CLASS()
c.remake_p_matrix()
```

Python : Step4_make_p_value_list.py

```
import csv

# Description : Generate list of p-values.
# Input : Step4_p_value_matrix_mod.csv file is ;abe;ed p-value matrix of cluster and activity.
# Output : Step4_p_value_list.csv file is list of p-values.

class make_p_value_list:
    def __init__(self):
        self.filename = "Step4_p_value_matrix_mod.csv"
        self.filename_out = "Step4_p_value_list.csv"

    def make_p_value_list(self):
        function_list = []
        with open(self.filename,mode='r') as f:
            reader = csv.reader(f,delimiter=",")
            for row in reader:
                for func in row:
                    function_list.append(func)
                break

        with open(self.filename_out,mode='w') as f_o:
            with open(self.filename,mode='r') as f:
                reader = csv.reader(f,delimiter=",")
                next(reader)
                for row in reader:
                    if len(row) > 0:
                        for i in range(len(row)):
                            if i > 1:
                                if float(row[i]) != 1.0:
                                    f_o.write(row[0] + "," + function_list[i] + "," + str(row[i])
+ "¥n")

c = make_p_value_list()
c.make_p_value_list()
```

Python : Step4_p_value_list_sort.py

```
import csv


# Description : Generate sorted p-value list.
# Input : Step4_p_value_list.csv file is list of p-values.
# Output : Step4_p_value_list_sorted.csv file is a list of p-values sorted in ascending order by p-value.


class CLASS:
	def __init__(self):
		self.filename = "Step4_p_value_list.csv"
		self.out_filename = "Step4_p_value_list_sorted.csv"


		self.data = [["","",0]]


	def file_sort_by_Pvalue(self):
		with open(self.filename,mode="r") as f:
			reader = csv.reader(f,delimiter = ",")
			for row in reader:
				self.data.append([str(row[0]),str(row[1]),float(row[2])])
			del self.data[0]
		self.data.sort(key=lambda x:x[2])


		with open(self.out_filename,mode="w") as f_o:
			for i in self.data:
				f_o.write(",".join(map(str,i)) + "¥n")


c = CLASS()
c.file_sort_by_Pvalue()
```

Python : Step4_p_value_select_by_FDR.py

```
# -*- coding: utf-8 -*-
import csv

# Description : Calculate the p-value corresponding to the FDR below the threshold and output the
# prediction result with that p-value.
# Input : Step4_p_value_list_sorted.csv file is a list of p-values sorted in ascending order by p-value.
# Output : Step4_p-value_list_by_FDR.csv file is predict result by FDR.

class CLASS:
    def __init__(self):
        self.filename = "Step4_p_value_list_sorted.csv"
        self.filename_for_define_threshold = "Step4_FDR.csv"
        self.out_filename = "Step4_p-value_list_by_FDR.csv"

        self.threshold_p_value = 0.0

    def define_threshold_p_value(self,FDR):
        with open(self.filename_for_define_threshold,mode="r") as f:
            reader = csv.reader(f,delimiter=",")
            threshold_FDR = float(FDR)
            temp = 0
            for row in reader:
                if row[1] != "FDR":
                    if float(row[1]) >= threshold_FDR:
                        self.threshold_p_value = temp
                        break
                    temp = float(row[0])

    def write_Metabolite_Function_List_under_threshold_Pvalue(self):
        with open(self.out_filename,mode="w") as f_o:
            with open(self.filename,mode="r") as f:
                reader = csv.reader(f,delimiter = ",")
                for row in reader:
                    if float(row[2]) <= self.threshold_p_value:
                        f_o.write(",".join(row) + "¥n")
```

```
c = CLASS()
c.define_threshold_p_value(0.05)
c.write_Metabolite_Function_List_under_threshold_Pvalue()
```

Python : Step4_make_CLUSTER-UM_List.py

```
# -*- coding: utf-8 -*-
import csv

# Description : Generate activity unknown metabolite list each cluster.
# Input : Step4_Cluster_with_Bioactivity.txt file is arranged cluster-metabolite-activity file.
# Output : Step4_CLUSTER-UM_List.csv file is list of cluster-activity unknown metabolite.

class CLASS():
    def __init__(self):
        self.filename = "Step4_Cluster_with_Bioactivity.txt"
        self.out_filename = "Step4_CLUSTER-UM_List.csv"

    def make_CLUSTER_UM_List(self):
        with open(self.out_filename,mode="w") as f_o:
            with open(self.filename,mode="r") as f:
                reader = csv.reader(f,delimiter="¥t")
                temp_str = ""
                temp_UM = []
                for row in reader:
                    if len(row) == 0:
                        f_o.write(temp_str + ",".join(temp_UM) + "¥n")
                        temp_UM = []
                    else:
                        if row[0][0:7] == "CLUSTER":
                            temp_row = row[0].split(" ")
                            temp_str = temp_row[0] + " " + str(temp_row[1]) + ","
                        elif len(row) == 1:
                            temp_UM.append(row[0])

c = CLASS()
c.make_CLUSTER_UM_List()
```

Python : Step4_make_UM-Function_List.py

```
# -*- coding: utf-8 -*-
import csv

# Description : Generate activity unknown metabolite-predicted activity list.
# Input : Step4_p-value_list_by_FDR.csv file is predict result by FDR.
#         Step4_CLUSTER-UM_List.csv file is list of cluster-activity unknown metabolite.
# Output : Step4_UM-Predicted_Function.csv file is activity unknown metabolite-predicted activity
# list.

class CLASS():
    def __init__(self):
        self.filename = "Step4_p-value_list_by_FDR.csv"
        self.filename2= "Step4_CLUSTER-UM_List.csv"
        self.out_filename = "Step4_UM-Predicted_Function.csv"

    def make_UM_Function_List(self):
        file_1 = []
        with open(self.filename,mode="r") as f:
            reader = csv.reader(f,delimiter=",")
            for row in reader:
                file_1.append(row)
        with open(self.out_filename,mode="w") as f_o:
            with open(self.filename2,mode="r") as f:
                reader = csv.reader(f,delimiter=",")
                for row in reader:
                    if not row[1] == "":
                        for row_file_1 in file_1:
                            if row[0] == row_file_1[0]:
                                for UM in row[1:]:
                                    f_o.write(str(UM) + "," + row_file_1[1] + "¥n")

c = CLASS()
c.make_UM_Function_List()
```