TBC2health: a database of experimentally validated health-beneficial effects of tea bioactive compounds

Shihua Zhang*, Hongdong Xuan*, Liang Zhang, Sicong Fu, Yijun Wang, Hua Yang, Yuling Tai, Youhong Song, Jinsong Zhang, Chi-Tang Ho, Shaowen Li and Xiaochun Wan

Corresponding authors. Xiaochun Wan, State Key Laboratory of Tea Plant Biology and Utilization, Anhui Agricultural University, Hefei 230036, China. Tel.: +86-551-65786468, Fax.: +86-551-65786183; E-mail: xcwan@ahau.edu.cn; Shaowen Li, College of Information and Computer science, Anhui Agricultural University, Hefei 230036, China. Tel.: +86-551-65786146; Fax.: +86-551-65786183; E-mail: shwli@ahau.edu.cn

*These authors contributed equally to this work.

Abstract

Tea is one of the most consumed beverages in the world. Considerable studies show the exceptional health benefits (e.g. antioxidation, cancer prevention) of tea owing to its various bioactive components. However, data from these extensively published papers had not been made available in a central database. To lay a foundation in improving the understanding of healthy tea functions, we established a TBC2health database that currently documents 1338 relationships between 497 tea bioactive compounds and 206 diseases (or phenotypes) manually culled from over 300 published articles. Each entry in TBC2health contains comprehensive information about a bioactive relationship that can be accessed in three aspects: (i) compound information, (ii) disease (or phenotype) information and (iii) evidence and reference. Using the curated bioactive relationships, a bipartite network was reconstructed and the corresponding network (or sub-network) visualization and topological analyses are provided for users. This database has a user-friendly interface for entry browse, search and download. In addition, TBC2health provides a submission page and several useful tools (e.g. BLAST, molecular docking) to facilitate use of the database. Consequently, TBC2health can serve as a valuable bioinformatics platform for the exploration of beneficial effects of tea on human health. TBC2health is freely available at http://camellia.ahau.edu.cn/TBC2health.

Key words: tea bioactive compound; health benefit; disease; phenotype; network
Introduction

In past decades, epidemiological, clinical and experimental studies have shown that good dietary habits play a critical role in preventing many diseases, such as cancers and cardiovascular diseases [1, 2]. One of the significant characteristics of a health-promoting diet is a high content of a wide range of micronutrients [3]. There is an increasing focus on how to use these natural compounds in the diet for the prevention and treatment of diseases. Of the numerous dietary substances, tea has been around for centuries. The health effects of tea bioactive compounds (TBCs), such as catechins, flavonoids, proanthocyanidins, phenolic acids, alkaloids, terpenoids, fatty acids, amino acids and carbohydrates, are widely documented in scientific literature [4-8].

Tea, brewed from the dried leaves of *Camellia sinensis*, is the most consumed beverage worldwide, after water [9]. Traditionally, tea can be divided into five types: green (non-fermented), white (lightly fermented), oolong (semi-fermented), black (fully fermented) and dark (post-fermented) according to the fermentation process [10]. There are some differences in the distribution and content of beneficial components among these teas. For example, compared with black tea, green and oolong teas have much higher levels of major catechins, such as (-)-epigallocatechin-3-gallate (EGCG), (-)-epicatechin gallate and (-)-epigallocatechin [11]. Among these types of teas, green tea is popular in Asian countries and has been most extensively studied in terms of health benefits [12]. Green tea primarily contains catechins, minor flavanols and polymeric flavonoids [13, 14]. Of these, EGCG is a major catechin and is recognized as a typical TBC in green tea. In past years, many studies report the role of EGCG in lowering the risk of cancer [15], cardiovascular disease [16], diabetes [17] and arthritis [18], and reducing body weight and body fat [19]. Much attention is also directed to black tea, the consumption of which accounts for nearly 80% of the overall tea beverage industry. The health properties of TBCs (e.g. thearubigins and theaflavins) in black tea, including anti-inflammatory [20], antioxidative [21] and anti-tumor activities [22], are also widely described. With an increasing focus on the tea health field, many efforts have been extended to search and explore the health effects of new TBCs from various parts of the tea plant, such as seeds, flowers and roots [23, 24]. For example, Matsuda et al. recently reported the anti-hyperlipidemic and anti-hyperglycemic effects of chasakasaponins I-III (principal saponins) in the flower bud of the tea plant [25]. These attempts enhance our understanding of the health effects of tea as a popular dietary material.

Despite notable progress in discovering different health properties of a large collection of TBCs, the health-promoting mechanisms of tea are still not completely understood. There is a large amount of literature resources that contain health-beneficial information about tea. Collecting and integrating the information in these articles into a cohesive database system gives researchers a complete picture of previous studies. The first database (named TMDB [26]) for tea small molecular compounds was published recently. In it the chemical relevance of each tea compound (such as compound structure and formula) can be systematically viewed. However, the vast majority of tea compounds in TMDB do not have experimentally validated health effects. The compound-oriented database design of TMDB omits the critical compound–disease association.

Considering the above concern, we developed a specialized database (entitled ‘TBC2health’) about the beneficial effects of TBCs on health by manually integrating widely scattered scientific literature. TBC2health allows users to browse, search and download detailed information on relationships between TBCs and diseases (or phenotypes). Several useful applications, such as network visualization and analysis, BLAST, physico-chemical property calculation, and molecular docking, have been combined to strengthen the database. Thus, TBC2health constitutes a unique and valuable repository to enable research of healthy mechanisms of tea.

Materials and methods

Data collection

The purpose of our TBC2health database is to provide a comprehensive information resource about experimentally validated health beneficial effects of TBCs. To this end, we collected detailed information about relationships between TBCs and diseases (or phenotypes) from published research articles. A detailed pipeline for the curation of tea bioactivity data is outlined as follows: (1) the search tool, i.e. SciFinder was used to retrieve tea health-related articles using a list of keywords such as ‘tea health’, ‘tea cancer’ and ‘tea disease’, (2) retrieved articles were thoroughly screened via full text reading to obtain relevant ones that describe the beneficial effects of TBCs on health, (3) after carefully reading these articles, high confident data information related to TBCs, diseases (or phenotypes) and evidence of their relationships was manually extracted, compiled and included into the TBC2health (see ‘Results’ section for details), and (4) it is notable that original research papers were used, not review papers in the tea health field, to avoid data redundancy and confusion.

Nomenclature standardization and classification

Research articles for data curation of TBC2health used varied descriptions for the three biological entities: TBC, disease and phenotype. For example, chemical name, system name and chemical formula for a certain TBC appeared in different publications. For clarity, popular chemical name and aliases of each TBC were manually summarized into 35 unique names (see Browse page for details). All the TBCs in TBC2health were classified by expert curation into 28 chemical groups. Disease classification found to be related to TBCs. The phenotypic effects of TBCs were manually summarized into 35 unique names (see Browse page for details). All the TBCs in TBC2health were classified by expert curation into 28 chemical groups. Disease classification used the method proposed by Goh et al. [27], who classified diseases into 22 classes according to the physiological system affected (see Browse page for details).

Chemical structure drawing

Using original publications, we manually produced the structures of TBCs documented in TBC2health using ISIS Draw (MDL Information Systems, Inc.). Structures of TBCs were further optimized by Sybyl (Tripos, Inc), with sybyl force field and default parameters [28]. All the TBC structures in mol2, SDF and PDB formats were made accessible for users to facilitate several useful applications, such as molecular docking, dynamic simulation and drug design. We also enabled the three-dimensional structure visualization of TBCs using the JavaScript molecular viewer jSmol (http://www.sciencegeek.net).
Database framework and web interface

We designed TBC2health as a relational database on an apache server. All the data were organized in a publicly available MySQL database as the back end, with a user-friendly web interface based on HTML, CSS and JavaScript programming languages as the front end. The architecture of TBC2health is shown in Figure 1.

Results

Data organization and statistics

As a central resource of bioactive relationships between TBCs and diseases (or phenotype), TBC2health offers detailed information about each database entry, such as TBC, disease (or phenotype), evidence of the relationship, CAS number, PubChem Compound Identifier (CID), source of compound (e.g. green, oolong, plant parts), compound type, target gene(s) and pathway(s) of compound, DO identifier, UMLS CUI, ICD-9-CM code, and the corresponding literature reference. In a details page, all these data fields can be viewed in three aspects: (i) compound information, (ii) disease (or phenotype) information, and iii) evidence and reference (see Help page of TBC2health for details).

The current release of TBC2health documents 1,338 relationships between 497 TBCs and 206 diseases (or phenotypes) using manual curation from >300 published articles. As seen in Figure 2A, describing five tea types, the health benefits of green tea are overwhelming. Of the total 497 TBCs, several compound types, such as flavan-3-ols, flavonoids, triterpenoidal saponins, phenolic acids and theaflavins, are prominently involved in these bioactive relationships (Figure 2B). Exploration of disease (phenotype) data indicates the significant health effects of tea, e.g. cancer prevention, metabolism promoting, antioxidation and anti-inflammation (Figure 2C and D).

As functional small molecules, the identification of target genes and pathways of TBCs is a critical step in the exploration of the health promoting mechanisms of tea. A large number of TBCs (54.2%) targeted on one gene, or several genes, and the expression of nearly all the target genes, was down-regulated. We observed that different types of TBCs might target the same gene(s) and trigger similar health promoting effects (Supplementary Table S1), providing valuable clues for the understanding of complicated health function of tea. Among the target pathways, it is noticeable that several disease-related signaling pathways, such as MAPK, ERK and TLR4, were mostly affected by TBCs in an inhibitory fashion (Supplementary Table S2).

Network visualization and topological analysis

With the data archived in TBC2health, a bipartite network can be reconstructed to explore the bioactive relationships between TBCs and diseases (or phenotypes). To achieve this, we developed a network visualization interface using Cytoscape Web [29], where the global bipartite network and direct interactions (i.e. sub-network) of a certain TBC (disease or phenotype) can be viewed (see Help page of TBC2health for details). We computed and provided several topological parameters for a certain TBC, disease or phenotype using the Cytoscape plugin NetworkAnalyzer [30]. Brief descriptions of these network parameters are also provided in the Web site. As indicated in Figure 3E, the direct interactions of EGCG are visualized in a network fashion, revealing multiple health functions of EGCG, such as cancer prevention, antioxidation and anti-inflammation. From a disease (or phenotype)-centered sub-network we can observe that health-promoting benefits related to a certain disease (or phenotype) are usually triggered by synergistic effects of no less than two different TBCs [31, 32] (see the corresponding entry details). For experimental biochemists, the network information reveals underlying biological connections, providing valuable information for further experimental designs.
Database utility

Browse
We developed a powerful browsing facility that allows for the manually curated data to be clearly viewed. Several logical categories, including tea infusions, plant parts, compound types, diseases, phenotypes and in vivo metabolites are listed for users that are also expanded into subcategories.

Search
In search page, three search fields are present that may be used together or separately: (1) compound, (2) disease and (3) phenotype (Figure 3A). Logical operators (AND and OR) are configured between two of these fields to allow rapid, targeted access to specific entries of interest. A user can query the database using standardized keywords, e.g. compound name, CAS number, PubChem CID, disease name and phenotype name, in the corresponding search fields (Figure 3B). TBC2health also offers a fuzzy search function. The fuzzy search function allows users to retrieve entries by the name of a compound, a disease or a phenotype even when the query name is not exactly clear. Once certain query name is received, the system searches in the corresponding data field (i.e. compound, disease and phenotype) for terminology that contains the query words. The matching terminologies are listed as multiple hits in the pop-up result page. From these retrieved results, users can manually check to get the exact one of interest through relevance to the query term (consider EGCG as an example in Figure 3C and D).

Submit and update
A submit page was implemented in the TBC2health to maintain the integrity of the database. We will conduct manual verification of the original publication(s) for data validation upon each submission from non-affiliated researchers. The submitted record(s) that pass this review process will be included in the database, and made available in the incoming release. In the future, we will incorporate the most recent data as soon as it is available to update the TBC2health.

Tools
We integrated several useful application tools (e.g. BLAST, molecular docking and physicochemical property calculation) to enable the database to be easily used in providing a one-stop service for scientific users.

Download and help
As a publicly released scientific database, TBC2health provides a page where the data can be downloaded as a whole or in a customized fashion. We also have a help page that describes the usage of TBC2health in detail via an example presentation and a schematic illustration.

Discussion
The tea plant (Camellia sinensis (L.) O. Kuntze) is well-known to contain catechins, vitamin C, tannins, caffeine and saponins and has been widely cultivated in Asian countries as a source of different types of tea, e.g. green, black and oolong. As a popular nonalcoholic beverage worldwide, tea has also been reported to possess remarkable anti-inflammatory and cancer chemopreventive properties in many animal models, cell culture systems and epidemiological studies [33, 34]. These health-promoting effects of tea are mediated by its various bioactive compounds, such as polyphenols, alkaloids and phenolic acids. For example, green tea polyphenols, the primary sources of green tea’s bioactivity, have been shown to possess potential neuroprotective effects in Parkinson’s disease, Alzheimer’s disease and ischemic stroke along with the confirmed biological activities in antioxidation and inducing cancer apoptosis [35, 36]. Despite these developments, detailed information about health-promoting effects of tea are scattered in published articles. The lack of a
curated database of health-related TBCs limits research in this field, and thus a cohesive database system is necessary for data deposit and further application.

To provide a comprehensive repository about the beneficial effects of TBCs on health, we manually curated the bioactive relationships of TBCs and diseases (or phenotypes) by reviewing published articles and developed the TBC2health database. TBC2health provides a user-friendly interface to accept queries and give detailed information about TBCs, diseases (or phenotypes) and evidence of their relationships. There is also a submit page that allows non-affiliated researchers to contribute novel entries. The incoming version is scheduled to be released on a monthly basis with more valuable resources constantly integrated into the database. To further database use for visiting users, we integrated several useful applications, such as network visualization and analyses, BLAST, physicochemical property calculation, and molecular docking. Consequently, TBC2health can serve as a valuable resource for experimental biochemists and clinical researchers in the tea health research community.

Figure 3. Search interface and example illustration in TBC2health. In search page, a user can conduct keyword-based data query in the compound, disease and phenotype fields separately or cooperatively (A). Three examples that used EGCG, skin cancer and oxidation were shown in this page (B). For EGCG, the direct interactions were visualized in a network fashion by a button clicking in the details page (C–E).
Key Points

- Manually curated bioactive relationships between TBCs and diseases (or phenotypes) are provided.
- Several useful applications, such as network visualization and analysis, BLAST and molecular docking, are integrated into the database for further use.
- A user-friendly interface is provided in this database from which users can browse, search and download information in different ways.
- Several interesting health-promoting mechanisms of TBCs have been discovered through network and statistical analysis.

Supplementary data

Supplementary data are available online at http://bib.oxfordjournals.org/.

Funding

This work was supported by National Key Basic Research Program of China (grant no. 2012CB722903); Anhui Major Demonstration Project for Leading Talent Team on Tea Chemistry and Health; Collaborative Innovation Center of Agri-forestry Forestry in Dabieshan Area; Earmarked Fund for Modern Agro-industry Technology Research System in Tea Industry (grant no. CARS-23); National Natural Science Foundation of China (grant nos. 31301248 and 31201335); China Postdoctoral Science Foundation (grant nos. 2013MS531498 and 2012M511397). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

References