Journal of Ethnopharmacology 208 (2017) 117–128

Contents lists available at ScienceDirect

Journal of Ethnopharmacology

journal homepage: www.elsevier.com/locate/jep

A framework for the targeted selection of herbs with similar efficacy by exploiting drug repositioning technique and curated biomedical knowledge

Sang-Jun Yea[a,b], Bu-Yeo Kim[c], Chul Kim[b,*], Mun Yong Yi[a,*]

[a] Graduate School of Knowledge Service Engineering, Korea Advanced Institute of Science and Technology, Republic of Korea
[b] K-herb Research Center, Korea Institute of Oriental Medicine, Republic of Korea
[c] KM Convergence Research Division, Korea Institute of Oriental Medicine, Republic of Korea

ARTICLE INFO

Keywords:
Systems biology
Traditional Chinese medicine
Medicinal herb
Similar efficacy
Drug repositioning

ABSTRACT

Ethno pharmacological relevance: Plants have been the most important natural resources for traditional medicine and for the modern pharmaceutical industry. They have been in demand in regards to finding alternative medicinal herbs with similar efficacy. Due to the very low probability of discovering useful compounds by random screening, researchers have advocated for using targeted selection approaches. Furthermore, because drug repositioning can speed up the process of drug development, an integrated technique that exploits chemical, genetic, and disease information has been recently developed. Building upon these findings, in this paper, we propose a novel framework for the targeted selection of herbs with similar efficacy by exploiting drug repositioning technique and curated modern scientific biomedical knowledge, with the goal of improving the possibility of inferring the traditional empirical ethno-pharmacological knowledge.

Materials and methods: To rank candidate herbs on the basis of similarities against target herb, we proposed and evaluated a framework that is comprised of the following four layers: *links, extract, similarity*, and *model*. In the framework, multiple databases are linked to build an herb-compound-protein-disease network which was composed of one tripartite network and two bipartite networks allowing comprehensive and detailed information to be extracted. Further, various similarity scores between herbs are calculated, and then prediction models are trained and tested on the basis of theses similarity features.

Results: The proposed framework has been found to be feasible in terms of link loss. Out of the 50 similarities, the best one enhanced the performance of ranking herbs with similar efficacy by about 120–320% compared with our previous study. Also, the prediction model showed improved performance by about 180–480%. While building the prediction model, we identified the compound information as being the most important knowledge source and structural similarity as the most useful measure.

Conclusions: In the proposed framework, we took the knowledge of herbal medicine, chemistry, biology, and medicine into consideration to rank herbs with similar efficacy in candidates. The experimental results demonstrated that the performances of framework outperformed the baselines and identified the important knowledge source and useful similarity measure.

1. Introduction

Plants have been the most important natural resources for treating diseases since ancient civilizations and still remain important for the modern pharmaceutical industry for areas such as new drug development (Sharma and Sarkar, 2013). These can be ascertainable from the well-established systems of traditional medicine in several countries and the fact that one-third of drugs that are currently available come from natural resources with plant origin (Strohl, 2000). Since the Nagoya Protocol has been applied to the traditional knowledge associated with genetic resources that are covered by the Convention on Biological Diversity, it has become more difficult to acquire non-indigenous medicinal herbs. As a result, the demand for alternative medicinal herbs with similar efficacy has increased (Nagoya protocol, 2017). Recently, in vivo and in vitro studies were carried out to determine and compare the anti-inflammatory effects of *Peucedanum praeruptorum* Dunn and *Peucedanum decursivum* (Miq.) Maxim. on allergic lung inflammation (Lee et al., 2016). Whereas the development of new drug has been suffering from high cost, long time, and high risk. Due to the fact that efficient and effective applications of natural

* Corresponding authors.
E-mail addresses: chulnice@kiom.re.kr (C. Kim), munyi@kaist.ac.kr (M.Y. Yi).

http://dx.doi.org/10.1016/j.jep.2017.06.048
Received 28 February 2017; Received in revised form 27 June 2017; Accepted 27 June 2017
Available online 05 July 2017

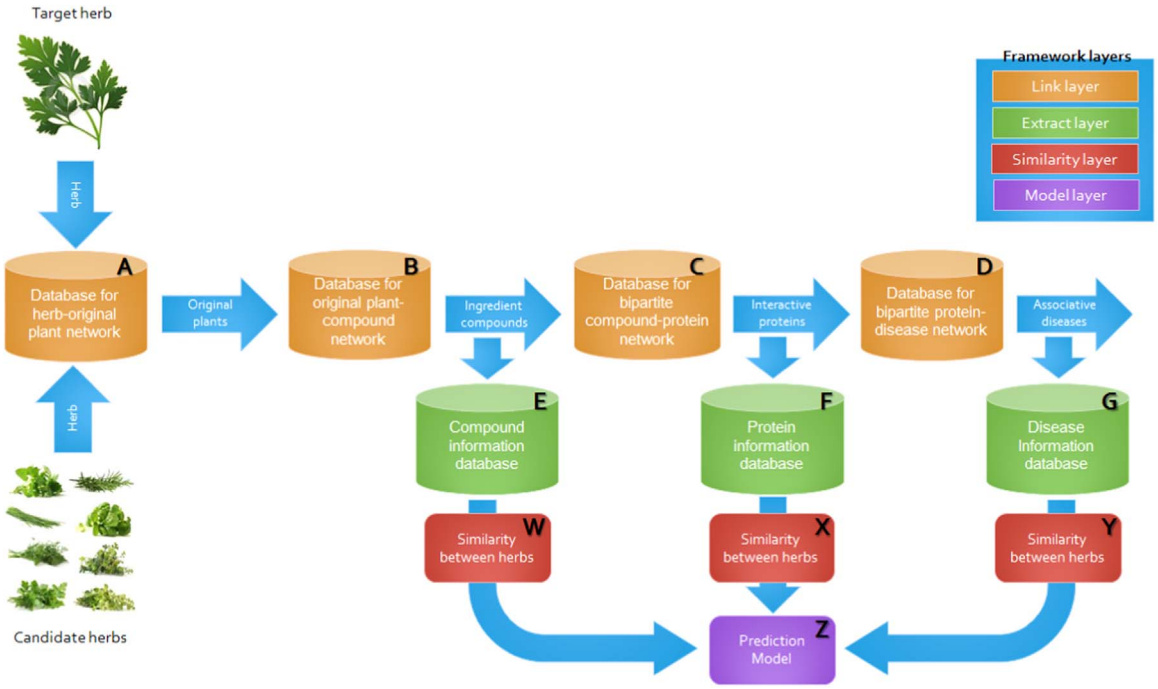

Fig. 1. The overall procedure of THED framework.

products will improve the drug discovery process and reduce the cost of drug development, various screening approaches are being developed in which natural products can be used in the drug discovery process (Harvey, 2008). One of these approaches is to adopt the same family and/or genus of plants or medicinal herbs with similar efficacy based on the assumption that they might have the same or similar bioactive ingredients and biomedical functions (Pan et al., 2013).

There is an approximate 1-in-10,000 probability of discovering useful compounds by random screening (Douwes et al., 2008), not to mention the fact that there is a huge cost involved in both time and expenditure to screen the vast number of randomly selected extracts. Thus researchers have advocated for the use of targeted selection approaches that employ phylogenetic, ecological, or ethno-pharmacological knowledge in the application of natural resources. To adopt a targeted selection approach in the discovery of new drugs, prior researchers have used Bayesian analysis (Weckerle et al., 2011), regression analysis (Douwes et al., 2008), the integration of ethnopharmacology and bioinformatics (Bernard et al., 2001), and the simple scoring system (Clark et al., 1997). To replace medicinal herbs in traditional medicine, targeted selection methods have been proposed, such as the simple mathematical and logical method (Fang et al., 2013), manual review (Zhang et al., 2012), and literature review (Medeiros et al., 2011). However, none of these studies utilized vast accumulated scientific biomedical knowledge to the fullest extent possible or they just used their knowledge of traditional medicine. In our previous study (Yea et al., 2016), a targeted selection technique was developed and evaluated over three validation datasets to rank herbs with similar efficacy by similarity scores calculated on the basis of medical subject headings (MeSH) extracted from articles in MEDLINE. It showed the possibility of inferring traditional empirical ethnopharmacological knowledge using modern scientific biomedical knowledge. However, it also had the following limitations: (1) it only used a non-curated biomedical database, (2) it did not show sufficient performance, and (3) it did not built sophisticated prediction model.

The inefficiencies connected to time and cost in new drug development has brought about the drug repositioning approach, which finds new or additional indication for existing drugs. Since drug repositioning can speed up the process of drug development, it has been in the limelight and as a result, various computational methods have been proposed (Terstappen and Reggiani, 2001; Paul et al., 2010). From amongst these methods, an integrated technique that constructs a comprehensive heterogeneous drug-molecule-disease network at distinct levels and on different scales has been recently developed providing systemic views in predicting new indications (Hurle et al., 2013; Zhang et al., 2014; Wu et al., 2013). These studies measured the similarities between the pertinent drug and disease information and combined it with a nonlinear optimization technique (Zhang et al., 2014), SVM with Kronecker product kernel (Wang et al., 2013a, 2013b), and a logistic regression classifier (Gottlieb et al., 2011), to rank the accumulated evidence for determining the connections between drug and disease. Finding herbs with similar efficacy can be thought of as being analogous to herb repositioning. However, none, as far as we know, have adopted the drug repositioning approach for selecting herbs with similar efficacy. Therefore, in this paper, we propose a novel method for the targeted selection of herbs with similar efficacy by exploiting drug repositioning technique (THED) based on curated biomedical knowledge. The curated biomedical knowledge is publicly available via an online database whose content has been collected by a number of experts via consulting, verifying, and aggregating existing sources (Buneman et al., 2008). We aimed to build a novel computational framework, called THED, for the targeted selection of herbs with similar efficacy; to evaluate the performance of THED; and to identify the relative importance of biomedical knowledge sources selected by prediction model in THED. In order to evaluate the performance of the proposed methods, we adopted the same three validation datasets and evaluation metrics employed by Yea et al. (2016).

2. Materials and methods

2.1. Overall procedure

The purpose of THED framework is to rank candidate herbs on the basis of similarities that are calculated against target herbs (i.e., to tell which candidate herb is more similar to the target herb in terms of efficacy). As depicted in Fig. 1, the THED framework is comprised four layers: *link*, *extract*, *similarity*, and *model*. The link layer is orange and composed of four databases (A, B, C, and D) to build an herb-