

KNApSack Family Databases: Integrated Metabolite–Plant Species Databases for Multifaceted Plant Research

Farit Mochamad Afendi^{1,2,7}, Taketo Okada^{3,7}, Mami Yamazaki^{4,7}, Aki Hirai-Morita¹, Yukiko Nakamura¹, Kensuke Nakamura¹, Shun Ikeda¹, Hiroki Takahashi¹, Md. Altaf-Ul-Amin¹, Latifah K. Darusman⁵, Kazuki Saito^{4,6} and Shigehiko Kanaya^{1,6,*}

¹Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma-shi, Nara, 630-0192 Japan

²Department of Statistics, Bogor Agriculture University, Bogor, Indonesia

³Faculty of Pharmaceutical Sciences at Kagawa Campus, Tokushima Bunri University, Sanuki-City, Kagawa, 769-2193 Japan

⁴Graduate School of Pharmaceutical Sciences, Chiba University, Inohana 1-8-1, Chiba, 260-8675 Japan

⁵Biopharmaca Research Center, Bogor Agriculture University, Bogor, Indonesia

⁶RIKEN Plant Science Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, 230-0045 Japan

⁷These authors contributed equally to this work

*Corresponding author: E-mail, skanaya@gtc.naist.jp; Fax, +81-743-72-5956
(Received October 11, 2011; Accepted November 16, 2011)

A database (DB) describing the relationships between species and their metabolites would be useful for metabolomics research, because it targets systematic analysis of enormous numbers of organic compounds with known or unknown structures in metabolomics. We constructed an extensive species–metabolite DB for plants, the KNApSack Core DB, which contains 101,500 species–metabolite relationships encompassing 20,741 species and 50,048 metabolites. We also developed a search engine within the KNApSack Core DB for use in metabolomics research, making it possible to search for metabolites based on an accurate mass, molecular formula, metabolite name or mass spectra in several ionization modes. We also have developed databases for retrieving metabolites related to plants used for a range of purposes. In our multifaceted plant usage DB, medicinal/edible plants are related to the geographic zones (GZs) where the plants are used, their biological activities, and formulae of Japanese and Indonesian traditional medicines (Kampo and Jamu, respectively). These data are connected to the species–metabolites relationship DB within the KNApSack Core DB, keyed via the species names. All databases can be accessed via the website http://kanaya.naist.jp/KNApSack_Family/. KNApSack WorldMap DB comprises 41,548 GZ–plant pair entries, including 222 GZs and 15,240 medicinal/edible plants. The KAMPO DB consists of 336 formulae encompassing 278 medicinal plants; the JAMU DB consists of 5,310 formulae encompassing 550 medicinal plants. The Biological Activity DB consists of 2,418 biological activities and 33,706 pairwise relationships between medicinal plants and their biological activities. Current statistics of the binary relationships between individual databases were characterized by the degree distribution analysis,

leading to a prediction of at least 1,060,000 metabolites within all plants. In the future, the study of metabolomics will need to take this huge number of metabolites into consideration.

Keywords: Database • Degree distribution • KNApSack family • Metabolomics • Scale-free • Species–metabolite relationship.

Abbreviations: DB, database; GZ, geographic zone; MS, mass spectrometry.

Introduction

The metabolomics of plants is developing rapidly (Bino et al. 2004, Macel et al. 2010, Saito and Matsuda 2010, Tohge and Fernie 2010), and it will be an important topic in the systems-biological studies of interactions between plants and humans, with the goal of achieving a holistic understanding of plant function and healthcare, including the activity of medicinal plants as well as interactions between plants and their environments (Verpoorte et al. 2005, Shyur and Yang 2008, Okada et al. 2010). Among the analytical technologies for plant metabolomes, mass spectrometry (MS) plays a powerful role in the comprehensive analysis of metabolites, because it can measure the masses of >2,000 compounds within one sample, yielding high-accuracy molecular weights (Moco et al. 2007, Sawada et al. 2009). Information on numerous metabolites from various organisms has been collected in several databases (DBs), including PubChem, most of which place emphasis on biological pathways (Bairoch 2000, Goto et al. 2000, Kanehisa et al. 2002). However, the relationships between metabolites and their biological origins have not yet been

Plant Cell Physiol. 53(2): e1(1–12) (2012) doi:10.1093/pcp/pcr165, available online at www.pcp.oxfordjournals.org

© The Author 2011. Published by Oxford University Press on behalf of Japanese Society of Plant Physiologists.

All rights reserved. For permissions, please email: journals.permissions@oup.com

addressed systematically. Therefore, we initiated the construction of a species–metabolite relationship DB, the KNApSack Core DB, based on the scientific literature.

Integration between the knowledge of plant usage and existing omics data, including metabolomics, is a major challenge for the usage of systems-biological approaches to understanding the effects of components in medicinal plants on human health (Qiu 2007, Okada et al. 2010). If information on medicinal and edible plants for individual geographic zones (GZs) could be accumulated in a single DB, we could use this DB to address systematically the generality of the usage of a given targeted plant around the world. Furthermore, data regarding the acquisition of metabolites by plants help us to understand the efficacy of plants on humans, at the molecular level. To achieve this goal, we have also accumulated information regarding medicinal plants from around the world, including blended herbal medicines such as Kampo in Japan and Jamu in Indonesia, based on their respective scholarly literatures. In blended herbal medicines, efficacy against complex diseases such as cancer and dementia will be enhanced by synergistic effects from medicinal plants, because each prescription uses hundreds of natural products (Ji et al. 2009). The accumulation of information regarding the blending of plants to make herbal medicines will also play an important role in the development of new ingredients from medicinal plants for use in human healthcare.

In this study, we designed and constructed the KNApSack Core DB, which contains 101,500 species–metabolite relationships encompassing 20,741 species and 50,048 metabolites. We also developed a search engine for use in metabolomics research, which makes it possible to search for metabolites

based on accurate masses, molecular formulae, metabolite names and/or mass spectra obtained in several ionization modes. We also designed a multifaceted plant usage DB, which encompasses binary relationships between plants and the GZs in which they are utilized, between plants and their biological activities, and between plants and formulae for herbal medicines in Kampo and Jamu medicinal systems.

Concept of the DB system KNApSack family

Fig. 1 shows the concept underlying the KNApSack family, which contains two types of binary relationships: the Metabolomics DB system and the Multifaceted Plant Usage DB. The Metabolomics DB system consists of the KNApSack Metabolomics Search Engine and the species–metabolite relationship DB (KNApSack Core DB), which are freely available at <http://kanaya.naist.jp/KNApSack/>. To facilitate access to metabolite information obtained from analytical techniques and stored in the DB and thereby support the interpretation of mass spectral data, we provided a tool (KNApSack Metabolomics Search Engine) that allows users to retrieve a list of candidate metabolites that correspond to a particular molecular weight on a mass spectrum. From such data, it is possible to obtain information regarding individual metabolites.

In the Multifaceted Plant Usage DB, medicinal/edible plants are related to (i) GZs where the plants are used (KNApSack WorldMap DB); (ii) the species–biological activity relationship DB (Biological Activity DB); (iii) formulae for Kampo in Japan (KAMPO DB); and (iv) formulae for Jamu in Indonesia

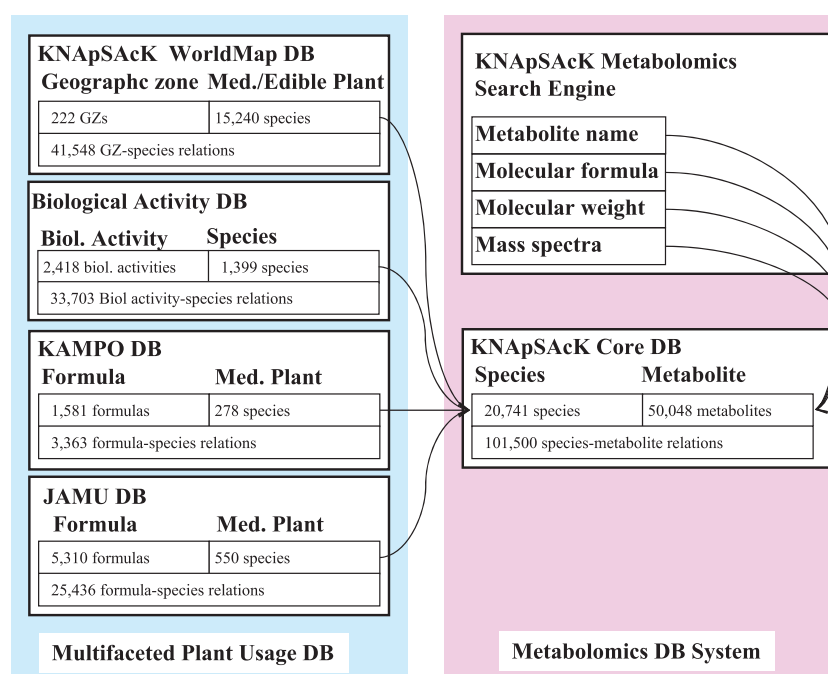


Fig. 1 The concept underlying the KNApSack family of databases. GZ stands for the geographic zone.

(JAMU DB). Those binary relationships are connected to the species–metabolite relationship DB (KNAPsACK Core DB) via the scientific name of each species. Thus, metabolites can be retrieved for a species selected by biological activity, GZs and/or the Kampo and Jamu formulae in which they are used. The Multifaceted Plant Usage DB of the KNAPsACK family can be easily extended by adding other systems (Fig. 1).

All databases can be accessed, and the KNAPsACK Search Engine can be downloaded, via the website http://kanaya.naist.jp/KNAPsACK_Family/. The main window for the KNAPsACK family is shown in Fig. 2A. An instruction manual with examples is also available via the website http://kanaya.naist.jp/KNAPsACK_Family/manual.pdf (at the bottom of the main window) and Supplement 1. Species can be selected by GZs (1 in Fig. 2A), biological activity (2), Kampo formula (3) and Jamu formula (4). A cross-search system (Skewed KNAPsACK; 5 in Fig. 2A) makes it possible to retrieve all information corresponding to a given species at once, i.e. GZs, Kampo and Jamu formulae and biological activity. Formulae and species relationships pertaining to Kampo and Jamu medicines can be retrieved via species and formula names, as shown in Fig. 2B and C, respectively. Details of individual retrieval systems are described below.

Search options of the KNAPsACK Metabolomics Search Engine for the KNAPsACK Core DB

The Metabolomics Search Engine provides tools for analyzing a user's own mass spectra data sets, stored in files prepared according to instructions downloadable using the KNAPsACK Metabolomics Search Engine. After pressing 6 in Fig. 2A, it leads to the window where a user can choose to download the database, or access the search window. By choosing WEB: Application Version, it leads to the search window shown in Fig. 3A, and the major panels for the system (Fig. 3B, C). Information on metabolites can be searched using the name of a metabolite or an organism (a in Fig. 3A), molecular weight within a specified margin (b), molecular formula (c) and/or multiple-format mass spectra detailing molecular weights and the intensity of each peak (d in Fig. 3B). As an example, the search results are displayed (h in Fig. 3A) as shown in the upper panel of Fig. 3C. Using mass spectrographic data, up to three spectra can be displayed and analyzed simultaneously (Fig. 3B). By selecting one of the available ionization modes (NH_4^+ , K^+ , Na^+ , H^+ and H^-) in Mode Selection (f in Fig. 3A), and the spectra in Select MS data (g in Fig. 3A), all masses in the files are displayed on the left side of the panel (left of Fig. 3B). When we select a mass isolated from the list, the black vertical line pointer moves to the position of the corresponding peak; simultaneously, possible metabolites corresponding to the selected molecular mass (within a specified margin) are shown in the upper panel (h in Fig. 3A). The margin for mass values is adjustable by the user (i in Fig. 3B). From a list

generated by a search (e.g. Fig. 3C), users may select a metabolite for detailed information regarding its molecular weight, formula, structure and biological functions (lower panel in Fig. 3C). The display of the molecular structure can be enlarged using the zoom-in button (j in Fig. 3C). This tool is applicable to any data that contain masses and their corresponding intensities. A tutorial for use of the KNAPsACK Metabolomics Search Engine is available via <http://kanaya.aist-nara.ac.jp/KNAPsACK/Manual/KNAPsACKManual.html>.

Multifaceted Plant Usage DB

KNAPsACK WorldMap DB. Users can retrieve medicinal/edible plants by selecting individual GZs. If the user inputs the name of a specific medicinal plant in the text box of the main window of the KNAPsACK WorldMap DB and then clicks the List button (Fig. 4A), the flags of the GZs where the target plant is utilized start blinking. The user can thereby visualize the extent of usage of a targeted plant around the world, as reported in the literature. Metabolites produced by individual species can be obtained if a KNAPsACK icon is displayed for that species. When the user clicks this icon, then the list of metabolites is displayed. When one of the metabolites is selected, the user can obtain the information on taxonomy as well as species producing the targeted metabolite (Fig. 4B).

Biological Activity DB. Users can retrieve medicinal/edible plants by inputting a search term pertaining to biological activity, including biochemical activities, diseases and human states in a text box. All of the data regarding biological activities are linked to relevant citations from the literature.

KAMPO DB. Kampo is Japanese traditional medicine. Kampo medicines consist of multiple raw herbs, referred to as Kampo formulae; the ratio of each herb in a formula is strictly fixed. The KAMPO DB contains two types of relationships: (i) groups of formulae used for identical medicinal purposes, connected to formula names; and (ii) formula names and ingredients, connected to plant names. Thus, we can retrieve formula names from the names of medicinal plants, as well as lists of ingredient plants from the groups of formulae and formula names (Fig. 2B).

JAMU DB. Jamu is Indonesian traditional medicine. Almost every region in Indonesia has its own traditional formula of Jamu; the variations depend on the plant resources in the region (Sangat et al. 2000, Adnyana and Soemardji, 2007, Pramono, 2007). By a program of the Indonesian Government, individual Jamu formulae have been systematized for use in the healthcare system. In addition to these traditional formulae, many Jamu are also produced commercially in Indonesia, and are registered in JAMU-DB, an official database at NA-DFC (National Agency for Drug and Food Control, http://www.pom.go.id/nonpublic/obat_tradisional/default.asp) of Indonesia. In the JAMU DB, we can retrieve formulae from the names of medicinal plants,

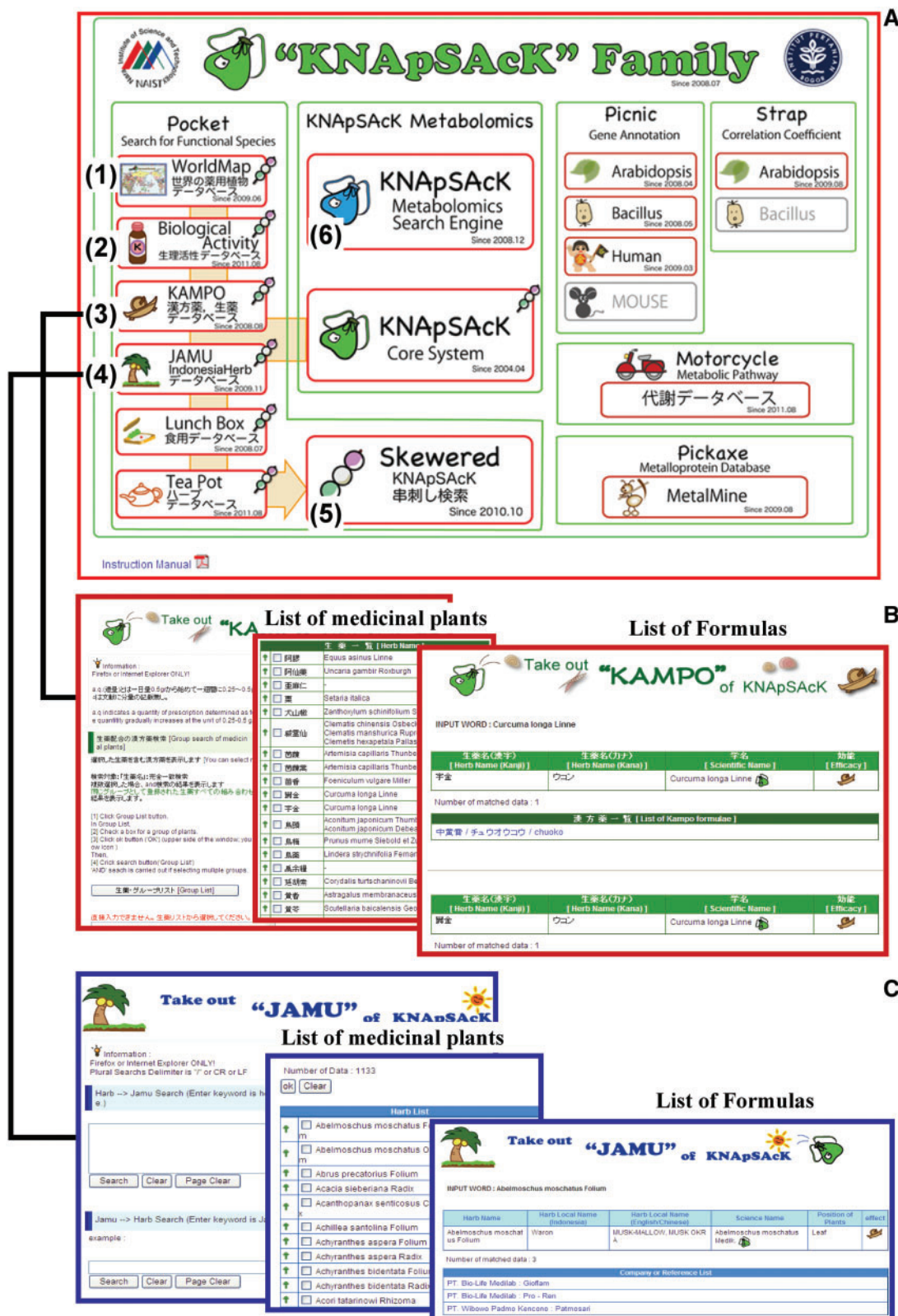


Fig. 2 KNAPsACK family database. (A) Main window for the KNAPsACK family. Users can access the geographic zone–species relationship DB by clicking the button for KNAPsACK WorldMap (1); biological activity, by clicking Biological Activity (2); Kampo medicine by clicking KAMPO (3); and Jamu medicines, by clicking JAMU (4). All information regarding geographic zone, Kampo and Jamu medicines can be retrieved by clicking Skewered KNAPsACK (5). The KNAPsACK Metabolomics Search Engine is available by choosing WEB application after clicking on KNAPsACK Metabolomics Search Engine (6). (B) Kampo medicine database. (C) Jamu medicine database. Medicinal plants and formulae can be obtained from the KAMPO and JAMU databases (B and C, respectively).

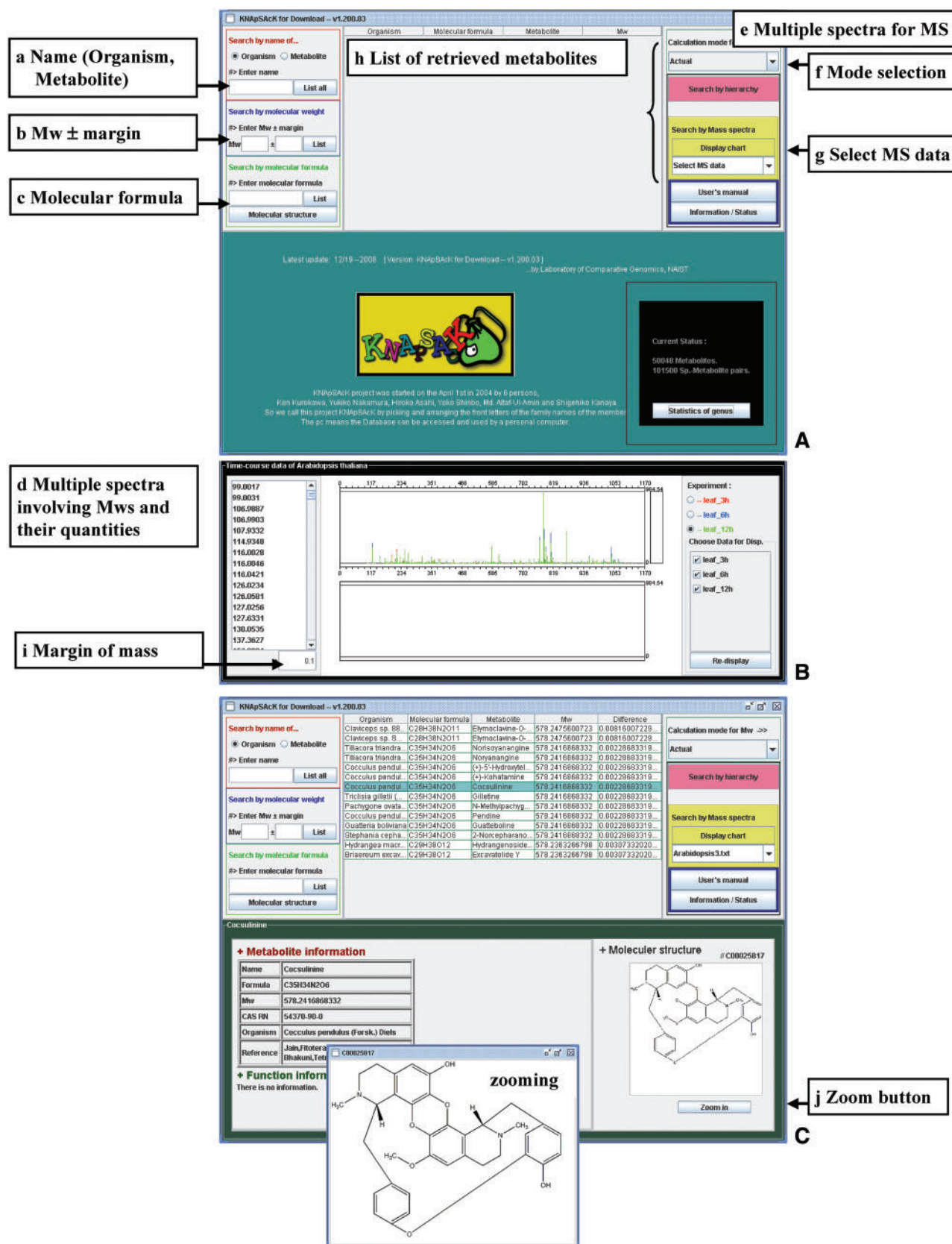


Fig. 3 KnapSack Metabolomics Search Engine (A) Main window. (B) Display of mass spectra. (C) Results of a metabolite search.

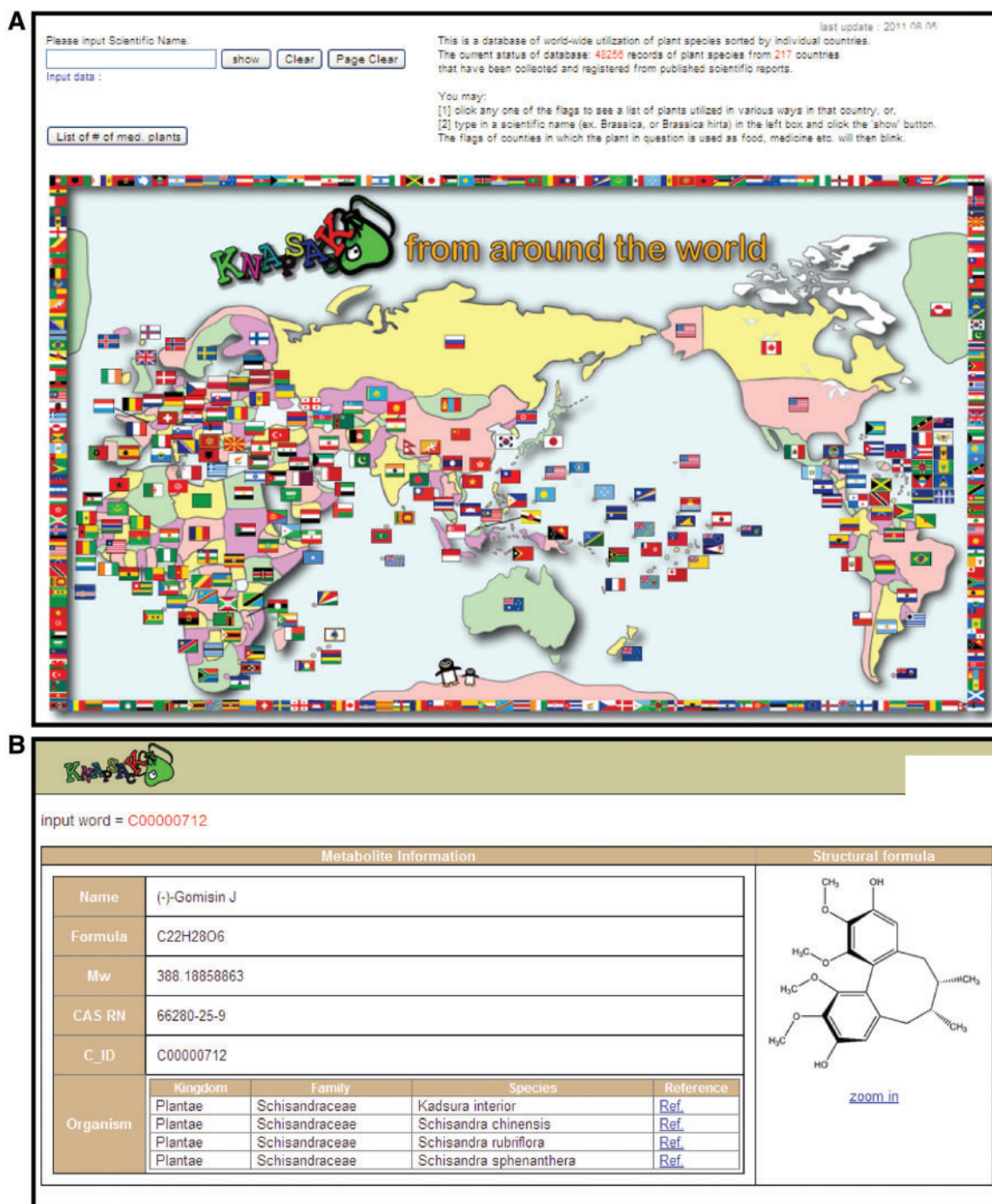


Fig. 4 (A) Window for searching plants by the geographical zone–species relationship DB (KNApSack WorldMap). Users input the scientific name in the text box and click Show, causing the geographical zone flags to blink. (B) List of plants retrieved by metabolites in the KNApSack Core DB. All databases in the Multifaceted Plant Usage DB are connected with the KNApSack Core DB.

and also the names of medicinal plants from formula names (Fig. 2C).

Skewed KNApSack. The skewed KNApSack system (5 in Fig. 2A) makes it possible to retrieve geological usage, formulae in Kampo and Jamu, as well as biological activities using the names of species and biological activities as search terms.

Current status and law unveiled in binary relationships of the KNApSack family DBs

The International Organization for Standardization (ISO3166) has defined 251 GZs based on the borders between nations and small islands. Medicinal/edible plants reported in the scientific literature have been classified into these GZs. The KNApSack

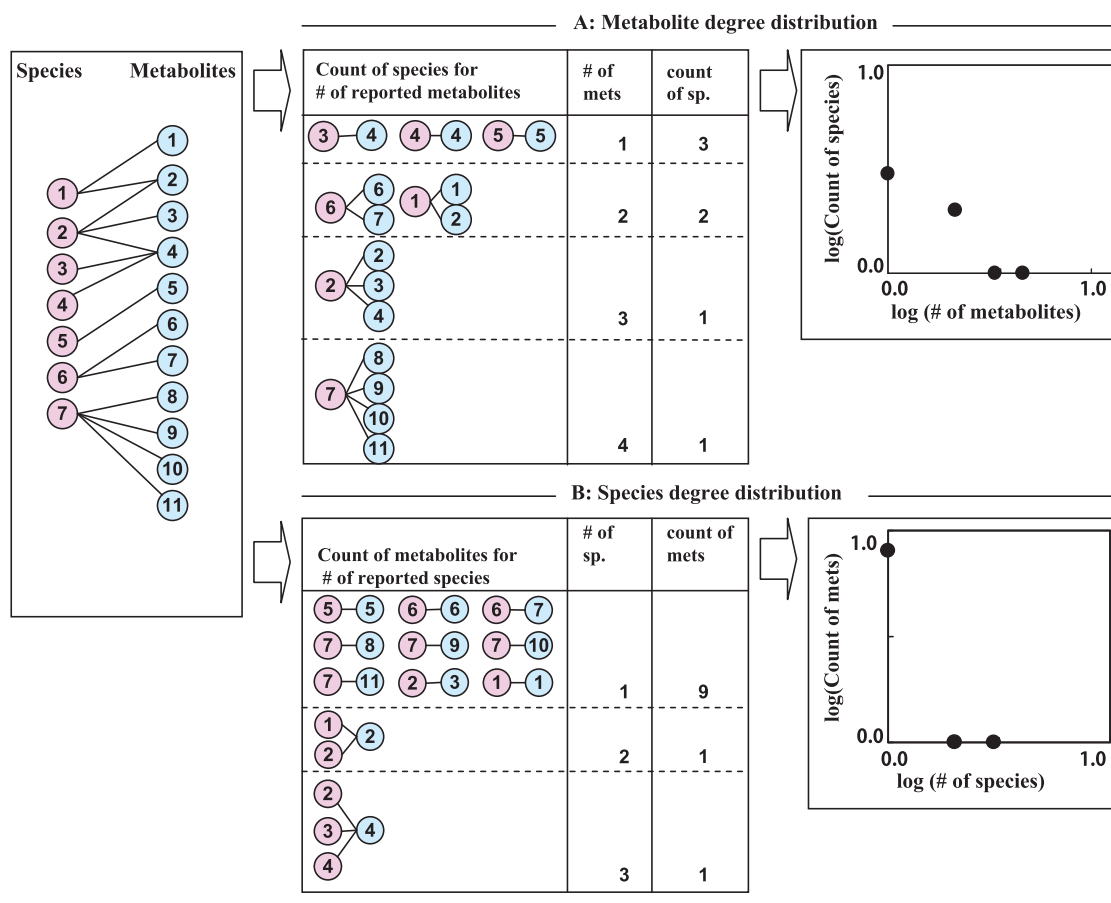


Fig. 5 Schematic representation of power law analysis in the bipartite graph. (A) Metabolite degree distribution. (B) Species degree distribution. Metabolites and species are represented by 'mets' and 'sp.', respectively. Circled numbers colored pink and blue are instances of species and metabolites.

WorldMap DB comprises 41,548 GZ–plant pair entries, including 222 GZs and 15,240 medicinal/edible plants (Fig. 1), covering 98% of the entire GZs (exceptions include ministates such as the Principality of Liechtenstein, Principality of Monaco, State of the City of the Vatican and the Principality of Andorra). The number of accumulated medicinal plants in the DB corresponds to 72% of the enumeration of the World Health Organization (21,000 medicinal species; Penso, 1980).

In the KNApSAck family DB, we represent data as binary relationships between two different attributes. In this study, the relationship between each pair of attributes was characterized by a power law analysis. Fig. 5 shows the procedure for power law analysis, using as an example a typical binary relationship between a species and a metabolite. We can obtain the number of species that produces a given set of reported metabolites (referred to as the metabolite degree distribution; Fig. 5A), as well as the number of metabolites for a given set of reported species (referred to as the species degree distribution; Fig. 5B). In this manner, two types of distributions can be obtained: the relationship on a log–log scale between numbers of metabolites and counts of species, and the relationship between numbers of

species and counts of metabolites. Species degree distributions for five attributes (GZs, formulae in Kampo and Jamu medicines, biological activities and metabolites) are shown on the left side of Figs 6A, C, E, 7A, C; degree distributions for the aforementioned attributes are shown on the right side of Figs. 6B, D, F, 7B, D.

Three Asian zones [the People's Republic of China (3,634 plants), the Republic of India (3,072) and the Kingdom of Thailand (1,888)], and two African zones [the Republic of Kenya (1936) and the Republic of Uganda (1,873)] have the largest numbers of reported medicinal/edible plants thus far. Both India and China are nations with their own holistic medicine systems, namely Ayurvedic medicine in India and TCM (traditional Chinese medicine) in China (Chan, 2005). The diversity of medicinal plants reported in the two African nations is comparable with that of the three Asian nations. The five most popular medicinal plants utilized throughout the world are *Ricinus communis*, *Psidium guajava*, *Jatropha curcas*, *Carica papaya* and *Zingiber officinale* at the latest release of the database (November 1, 2011).

A linear decreasing trend in the count of species with respect to the number of GZs can be observed on a log–log

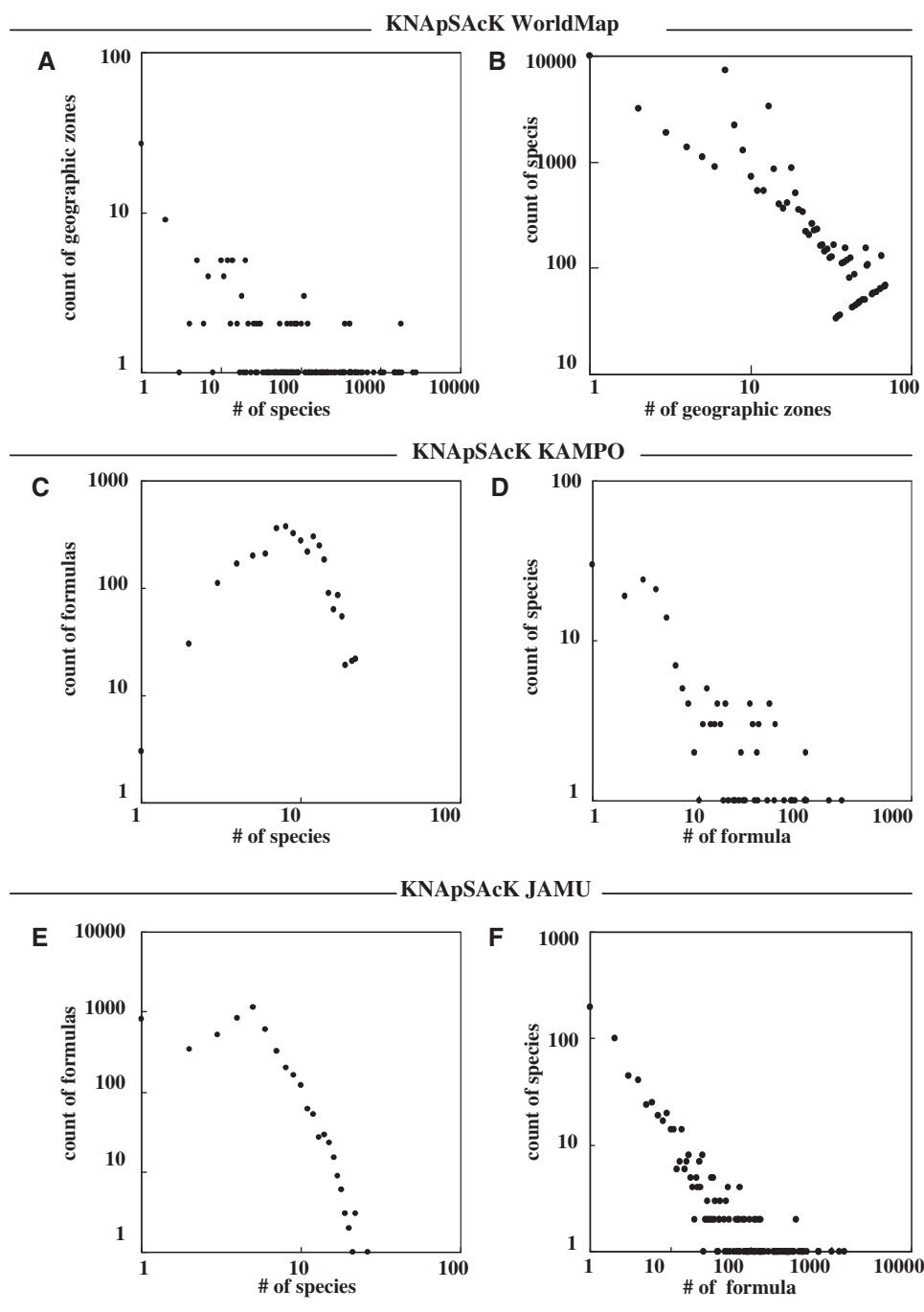


Fig. 6 Degree distributions for geographical zone–species relationships (A and B) and formula–species relationships in Kambo medicine (C and D) and Jamu medicine (E and F).

scale plot (**Fig. 6B**). This trend is called scale-free degree distribution, and can be regarded as a property of the relationship between the two attributes. We can offer a preliminary explanation of this scale-free property of the species–metabolite relationship. **Fig. 7D** shows that the metabolite degree distribution is scale free. This property is present for the following reason: at each step, when a new species (vertex in **Fig. 5**) is added, for each of the edges of the new species, either a new metabolite is

added or one is picked from among the existing ones, using preferential attachment (such a model is also referred to as growing bipartite graphs using a preferential rewiring process; Ohkubo *et al.* 2005, Guillaume and Latapy, 2006); it can be formally proven that this process exactly follows the original AB model (Albert *et al.* 1999). A near-perfect scale-free property is observed in the metabolite degree distribution (**Fig. 7D**), since a certain number of metabolites determined in a new

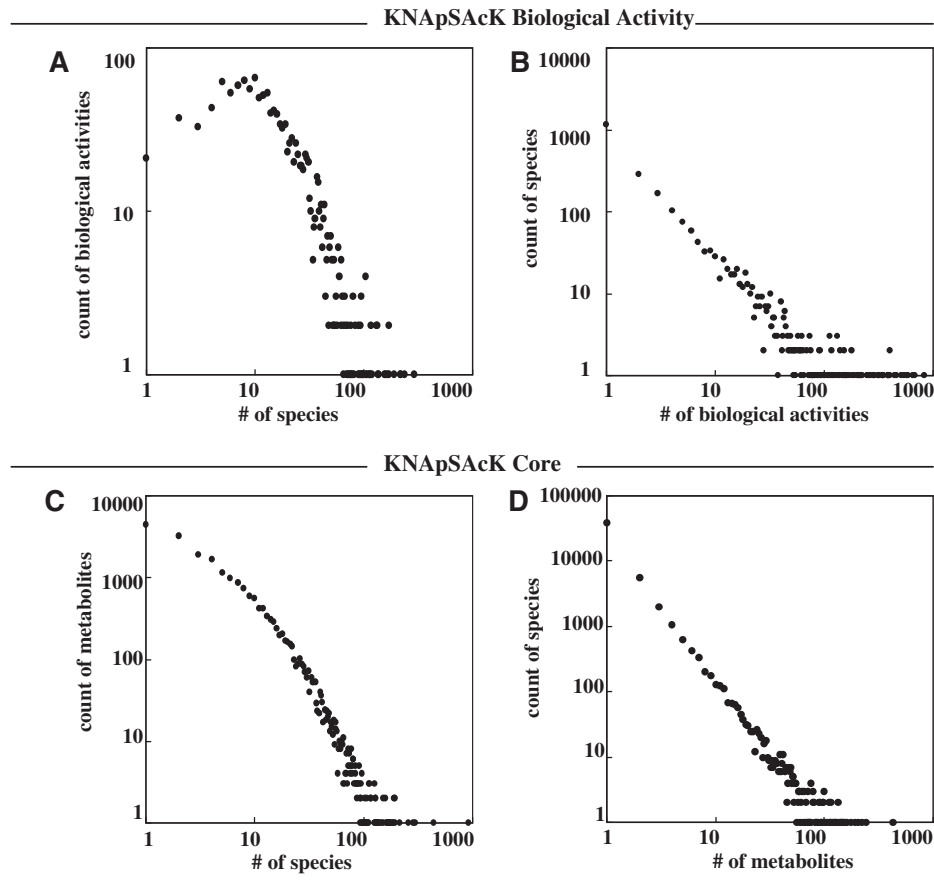


Fig. 7 Degree distributions for species–activity relationships (A and B) and species–metabolite relationships (C and D). In the species degree distribution, a linear relationship between the interval 10 and 1,000 on the x-axis (# of species) can be observed, corresponding to $\log [\text{Count of metabolites}] = -2.155 \log [\# \text{ of species}] + 4.813$. Spearman correlation coefficient, $r = -0.93$.

species have already been determined in closely related species, i.e. in species from the same genera, corresponding to preferential attachment.

A trend toward a scale-free property in GZs (Fig. 6B) can also be explained via the growth process of the bipartite graph: when a new medicinal plant is determined within a GZ, then it is preferentially attached either to nearby GZs or to GZs with high diversity.

National health insurance in Japan has approved 1,581 primary formulae of ingredients, which are classified into 336 formula names encompassed by 278 medicinal plants. The number of plant species included in Kampo formulae has a peak between seven and eight (Fig. 6C). The degree distributions of Fig. 6C and D do not exhibit the scale-free property. Kampo medicine was originally imported from China, but during the Edo period from 1600 onwards, the seclusion of Japan from the outside world led to a reduction in the number of medicinal plants used, from the initial pharmacopoeia of thousands of Chinese crude drugs (Watanabe et al. 2011). Thus, it could be argued that an artificial reduction of the number of medicinal plants used in Kampo medicines distracted the scale-free property of the formula degree

distribution, and that Kampo has therefore become a unique medicinal system in this respect.

Numerous Jamu formulae are used in Indonesia, and >7,000 commercial Jamu are registered at NA-DFC. Of these, we have accumulated 5,310 formulae encompassing 550 medicinal plants and 12 morphological segments. Formula names and plant ingredients are related in the JAMU DB. The number of species included in Jamu peaks at five (Fig. 6E). A trend in the formula degree distribution that it decreases linearly is clearly observed (Fig. 6F). This can be explained by the growth process of bipartite graphs as explained above: a new formula using new medicinal plants is developed based on knowledge of the main ingredients, supporting ingredients and those that are added simply to improve the taste of the Jamu (Beers, 2001); this process corresponds to preferential attachment. In addition, the formula degree distribution of Jamu is different from that of Kampo. It could be predicted that a reduction process of medicinal plants plays a role in the development of new types of Jamu formula by comparing the situation of Kampo formula (Fig. 6D).

No scale-free property is observed in the species degree distributions of Kampo and Jamu, because the numbers of

medicinal plants included in Kampo and Jamu medicines are limited (Fig. 6C, E). Surprisingly, 98 plants are shared between Kampo (278 species) and Jamu (550 species), corresponding to 35% of the medicinal plants in Kampo and 17% of those in Jamu. Fig. 8A shows the relationship between the number of GZs and the number of metabolites reported in these 98 shared plants. Plants utilized in the largest number of GZs are *Z. officinale*, *Foeniculum vulgare*, *Sesum indicum*, *Cyperus rotundus* and *Curucuma longa*, represented by red circles; plants with the largest number of reported metabolites are *Glycine max*, *Glycyrrhiza uralensis*, *Sophora flavescens*, *Oryza sativa* and *Panax ginseng*, represented by blue circles. Those plants are well studied at the molecular level or are used in multiple regions, i.e. they represent medicinal plants that are common around the world. Thus, the generality of usage of medicinal plants can be estimated by distinct criteria: usage in GZs and reported metabolites.

Relationships between medicinal plants and biological activities are contained in the Biological Activity DB. Currently, we have accumulated 2,418 biological activities and 33,706 pairwise relationships between medicinal plants and their biological activities. Most of the medicinal plants have 5–10 biological activities (Fig. 7A), and a clear scale-free property is observed in the biological activity degree distribution (Fig. 7B). Consequently, there is a trend toward a scale-free property in the right panels, in contrast to the left panels (Figs. 6, 7), which can be explained by the concept that a preferential attachment is acquired when a new species is added.

How many plant metabolites are there on Earth?

The latest release of the KNApSack Core DB contains 101,500 species–metabolite relationships involving 50,048 metabolites and 20,741 species; this exceeds the total number of secondary metabolites estimated by De Luca and St Pierre (2000). Using the KNApSack Core DB, we tried to predict the total number of metabolites on Earth based on the scale-free property of the statistics of the relationships between species and metabolites. Many biological systems, including neural networks (Watts and Strogatz 1998), metabolic pathways (Jeong et al. 2000, Bike, 2001) and protein domain networks (Rzhetsky and Gomez, 2001, Wuchty, 2001), appear to conform to the scale-free network model, which can be explained by evolutionary processes in biological systems (Wolf et al. 2002). Thus, the species degree distribution in Fig. 7C could be scale free. In the current release of the KNApSack DB, the scale-free trend is observed in the species degree distribution, but the slope is smaller in the low-degree region on the x-axis compared with that in the high-degree region (Fig. 7C). The slope in the high-degree region should be held in species degree distribution in Fig. 7C. Using the linear relationship in the interval between 10 and 1,000 on the x-axis (number of species), we estimated that there potentially exist 98,800 unique metabolites among

the 20,741 species in the KNApSack Core DB. Thus, on average, 4.7 (=98,800/20,741) unique metabolites are biosynthesized in individual species.

The scientific literature so far recognizes the existence of at least 223,300 plant species (Scotland and Wortley 2003). Thus, the total number of metabolites in plant species can be estimated as 1,060,000 (=98,800/20,741 × 223,300), comparable with the reported range for metabolite number (200,000–1,000,000; Saito and Matsuda 2010).

Fig. 8B shows the relationship between the numbers of plant species and medicinal plants. Three countries (Korea, Pakistan and Japan, denoted by red triangles) have a very small number of plants, but the ratio of medicinal species is relatively higher. The red line in Fig. 8B represents the maximum ratio of the number of medicinal plants to the number of

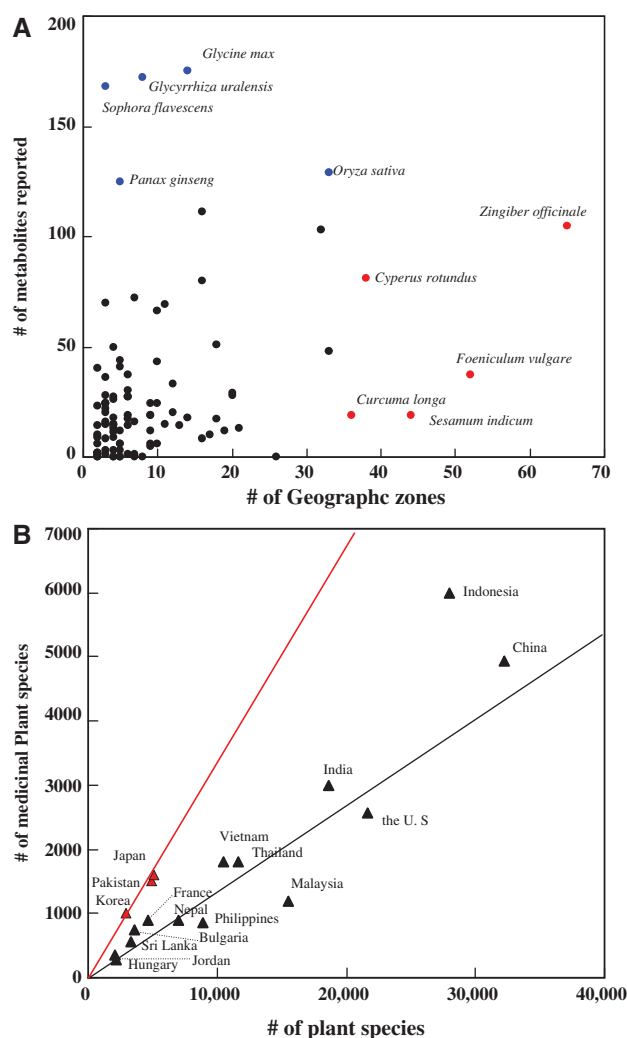


Fig. 8 (A) Relationship between the number of geographical zones and the number of metabolites reported in plants utilized in both Kampo and Jamu medicines. (B) Relationship between the number of wild plants and the number of medicinal plants in different countries. Data are from Iwatsuki et al. (1995), Mitsuhashi et al. (1988) and Schippmann et al. (2002).

wild plants, which is 0.35, with an estimated average ratio of 0.13 (slope represented by a black line). Therefore, assuming 223,300 plant species exist on Earth, the estimated number of medicinal plants ranges from 29,029 ($223,300 \times 0.13$) to 78,155 ($223,300 \times 0.35$), and the estimated number of metabolites expected to have some efficacy in humans ranges from 136,436 ($29,029 \times 4.7$) to 367,328 ($78,155 \times 4.7$). Metabolomics should take into consideration the challenge of isolating this huge number of secondary metabolites and systematizing the effects of natural products on organisms (including humans), using systems-biological approaches. Furthermore, metabolomics should play a role in comprehensively understanding species–species relationships based on natural products, and sustainable society in view of the chemistry between ecology and genetics (Macel et al. 2010). We believe that the DBs reported in this study will make a significant contribution to plant metabolomics.

Supplementary data

Supplementary data are available at PCP online.

Funding

This work was supported by the National Bioscience Database Center in Japan; the Ministry of Education, Culture, Sports, Science and Technology of Japan [Grant-in-Aid for Scientific Research on Innovative Areas ‘Biosynthetic Machinery: Deciphering and Regulating the System for Creating Structural Diversity of Bioactive Metabolites (2007)’].

References

- Adnyana, I.K. and Soemardji, A.A. (2007) Evaluation of Pharmacological Efficacy of Jamu Medicine Institute of Natural Medicine, University of Toyama, Toyama.
- Albert, R., Jeong, H. and Barbási, A.-L. (1999) Internet: diameter of the world-wide web. *Nature* 401: 130–131.
- Bairoch, A. (2000) The ENZYME database in 2000. *Nucleic Acids Res.* 28: 304–304.
- Bike, S. and Peterson, C. (2001) Topological properties of citation and metabolic networks. *Phys. Rev. E* 64: 036106.
- Bino, R.J., Hall, R.D., Fiehn, O., Kopka, J., Saito, K., Draper, J. et al. (2004) Potential of metabolomics as a functional genomics tool. *Trends Plant Sci.* 9: 418–425.
- Beers, S.J. (2001) *Jamu: The Ancient Indonesian Art of Healing*. Periplus Editions Ltd., Vermont.
- Chan, K. (2005) Chinese medicinal materials and their interface with Western medical concepts. *J. Ethnopharmacol.* 96: 1–18.
- De Luca, V. and St Pierre, B. (2000) The cell and developmental biology of alkaloid biosynthesis. *Trends Plant Sci.* 5: 168–173.
- Goto, S., Okuno, Y., Hattori, M., Nishioka, T. and Kanehisa, M. (2002) LIGAND: database of chemical compounds and reactions in biological pathways. *Nucleic Acids Res.* 30: 402–404.
- Guillaume, J.-L. and Latapy, M. (2009) Bipartite graphs as models of complex networks. *Physica A* 371: 795–813.
- Iwatsuki, K., Yamazaki, T., Boufford, D.E. and Ohba, H. (1995) *Flora of Japan*, Vol. 1 Pteridophyta and Gymnospermae, Kodansha Ltd., Tokyo.
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N. and Barabási, A.L. (2000) The large-scale organization of metabolic networks. *Nature* 407: 651–654.
- Ji, H.-F., Li, X.-J. and Zhang, H.-Y. (2009) Natural products and drug discovery. *EMBO Rep.* 10: 194–199.
- Kanehisa, M., Goto, S., Kawashima, S. and Nakaya, A. (2002) The KEGG databases at GenomeNet. *Nucleic Acids Res.* 30: 42–46.
- Macel, M., van Dam, N.M. and Keurentjes, J.B. (2010) Metabolomics: the chemistry between ecology and genetics. *Mol. Ecol. Resources* 10: 583–593.
- Mitsuhashi, H., Okada, M., Nunome, S., Terabayashi, S., Miki, E., Fujita, T. et al. (1988) *Illustrated Medicinal Plants of the World in Colour*. Hokuryukan Co. Ltd., Tokyo.
- Moco, S., Bino, R.J., De Vos, C.H. and Vervoort, J.V. (2007) Metabolomics technologies and metabolite identification. *Trends Anal. Chem.* 26: 855–866.
- Ohkubo, J. (2005) Generation of complex bipartite graphs by using a preferential rewiring process. *Phys. Rev.* 72: 036120.1–10.
- Okada, T., Afendi, F.M., Md., Altaf-UI-Amin, Takahashi, H., Nakamura, K. and Kanaya, S. (2010) Metabolomics of medicinal plants: the importance of multivariate analysis of analytical chemistry data. *Curr. Comput. Aided Drug Des.* 6: 179–196.
- Penso, G. (1980) *WHO Inventory of Medicinal Plants Used in Different Countries*. WHO, Geneva.
- Pramono, S. (2007) *Jamu in Indonesian Daily Life and Industry*. Institution of Natural Medicine, University of Toyama, Toyama.
- Qiu, J. (2007) Traditional medicine: a culture in the balance. *Nature* 448: 126–128.
- Rzhetsky, A. and Gomez, S.M. (2001) Birth of scale-free molecular networks and the number of distinct DNA and protein domains per genome. *Bioinformatics* 17: 988–996.
- Saito, K. and Matsuda, F. (2010) Metabolomics for functional genomics, systems biology, and biotechnology. *Annu. Rev. Plant. Biol.* 61: 463–489.
- Sangat, H.M., Zuhud, E.A.M. and Damayanti, E.K. (2000) *Kamus Penyakit dan Tumbuhan Obat Indonesia [Etnofitomedika]*, Jakarta: Yayasan Obor Indonesia (in Javanese).
- Sawada, Y., Akiyama, K., Sakata, A., Kuwahara, A., Otsuki, H., Sakurai, T. et al. (2009) Widely targeted metabolomics based on large-scale MS/MS data for elucidating metabolite accumulation patterns in plants. *Plant Cell Physiol.* 50: 37–47.
- Schippmann, U., Leaman, D.J. and Cunningham, A.B. (2002) Impact of Cultivation and Gathering of Medicinal Plants on Biodiversity: Global Trends and Issues. Biodiversity and the Ecosystem Approach in Agriculture, Forestry and Fisheries Inter-Departmental Working Group on Biological Diversity for Food and Agriculture, Rome.
- Scotland, R.W. and Wortley, A.H. (2003) How many species of seed plants are there? *Taxon* 52: 101–104.
- Shyur, L.F. and Yang, N.S. (2008) Metabolomics for phytomedicine research and drug development. *Curr. Opin. Chem. Biol.* 12: 66–71.
- Tohge, T. and Fernie, A. (2010) Combining genetic diversity, informatics and metabolomics to facilitate annotation of plant gene function. *Nat. Protoc.* 5: 1210–1227.
- Verpoorte, R., Choi, Y.H. and Kim, H.K. (2005) Ethnopharmacology and systems biology: a perfect holistic match. *J. Ethnopharmacol.* 100: 53–56.

Watanabe, K., Matsuura, K., Gao, P., Hottenbacher, L., Tokunaga, H., Nishimura, K. et al. (2011) Traditional Japanese Kampo medicine: clinical research between modernity and traditional medicine—the state of research and methodological suggestions for the future. *Evid. Based Complement. Alternat. Med.* (in press).

Watts, D.J. and Strogatz, S.H. (1998) Collective dynamics of ‘small-world’ networks. *Nature* 393: 440–442.

Wolf, Y.I., Karev, G. and Koonin, E.V. (2002) Scale-free networks in biology: new insights into the fundamentals of evolution? *BioEssays* 24: 105–109.

Wuchty, S. (2001) Scale-free behavior in protein domain networks. *Mol. Biol. Evol.* 18: 1694–1702.