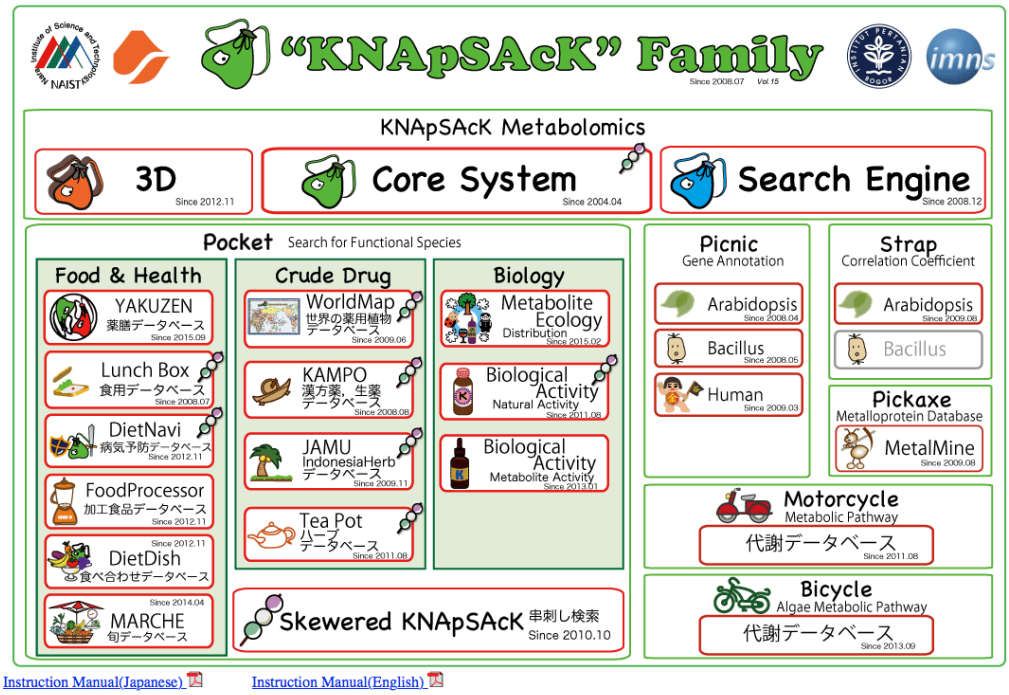
Instruction manual for KNApSAcK family

All databases can be accessed, and KNApSAcK Search Engine can be downloaded, via the website http://kanaya.naist.jp/KNApSAcK_Family/.

Please use the following reference for citation:

Farit Mochamad Afendi, Taketo Okada,, Mami Yamazaki, Aki-Hirai-Morita, Yukiko Nakamura, Kensuke Nakamura, Shun Ikeda, Hiroki Takahashi, Md. Altaf-Ul-Amin, Latifah, Darusman, Kazuki Saito, Shigehiko Kanaya, "KNApSAcK Family Databases: Integrated Metabolite-Plant Species Databases for Multifaceted Plant Research," Plant Cell Physiol., 53, e1(1-12), (2012). doi: 10.1093/pcp/pcr165.



Panel Main　Main window of "KNApSAcK" Family

The purpose of the KNApSAcK Metabolomics is to search metabolites from MS peak, molecular weight and molecular formula, and species. It consists of KNApSAcK Metabolomics Search Engine and KNApSAcK Core System.

A1 KNApSAcK Core System

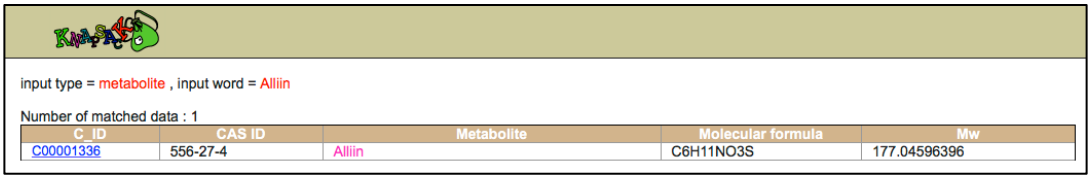
KNApSAcK Core System search metabolites and species using KNApSAcK Core DB which consists of species-metabolite relations. KNApSAcK Core system (**Panel 1**) can be accessed by clicking KNApSAcK Core System in the main window.

Panel 1

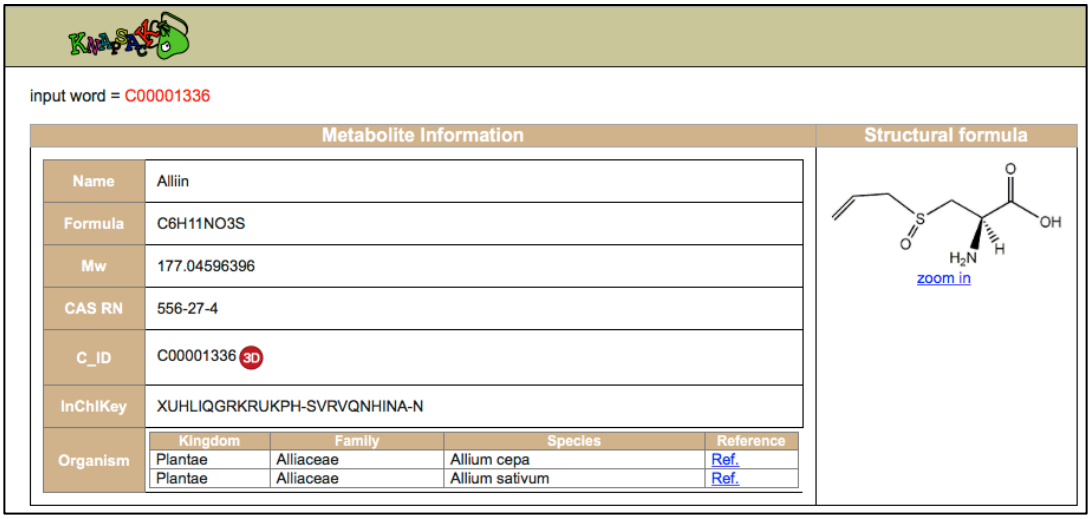
(1) Search for metabolites by molecular information

To search information about a metabolite, users should select radio button corresponding to its name, molecular formula, C_ID (identifier in metabolites in KNApSAcK Core DB) or CAS_ID, and input corresponding information and then click the “List” button. For example, a user selects radio button Metabololite, inputs “Alliin”, and clicked List button (**Panel 2**) then, **Panel 3** is obtained. Metabolite-species relations are listed when C_ID in Panel 3 is clicked (**Panel 4**).

Panel 2



Panel 3



Panel 4

(2) Search for Metabolites by species information

When a user selects radio button “Organism” and input a species name then they can obtain metabolite information. For example, a user selects radio button “Organism” inputs “Allium cepa”, and clicks List button (**Panel 5**) then, **Panel 6** is obtained. Metabolite-species relations are listed when a C_ID in Panel 6 is clicked.

Panel 5

input type = organism , input word = Allium cepa

Number of matched data : 55

C_ID	CAS ID	Metabolite	Molecular formula	Mw	Organism
C00000001	545-97-1	Gibberellin A1	C19H24O6	348.1572885	Allium cepa
C00000004	468-44-0	Gibberellin A4	C19H24O5	332.16237388	Allium cepa
C00000008	7044-72-6	Gibberellin A8	C19H24O7	364.15220312	Allium cepa
C00000009	427-77-0	Gibberellin A9	C19H24O4	316.16745925	Allium cepa
C00000012	1164-45-0	Gibberellin A12	C20H28O4	332.19875938	Allium cepa
C00000015	13744-18-8	Gibberellin A15	C20H26O4	330.18310932	Allium cepa
C00000020	19143-87-4	Gibberellin A20	C19H24O5	332.16237388	Allium cepa
C00000034	32630-92-5	Gibberellin A34	C19H24O6	348.1572885	Allium cepa
C00000044	36434-15-8	Gibberellin A44	C20H26O5	346.17802394	Allium cepa
C00000051	56978-14-4	Gibberellin A51	C19H24O5	332.16237388	Allium cepa
C00000218	6894-38-8	(-)-Jasmonic acid	C12H18O3	210.12559444	Allium cepa
C00000747	1187-84-4	L-S-methylcysteine	C4H9NO2S	135.03539927	Allium cepa
C00001242	539-86-6	Allicin	C6H10OS2	162.0173064	Allium cepa
C00001243	2179-57-9	Diallyl disulfide	C6H10S2	146.02239178	Allium cepa
C00001266	32157-29-2	Propanethial S-oxide	C3H6OS	90.01393554	Allium cepa
C00001267	107-03-9	Propane-1-thiol	C3H8S	76.03467099	Allium cepa
C00001336	556-27-4	Alliin	C6H11NO3S	177.04596396	Allium cepa
C00001389	23315-20-0	S-[(E)-Prop-1-enyl]-L-cysteine S-oxide	C6H11NO3S	177.04596396	Allium cepa
C00001495	73-03-0	Cordycepin	C10H13N5O3	251.10183932	Allium cepa
C00002374	7084-24-4	Cyanidin 3-O-glucoside	C21H21O11+	449.10838652	Allium cepa
C00002378	2611-67-8	Cyanin	C27H31O16	611.16120995	Allium cepa
C00002665	108-73-6	Phloroglucinol	C6H6O3	126.03169406	Allium cepa
C00002689	498-02-2	Acetovanillone	C9H10O3	166.06299419	Allium cepa

Panel 6

(3) Search for metabolites by metabolite and species information

In case of search for metabolite by a keyword related with both metabolite and species information, users should select radio button “All Types”, input a keyword, and click the list button. For example, a user selects radio button “All Types” inputs a keyword “alli”, and clicks the List button (**Panel 7**), then **Panel 8** is obtained. The targeted keyword is highlighted.

Panel 7

input type = all , input word = alli

Number of matched data : 957

C_ID	CAS ID	Metabolite	Molecular formula	Mw	Organism or InChIKey
C00000001	545-97-1	Gibberellin A1	C19H24O6	348.1572885	Allium cepa
C00000004	468-44-0	Gibberellin A4	C19H24O5	332.16237388	Allium cepa
C00000008	7044-72-6	Gibberellin A8	C19H24O7	364.15220312	Allium cepa
C00000009	427-77-0	Gibberellin A9	C19H24O4	316.16745925	Allium cepa
C00000012	1164-45-0	Gibberellin A12	C20H28O4	332.19875938	Allium cepa
C00000015	13744-18-8	Gibberellin A15	C20H26O4	330.18310932	Allium cepa
C00000020	19143-87-4	Gibberellin A20	C19H24O5	332.16237388	Allium cepa
C00000034	32630-92-5	Gibberellin A34	C19H24O6	348.1572885	Allium cepa
C00000044	36434-15-8	Gibberellin A44	C20H26O5	346.17802394	Allium cepa
C00000051	56978-14-4	Gibberellin A51	C19H24O5	332.16237388	Allium cepa
C00000125	4356-52-9	Glucobrassicin	C16H20N2O9S2	448.06102171	Reseda crystallina
C00000149	6750-60-3	Spathulenol	C15H24O	220.18271539	Callicarpa americana
C00000149	6750-60-3	Spathulenol	C15H24O	220.18271539	Callicarpa japonica
C00000152	7400-08-0	p-Coumaric acid	C9H8O3	164.04734412	Calligonum leucocladum

Panel 8

A2 KNApSAcK Metabolomics Search Engine

Introduction

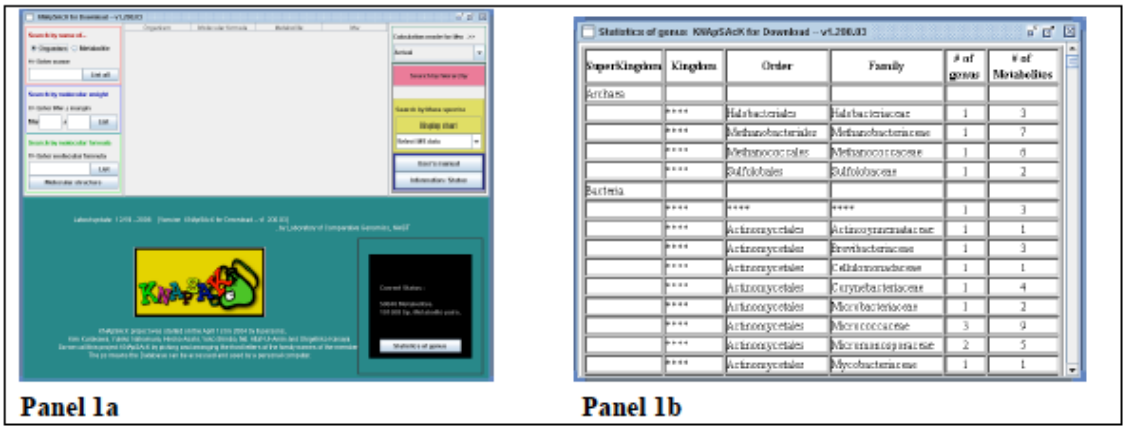
The KNApSAcK package when installed in the user's computer provides tool for analyzing his/her own datasets of mass spectra that are prepared according to a particular format, as well as for retrieving information on metabolites by entering the name of a metabolite, the name of an organism,
molecular weight or molecular formula. A list of metabolites that are associated to a taxonomic class can be obtained by search with the taxonomic name, from which information of individual metabolites can be retrieved.

Instrallation of KNApSAcK database

If and when a user wants to customize KNApSAcK Metabolomics Search Engine to use for some purpose, Java j2sdk-1.4.2 is required to be installed in the user's computer. First, the compressed file, KNApSAcK_database.zip is to be downloaded from http://kanaya.naist.jp/KNApSAcK/. Under KNApSAcK_database folder, there are two folders (spectrum data and taxonomic files), and two files (KNApSAcK.jar and knapsack.gif). User can access KNApSAcK database by clicking KNApSAcK.jar.

1. Search Options of the KNApSAcK database

The Main window of KNApSAcK is shown in Panel Main. Information on metabolites contained in the database can be searched by entering the name of metabolite, organism (scientific name), molecular weight or molecular formula. The search result is listed in the middle of the upper half of the panel. The numbers of metabolites and metabolite-species relations compiled in the present version of the database are displayed in the lower right corner of the panel. Detail information of the accumulated data in the database, for example the number of metabolites and genus in each family can be retrieved by clicking "Statistics of genus" button (Panel 1b).

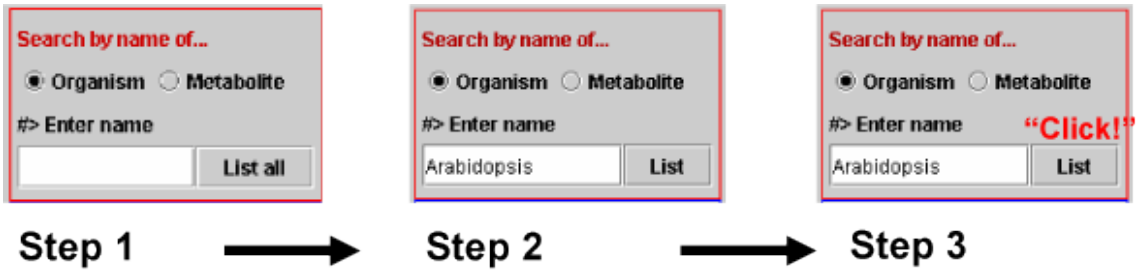


(1) Search by the name of metabolite or organism (Red panel in left side)

In case of search by organism (scientific name) or metabolite name, small and capital letters are not distinguished.

(1a) Search by the name of an organism (scientific name)

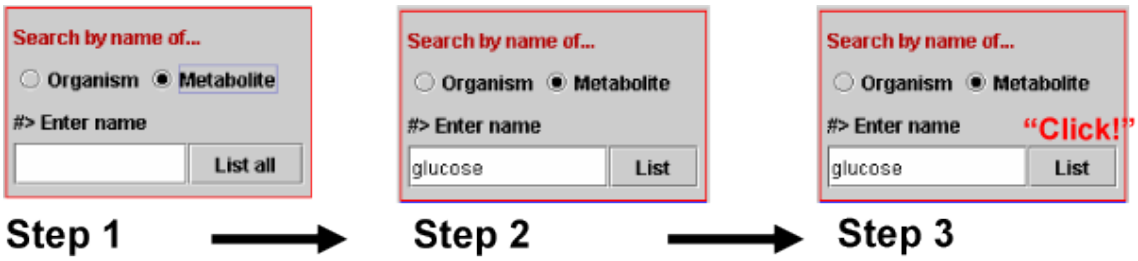
Select Organism (Step 1 in Scheme 1a), enter organism name (Step 2) and click the List button (Step 3). The entered organism name in matched to those in the database. Organism name can be the name of a species on a genus. If we input "Ara" then metabolites associated to species name with "ara" are listed. For example, metabolites associated to Arabidopsis thaliana, Marah macrocarpus and so on are listed.



Scheme 1a

(1b) Search by the name of a metabolite

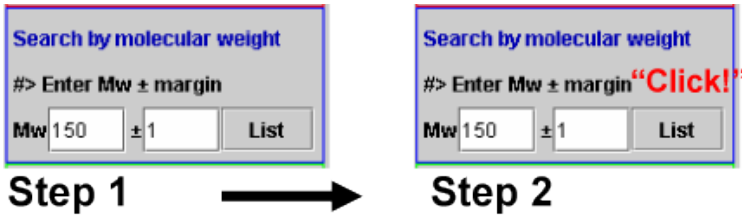
Select Metabolite (Step 1 in Scheme 1b), enter metabolite name (Step 2) and click the List button (Step 3). The entered metabolite name is matched perfectly to those in the database. If we input "glucose" then metabolites with "glucose" are listed. For example, ADP-D-glucose, D-glucose 6-phosphate and so on.



Scheme 1b

(2) Search by molecular weight (blue panel in left side)

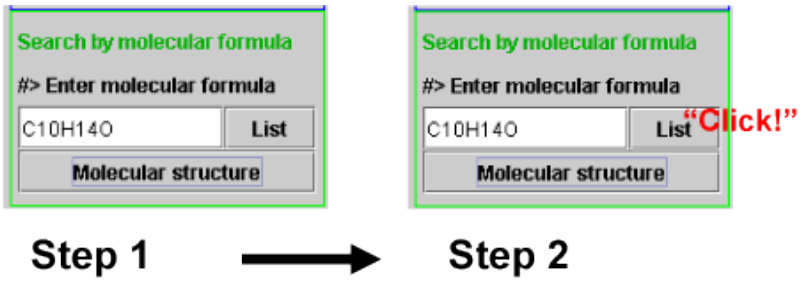
If we enter desired molecular weight (say 150) and a margin value (say 1) (Step 1 in Scheme 2), and click the List button (Step 2), the metabolites whose molecular weight are within the range 149-151 are listed.



Scheme 2

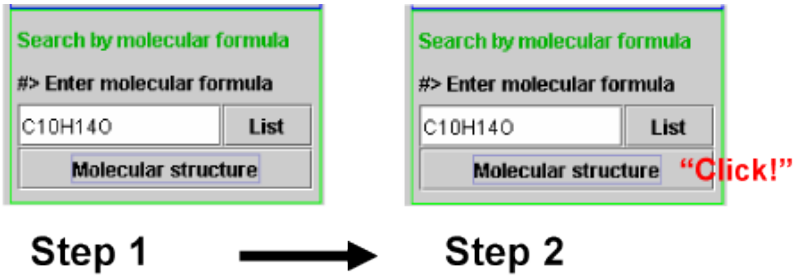
(3) Search by molecular formula

Metabolite names and origins of the metabolites are listed by molecular formula search. Enter molecular formula (Step 1 in Scheme 3a) and click the List button (Step 2).



Scheme 3a

When users are interested to know molecular structures corresponding to a molecular formula, molecular structure button should be clicked after entering the molecular formula and the molecular structure in displayed in a separate panel.

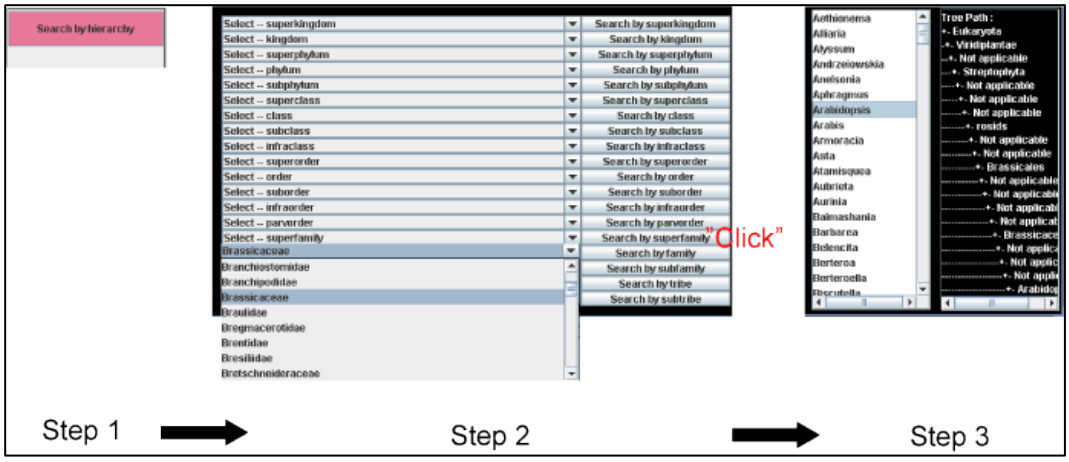


Scheme 3b

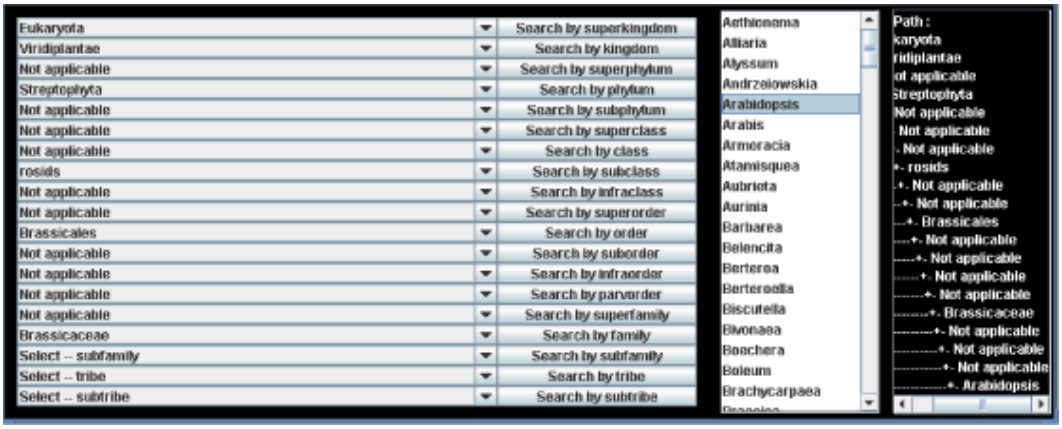
(4) Search by hierarchy

Click "Search by hierarchy" button (pink) in the right side of the panel (Step 1 in Scheme 4) and then hierarchy table appears in the bottom of the panel. Next, select any taxonomic name in any hierarchical level and click the Search button (Step 2), then genus names belonging to the selected taxonomy are listed on the right side. Next select a genus name (Step 3), then Organism names, Molecular formulae, Metabolite names and Molecular weights are listed in the upper panel. As an

example, when Brassicaceae under the taxonomic level of family is selected, the corresponding upper taxonomical levels are automatically assigned in the panel (Panel 2), that is, order, subclass, phylum, kingdom and superkingdom are automatically changed to "Brassicales", "rosides", "Streptophyta", "Viridiplantae" and "Eukaryota" respectively.



Scheme 4



Panel 2

(5) Search of compounds in mass spectra

Format of mass spectra data set

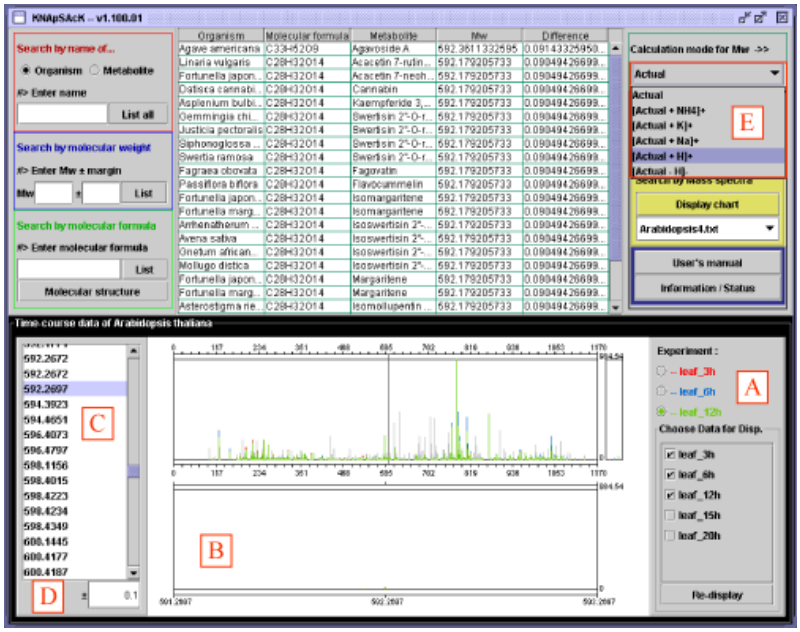
The data set must be constructed as a text file. The first line, comment line for mass spectra data, must be started by ":". The second line, attribute of individual mass spectra data, must also be started by ":". Each column is separated by a tab. The first column corresponds to m/z, and the second to last columns correspond to experimental conditions defined by user. Each line from the third to the last contains values of m/z and corresponding intensities in the individual experimental conditions. In the following example, Comment line is ":Arabidopsis T87 14days-Negative mode Scaling". In

the second line, ":Mean_Mass" is described by default, "Light", "Dark" and "Light_2" correspond to experimental conditions defined by user. The following lines contain m/z and corresponding intensities in Light, Dark and Light_2 conditions.

:Arabidopsis T87 14days-Negative mode Scaling			
:Mean_Mass	Light	Dark	Light_2
72.991712720809	0.149765559139204	0.166692818745594	0.151……
73.657442634306	0.106314886454242	0.101988938554578	0.104……
95.021470016800	0.087191317809083	0.000000000000000	0.095……
95.512902956934	0.133837666739480	0.000000000000000	0.115……
109.483588624006	0.198405127144166	0.298106748007966	0.200……
…… m/z	……	……	
……	……	……	

Format of mass spectra data

KNApSAcK_database folder contains two folders (spectrum data and taxonomic files), and two files (KNApSAcK.jar and ReadMe(KNApSAcK).txt). Save a file of mass specta data with required format as described above to spectrumdata folder, then Click KNApSAcK.jar. Select the file you want to analyze in Display chart (corresponding to Step 3 in Scheme 5a), then chart of mass spectra (Panel 3) is obtained.



Panel 3

Up to three spectra can be displayed and analyzed simultaneously by the proposed system. The spectra selected are overlaid with different colors and shown in the middle panel. Any spectrum can be brought to the front by spectrum selection (A in Panel 3). Any region of the spectra can be

enlarged by stretching the cursor horizontally and shown in the lower panel (B). All masses in the files are displayed on the left side of the panel (C). When a mass is selected from the list, a black vertical line pointer moves to the position of the peak of the mass on the spectra, and simultaneously possible metabolites corresponding to that molecular mass or masses close to that are shown in the upper panel. The margin value is changeable (D). As it is helpful to show the mass value with the value of an additive ion such as H+ and K+ depending on the solvent used for sample preparation, the species of additive ions are selectable (E). When a user set [Actual - H]- in calculation mode for Mw and select a m/z value, 95.02147002, the database system regards the molecular weight without ionization as 96.0292951019 = 95.02147002 + monoisotopic molecular weight for hydrogen ion (1.0078250319) and retrieve metabolites corresponding to this molecular weight.

B. Multifaceted Plant Usage DB

Introduction

In Multifaceted Plant Usage DB, medicinal/edible plant are related to (B1) geographical zones where the plants are used (World Map DB); (B2) species-biological activity relationship DB (Biological Activity DB); (B3) formulas for Kampo in Japan (KAMPO DB); (B4) formulas for Jamu in Indonesia (JAMU DB); (B5) edible species DB (on going, Lunch Box); and (B6) medicinal/edible herbs (on going, TeaPot). B1 to B4 are documented in this manual.

B1. World Map DB

In World Map DB system, users can search for geographic zones by a targeted plant and search for plants by geographic zones.

(1) Search for geographic zones by a targeted plant

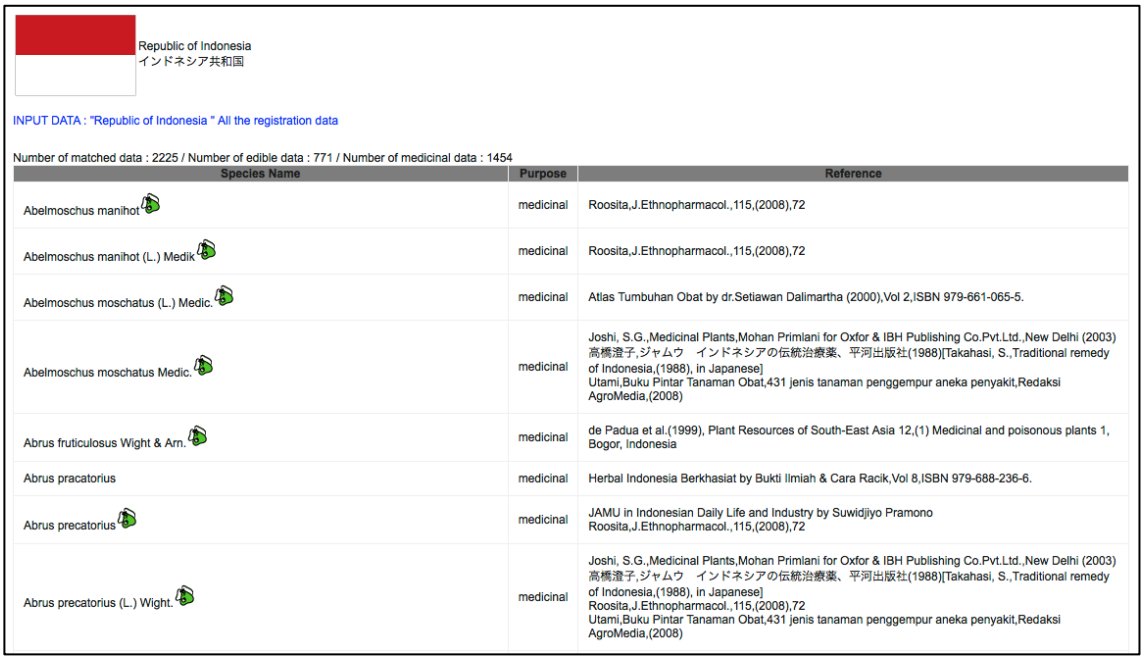
For example, when a user input “Allium cepa” and click show button (Panel 12), then national flags corresponding to geographic zones that use Allium cepa are blinked.

Panel 12

(2) Search for plants by a targeted geographic zone

When a user clicks on the nation flag of a geographic zone, then a list of medicinal/edible plants is obtained. Then by selecting KNApSAcK icon, users can obtain metabolites reported in a targeted species. In **Panel 13**, if Indonesia flag is clicked, then a list of medicinal plants utilized in Indonesia is obtained as shown in **Panel 14**. Then information on metabolites can be obtained by clicking the KNApSAcK icon "  ".

Panel 13



Republic of Indonesia
インドネシア共和国

INPUT DATA : "Republic of Indonesia " All the registration data

Number of matched data : 2225 / Number of edible data : 771 / Number of medicinal data : 1454

Species Name	Purpose	Reference
Abelmoschus manihot	medicinal	Roosita,J.Ethnopharmacol.,115,(2008),72
Abelmoschus manihot (L.) Medik	medicinal	Roosita,J.Ethnopharmacol.,115,(2008),72
Abelmoschus moschatus (L.) Medic.	medicinal	Atlas Tumbuhan Obat by dr.Setiawan Dalimartha (2000),Vol 2,ISBN 979-661-065-5.
Abelmoschus moschatus Medic.	medicinal	Joshi, S.G.,Medicinal Plants,Mohan Primlani for Oxfor & IBH Publishing Co.Pvt.Ltd.,New Delhi (2003) 高橋澄子,ジャムウ　インドネシアの伝統治療薬、平河出版社(1988)[Takahasi, S.,Traditional remedy of Indonesia,(1988), in Japanese] Utami,Buku Pintar Tanaman Obat,431 jenis tanaman penggempur aneka penyakit,Redaksi AgroMedia,(2008)
Abrus fruticulosus Wight & Arn.	medicinal	de Padua et al.(1999), Plant Resources of South-East Asia 12,(1) Medicinal and poisonous plants 1, Bogor, Indonesia
Abrus pracatorius	medicinal	Herbal Indonesia Berkhasiat by Bukti Ilmiah & Cara Racik,Vol 8,ISBN 979-688-236-6.
Abrus precatorius	medicinal	JAMU in Indonesian Daily Life and Industry by Suwidjiyo Pramono Roosita,J.Ethnopharmacol.,115,(2008),72
Abrus precatorius (L.) Wight.	medicinal	Joshi, S.G.,Medicinal Plants,Mohan Primlani for Oxfor & IBH Publishing Co.Pvt.Ltd.,New Delhi (2003) 高橋澄子,ジャムウ　インドネシアの伝統治療薬、平河出版社(1988)[Takahasi, S.,Traditional remedy of Indonesia,(1988), in Japanese] Roosita,J.Ethnopharmacol.,115,(2008),72 Utami,Buku Pintar Tanaman Obat,431 jenis tanaman penggempur aneka penyakit,Redaksi AgroMedia,(2008)

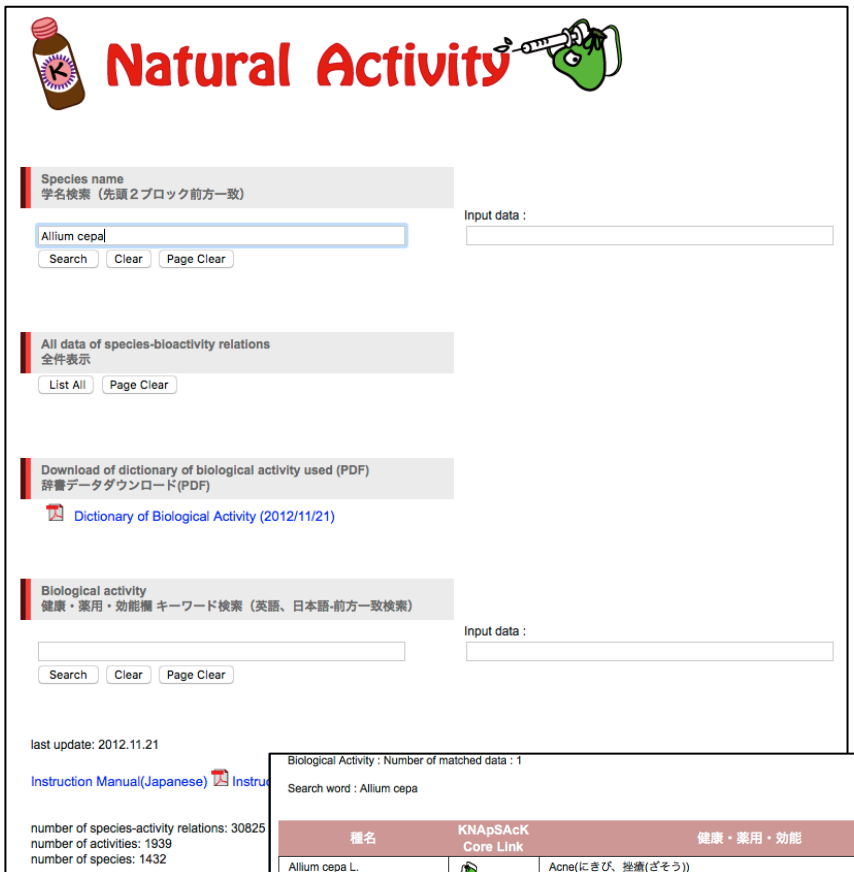
Panel 14

B2 Biological Activity DB

In Biological Activity DB system, users search for plants by a targeted biological activity.

(1) Search of Biological Activity by a species

Panel 15 is obtained by clicking Biological Activity in the main window. Users input species name in species name text box and click Search button, then a list of biological activity is obtained (Panel 16). Clicking KNApSAcK icon, users obtain metabolite information. Biological activities used in this system are downloadable from Dictionary of biological activity (shown in the middle of the window.)



Panel 15

Panel 16

(2) Search for species by its related biological activity

In the page of Biological Activity, users input a biological activity (for example, Adrenergic) in biological activity text box (**Panel 17**) and click Search button, then a list of species is obtained (**Panel 18**).

Panel 17

Panel 18

B3. KAMPO DB

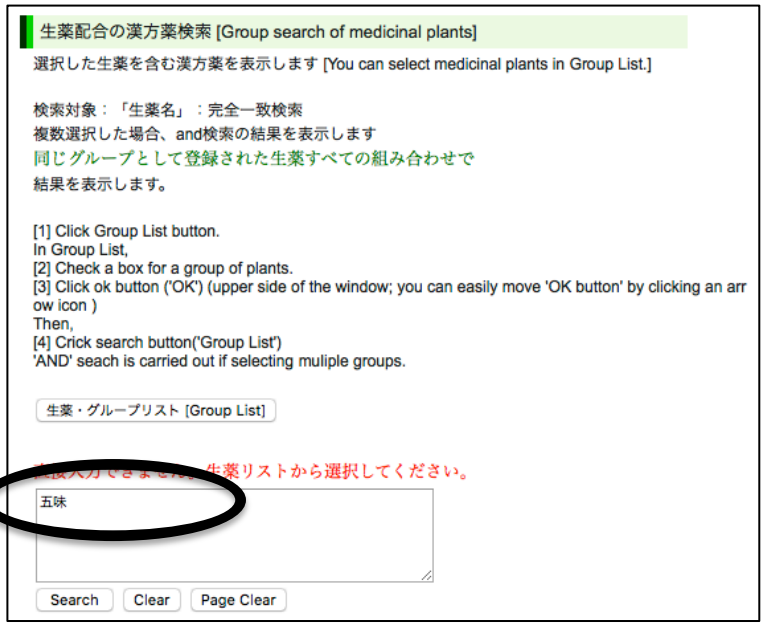
Introduction

In KAMPO DB system, users can search for medicinal plants by a blend name called a formula and search for a list of formulas by a targeted medicinal plant.

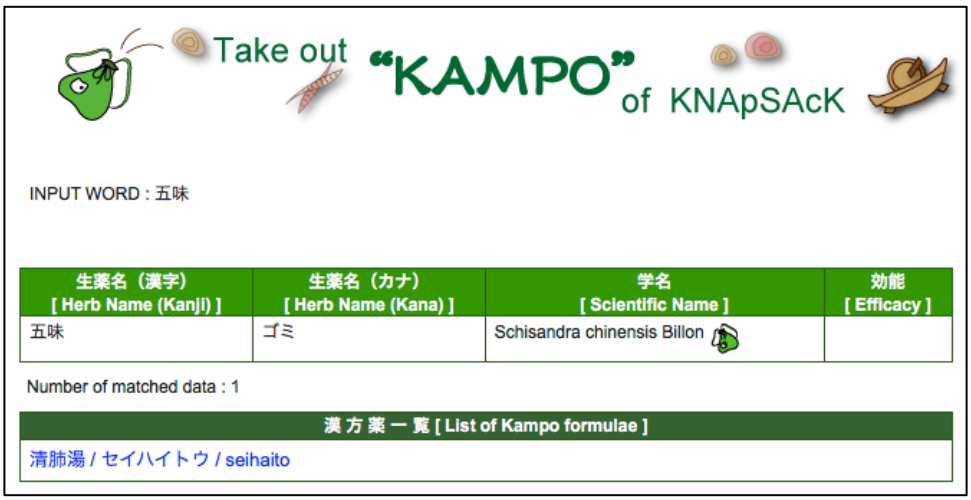
(1) Search for formula name by medicinal plant

KAMPO DB can be accessed by clicking KAMPO in the main widow. Users should select species names by clicking a box for a plant in Group List. The Group List is obtained by clicking [Group List] in Panel 19. By clicking a box corresponding to species name (for example, Schisandra chinensis is selected; Panel 20) and then clicking ok button, Schisandra chinensis is automatically written in Kanji character (Panel 21). After that by clicking search button, users can obtain a list of formulas (Panel 22; list of Kampo formula). When users select a formula in a list of Kampo formulae, then, information on blend herbal medicines can be obtained.

Panel 19 Panel 20



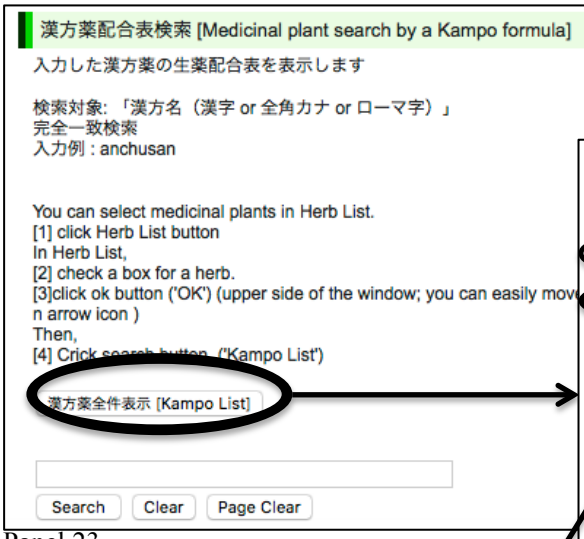
Panel 21



Panel 22

(2) Search for medicinal plants by a formula name

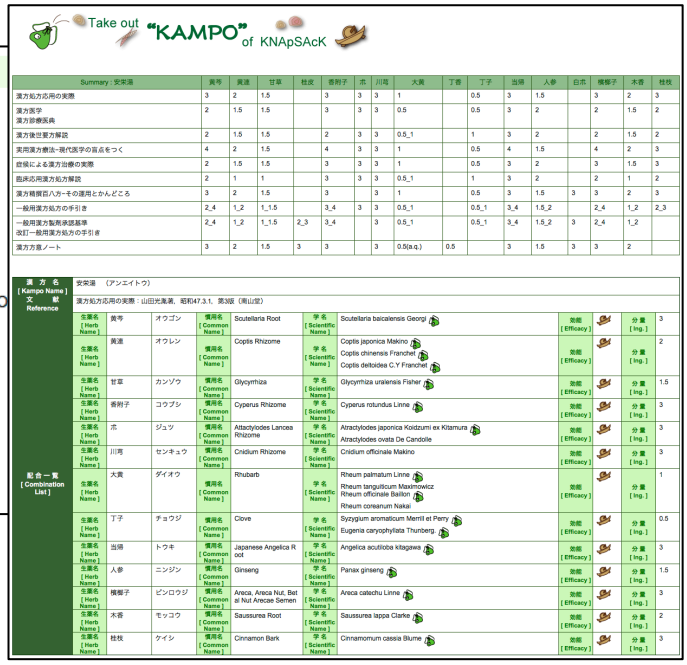
In search of medicinal plants corresponding to a targeted formula, a user should slelect the formula name by using the [Kampo List](**Panel 23**). An example of [Kampo List] is shown in **Panel 24**. The user should select one or more formula names by clicking radio button and click the “ok” button, then selected formula name is automatically written in the text box (**Panel 25**). Next, if the search button is clicked, then species included in the selected formulas are listed as **Panel 26**.



Panel 23

Panel 24

Panel 25



Panel 26

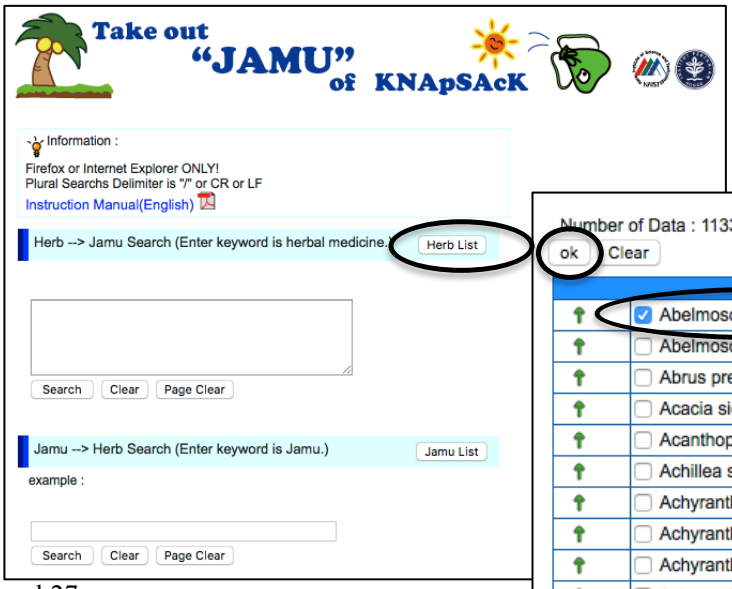
B4. JAMU DB

Introduction

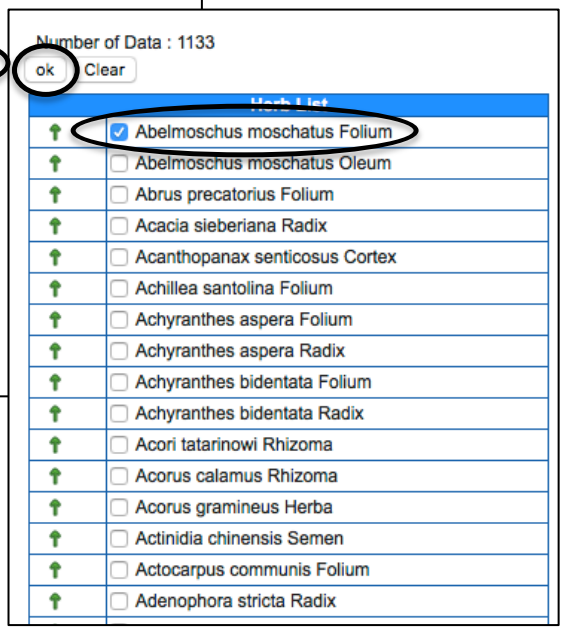
In JAMU DB system, users can search for medicinal plants by a formula and search for a list of formulas by a targeted medicinal plant.

(1) Search for formula name by medicinal plant

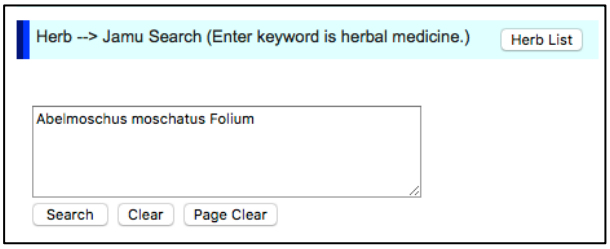
JAMU DB can be accessed by clicking JAMU in the main widow. Users should select species names by clicking a box for a plant in Group List. The Group List is obtained by clicking [Herb List] in Panel 27. By clicking a box corresponding to species name (for example, Abelmoschus moschatus is selected; Panel 28) and then clicking ok button, Abelmoschus moschatus is input in the text box (Panel 29). After that by clicking search button, users can obtain a list of formulas (Panel 30; list of JAMU formula). When users select a formula in a list of JAMU formulae, then, information on blend herbal medicines can be obtained.



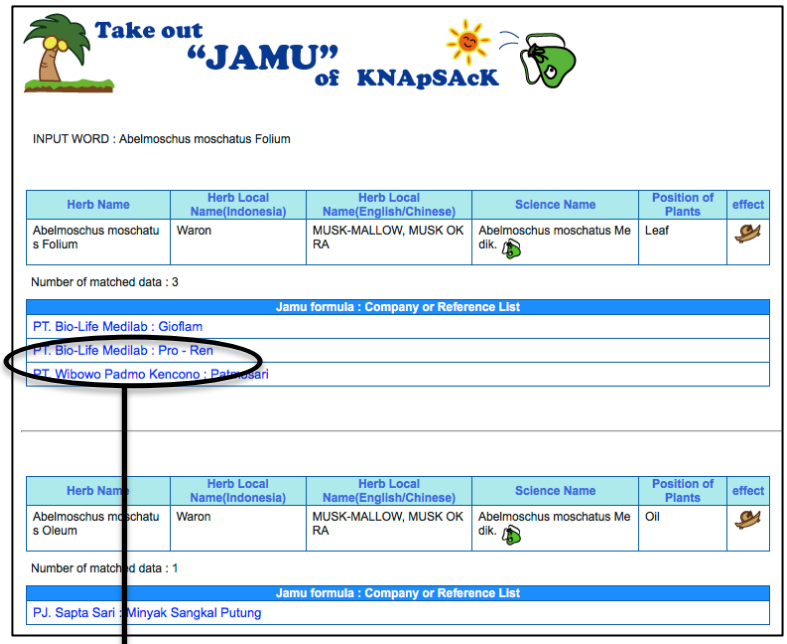
Panel 27



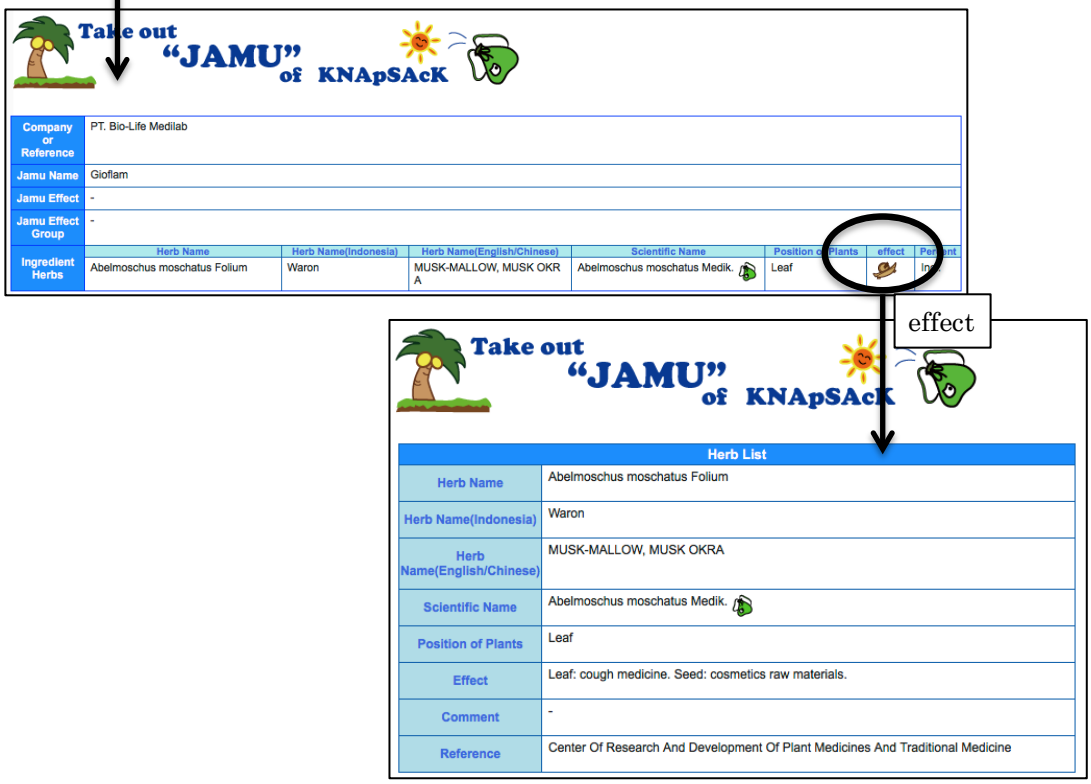
Panel 28



Panel 29

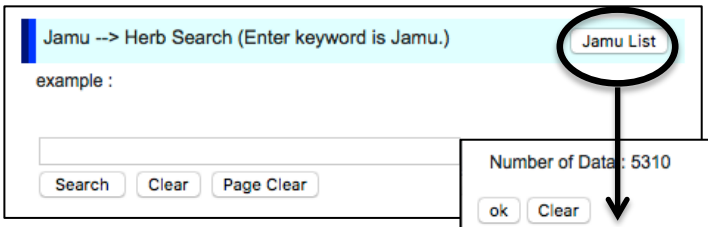


Panel 30

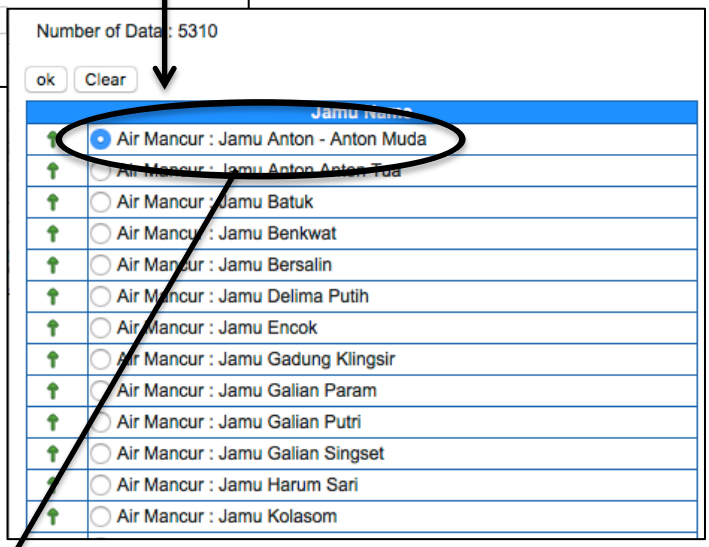


(2) Search for medicinal plants by a formula name

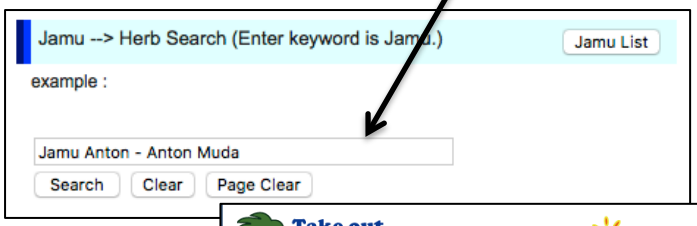
In search of medicinal plants corresponding to a targeted formula, a user should slelect the formula name by using the [JAMU LIST](**Panel 31**). An example of [JAMU List] is shown in **Panel 32**. The user should select one or more formula names by clicking radio button and click the "ok" button, then selected formula name is automatically written in the text box (**Panel 33**). Next, if the search button is clicked, then species included in the selected formulas are listed as **Panel 34**.



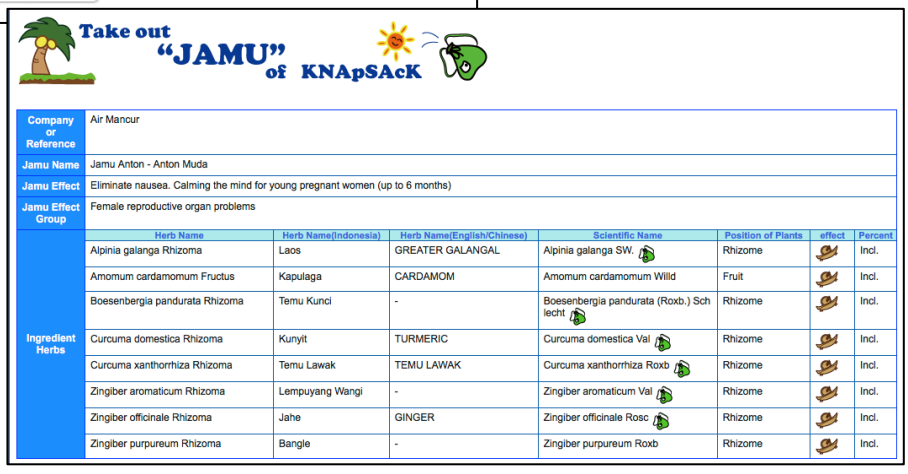
Panel 31



Panel 32



Panel 33



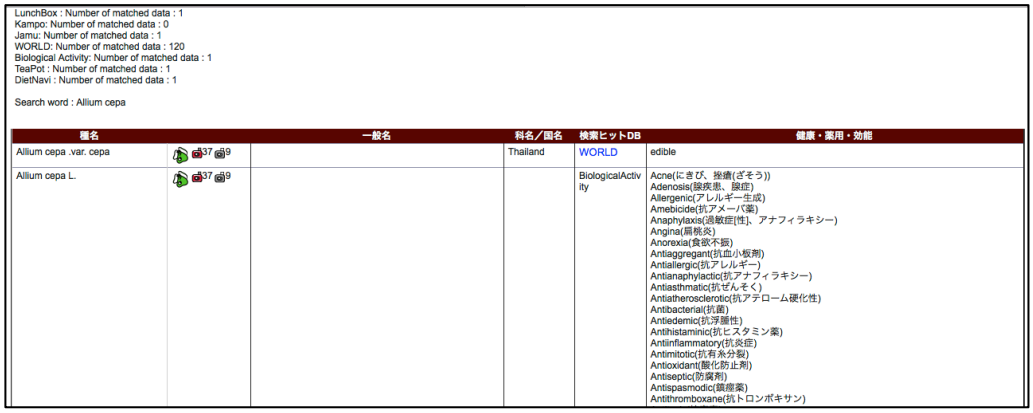
Panel 34

C. Cross-serach system (Skewered KNApSAcK)

(1) Cross search by species name

Skewered KNApSAcK makes it possible to search geographic usage, biological activity, and usage of Kampo and Jamu formulas by species name. Skewered KNApSAcK can be accessed by clicking Skewered KNApSAcK in the main widow (**Panel 35**). Users should input species name and then click Search button. Then geographic usage, biological activity, and usage of Kampo and Jamu formulas are listed as shown in **Panel 36**.

Panel 35



Panel 36

(2) Cross search by biological activity

In Skewered KNApSAcK, users can search for a list of species by a biological activity. Users should input species name in the text box and click Search button (**Panel 37**), then a list of species possessing the target activity (**Panel 38**). Then by clicking a species, users can obtain information about geographic usage, biological activity, and usage of Kampo and Jamu formulas by species name as same way of **(1) Cross search by species name**.

Panel 37

Panel 38

D. KNApSAcK

KNApSAcK project was started on the April 1st in 2004 by 6 persons, Ken Kurokawa, Yukiko Nakamura, Hiroko Asahi, Yoko Shinbo, Md. Altaf-Ul-Amin and Shigehiko Kanaya. So we call this project KNApSAcK by picking and arranging the front letters of the family names of the members. The pc means the Database can be accessed and used by a personal computer. Currently many researchers have contributed to construct KNApSAcK Family and to accumulated data.

Afend, Farit Mochamad (Graduate School of Information Science, NAIST; Biopharmaca Research Center, Bogor Agriculture Univ.)

Alatf-Ul-Amin, Md. (Graduate School of Information Science, NAIST)

Asahi, Hiroko (Graduate School of Information Science, NAIST)

Darusman, Latifah K. (Biopharmaca Research Center, Bogor Agriculture Univ.)

Hirai-Morita, Aki (Graduate School of Information Science, NAIST)

Ikeda, Shun (Graduate School of Information Science, NAIST)

Kanaya, Shigehiko (Graduate School of Information Science NAIST)

Kurokawa, Ken (Graduate School of Information Science, NAIST)

Nakamura, Kensuke (Graduate School of Information Science, NAIST)

Nakamura, Yukiko (Graduate School of Information Science, NAIST)

Okada, Takeo (Faculty of Pharmaceutical Sciences, Tokushima Bunri Univ.)

Parvin, Aziza Kawsar (Graduate School of Information Science, NAIST)

Saito, Kazuki (Gradulate School of Pharmaceutical Sciences, Chiba Univ.; RIKEN Plant Science Center)

Shibata, Daisuke (Kazusa DNA Research Institute)

Suzuki, Hideyuki (Kazusa DNA Research Institute)

Takahashi, Hiroki (Graduate School of Information Science, NAIST)

Yamazaki, Mami (Gradulate School of Pharmaceutical Sciences, Chiba Univ.)