

TCM-Suite: 一个全面的中药成分鉴定和网络药理分析平台

杨朋硕,郎继东,李洪军,卢璟祥,林汉阳,田埂,白虹,杨家亮,宁康

华中科技大学 生命科学与技术学院
北京元码基因科技
赛昆思生物科技



Yang, Pengshuo, Lang, Jidong, Li, Hongjun, Lu, Jinxiang, Lin, Hanyang, Tian, Geng, Bai, Hong, Yang, Jialiang, and Ning, Kang. 2022. “TCM-Suite: A comprehensive and holistic platform for Traditional Chinese Medicine component identification and network pharmacology analysis.” *iMeta*. e47. <https://doi.org/10.1002/imt2.47>

Content

- **Introduction**

- Material

- Result



Introduction

Traditional Chinese Medicine:

More than 3,500 years of Chinese medical practice

Holds great potential for treatment of **varied diseases**

Evaluated as a great resource for **drug research**



The quality evaluation of TCM

Composition identification

Active Ingredients Identification

Mix system

Tool:

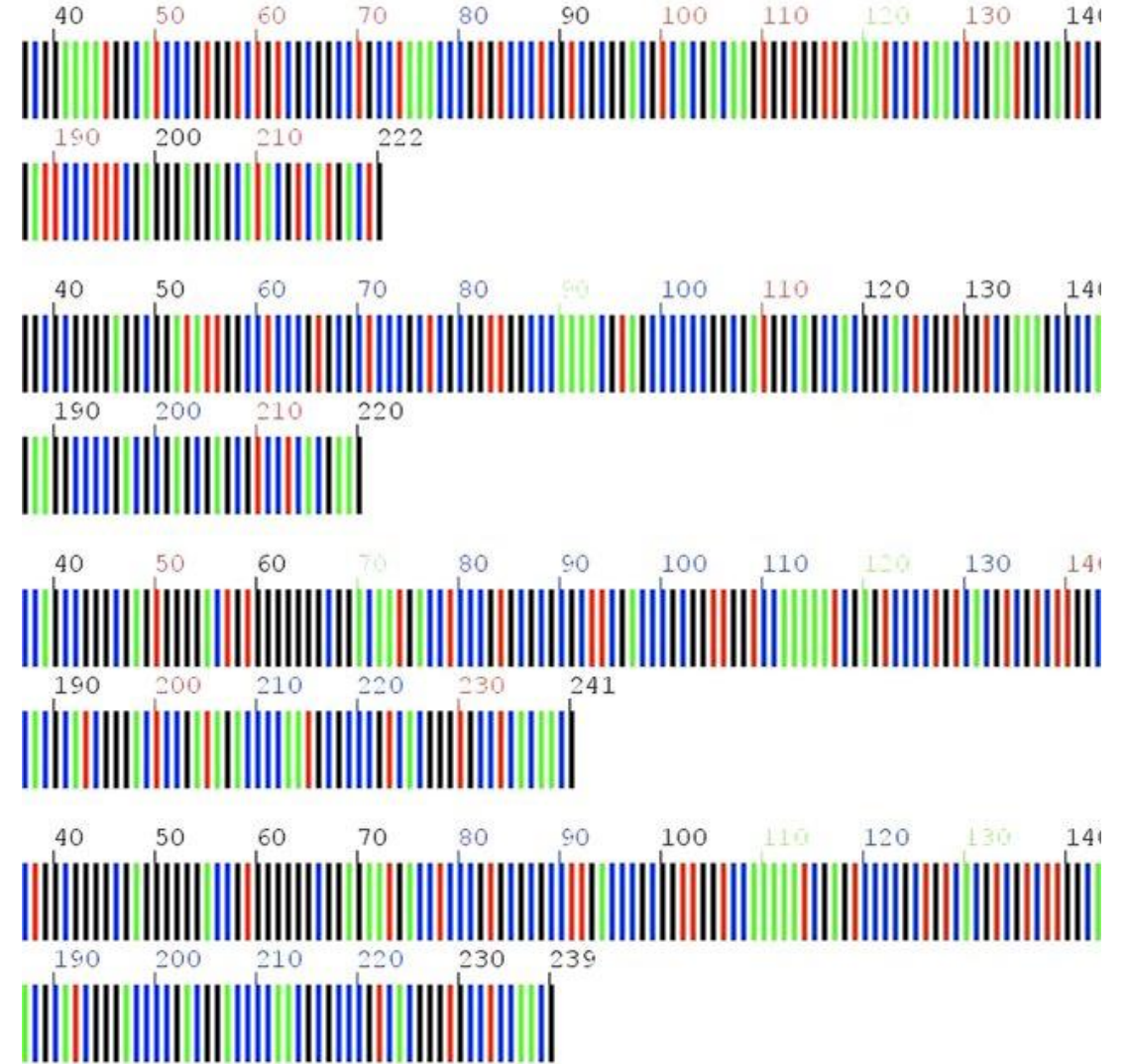
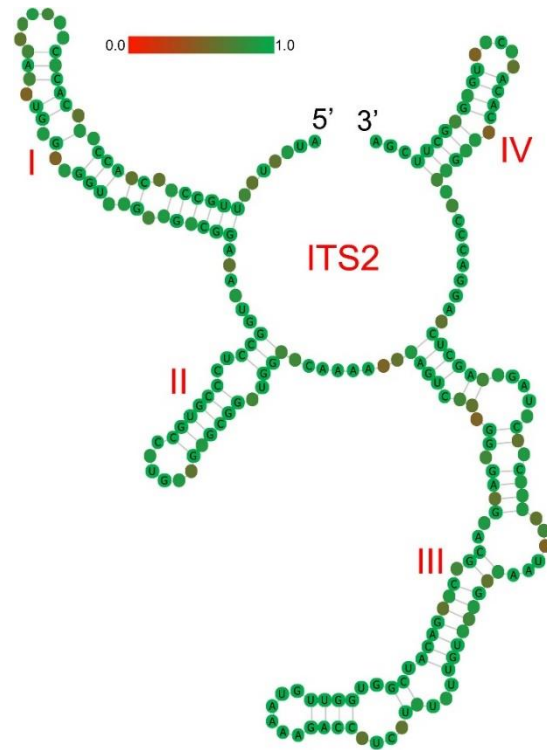
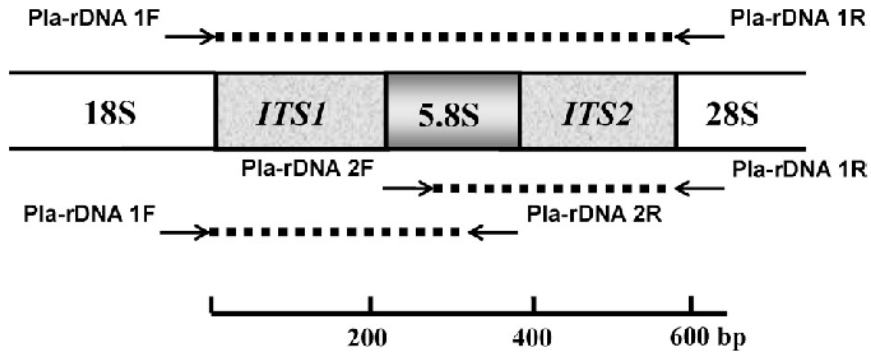
High-throughput sequencing

Bioinformatics



TCM identification: Barcode gene

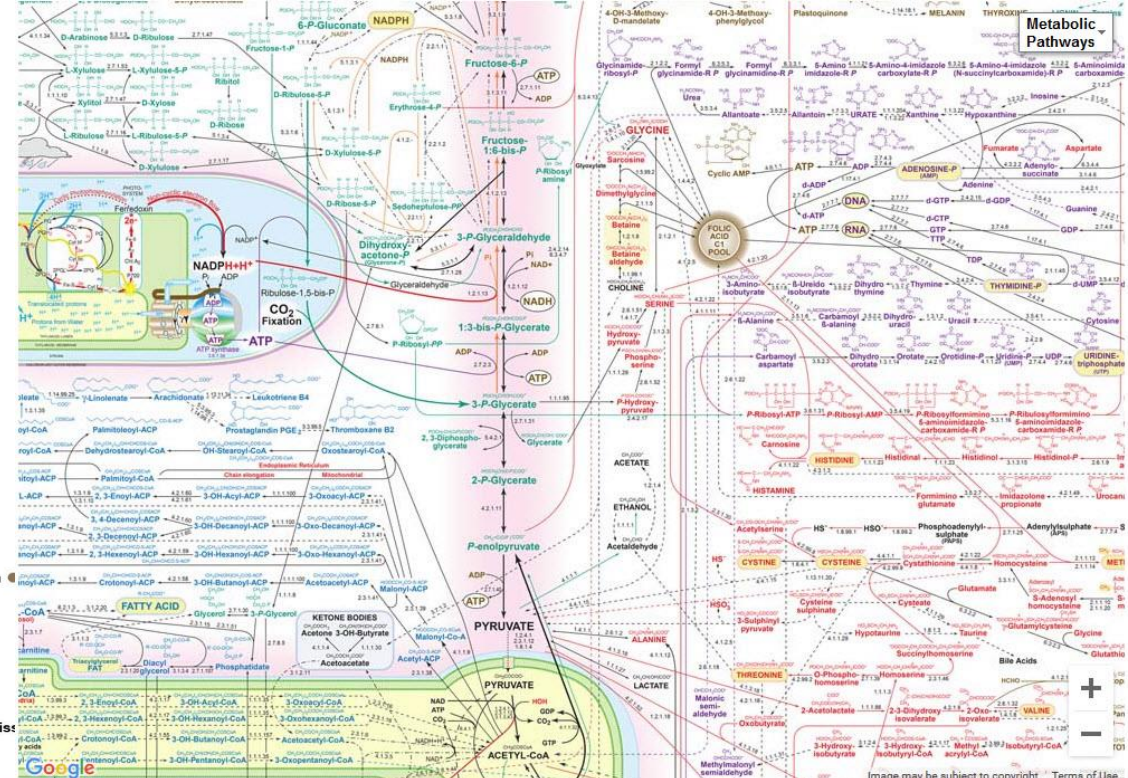
ITS2
matK
trnH-psbA
trnL
rpoc1
yfc1



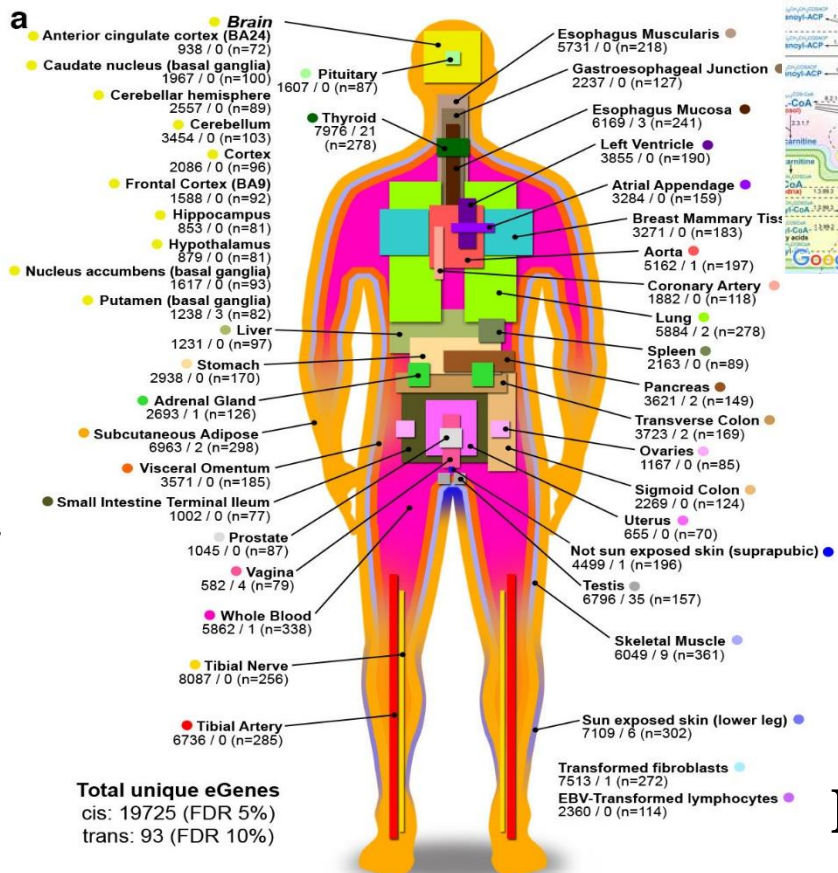
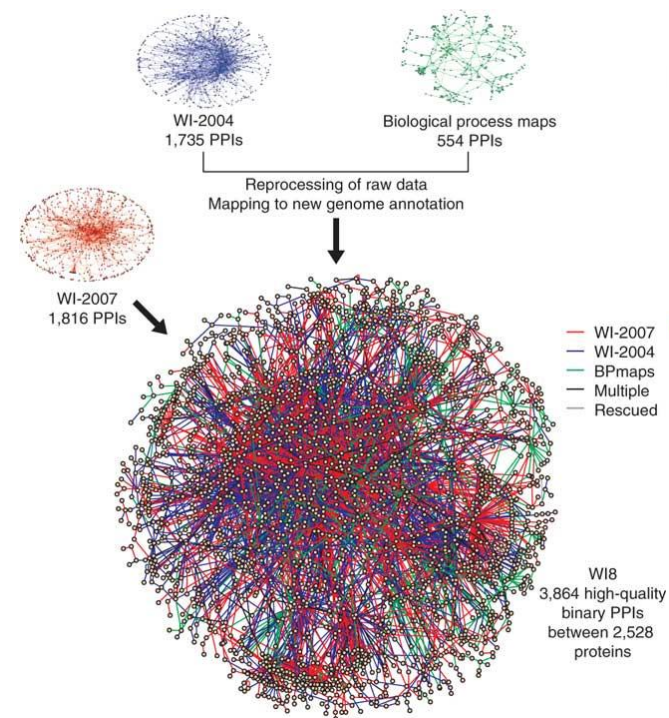
Network Pharmacology

System level: medicine-metabolite-genes-disease

Selects the specific signal node for multi-target drug molecular design.



Human metabolite network




Protein-Protein Interaction network

Human gene and disease

Herb-compound-target protein/gene-disease

TCMID
Traditional Chinese Medicines Integrated Database

Home Search Help Download Contact



Herb name ex: Citrus Reticulata/chenpi/陈皮 Search Advanced

TCMSP : Traditional Chinese Medicine Systems Pharmacology Database and Analysis Platform

TCMSP is a unique systems pharmacology platform of Chinese herbal medicines that captures the relationships between drugs, targets and diseases. The database includes **chemicals**, **targets** and **drug-target networks**, and **associated drug-target-disease networks**, as well as pharmacokinetic properties for natural compounds involving **oral bioavailability**, **drug-likeness**, **intestinal epithelial permeability**, **blood-brain-barrier**, **aqueous solubility** and etc. This breakthrough has sparked a new interest in the search of candidate drugs in various types of traditional Chinese herbs.

The authors are grateful to the [TTD](#) (Therapeutic Target Database), [PharmGKB](#) and [PubChem](#) database. The network viewer in our website is based on CytoscapeWeb, an interactive web-based network browser. It is freely available at <http://cytoscapeweb.cytoscape.org/>.

Welcome to **STITCH**

Chemical-Protein Interaction Networks

ORGANISMS	CHEMICALS	PROTEINS	INTERACTIONS
2031	0.5 mio	9.6 mio	1.6 bn

SEARCH

Welcome to **STRING**

Protein-Protein Interaction Networks
Functional Enrichment Analysis

ORGANISMS	PROTEINS	INTERACTIONS
5090	24.6 mio	>2000 mio

SEARCH



Genetic Association Database

The Genetic Association Database is a database of genetic association data from complex diseases and disorders. After serving the scientific community for more than 10 years, the Genetic Association Database (GAD) has been retired and all data is "frozen" as of 09/01/2014. However, all GAD data as of 08/18/2014 will continue to be available via the following download links.

Text Format: [data.zip](#)
SQL Format: [sql.zip](#)

The limitation of current TCM research

Lack comprehensive TCM biological composition identification

Redundancy in multiple databases

Holistic pipeline is still lacking for TCM analysis





TCM-Suite

A comprehensive and holistic platform for
TCM composition identification(Holmes-Suite) and network pharmacology
analysis(Watson-Suite)

Formula

Search Formula by English Name, PinYin Name, or Chinese Name ...

Example Formula: [Hu Gan Pills](#) 护肝丸

[BROWSE PHARMACOLOGY DATABASE](#)

[SEARCH TCM MARKER GENE](#)



Content

- Introduction

- **Material**



- Result

Material



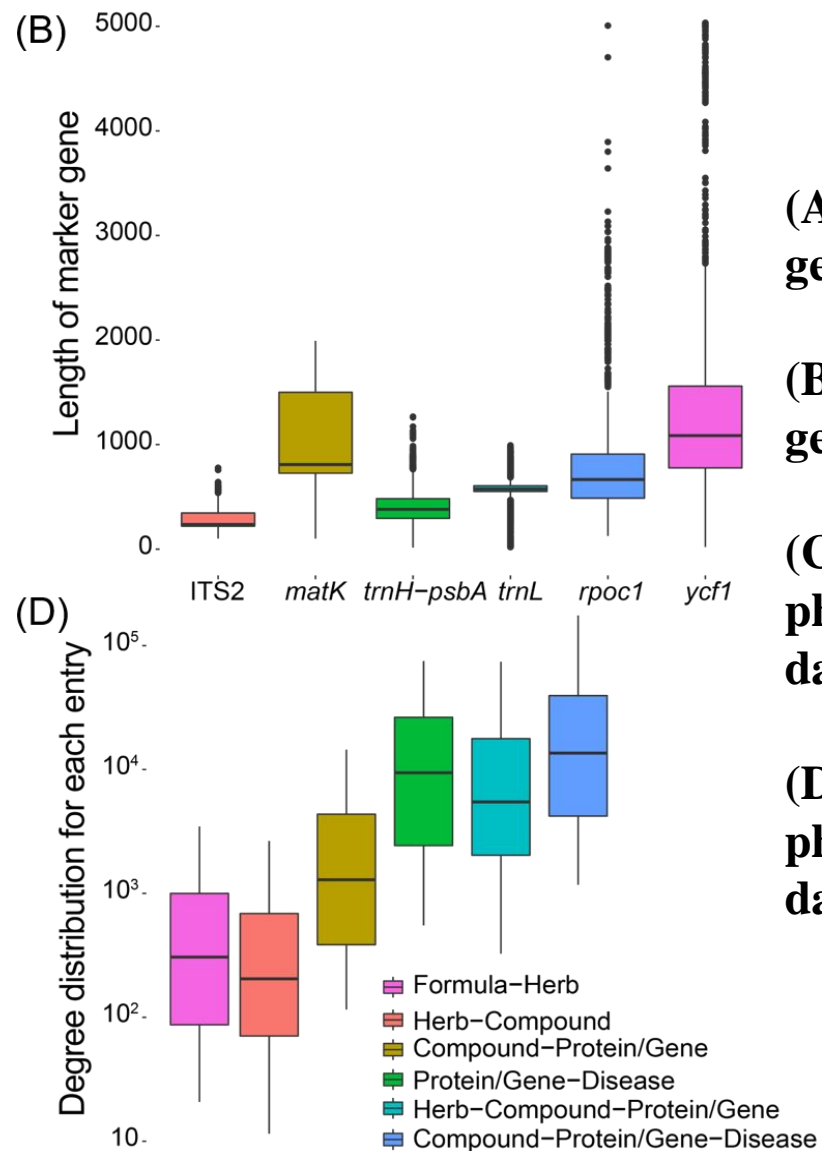
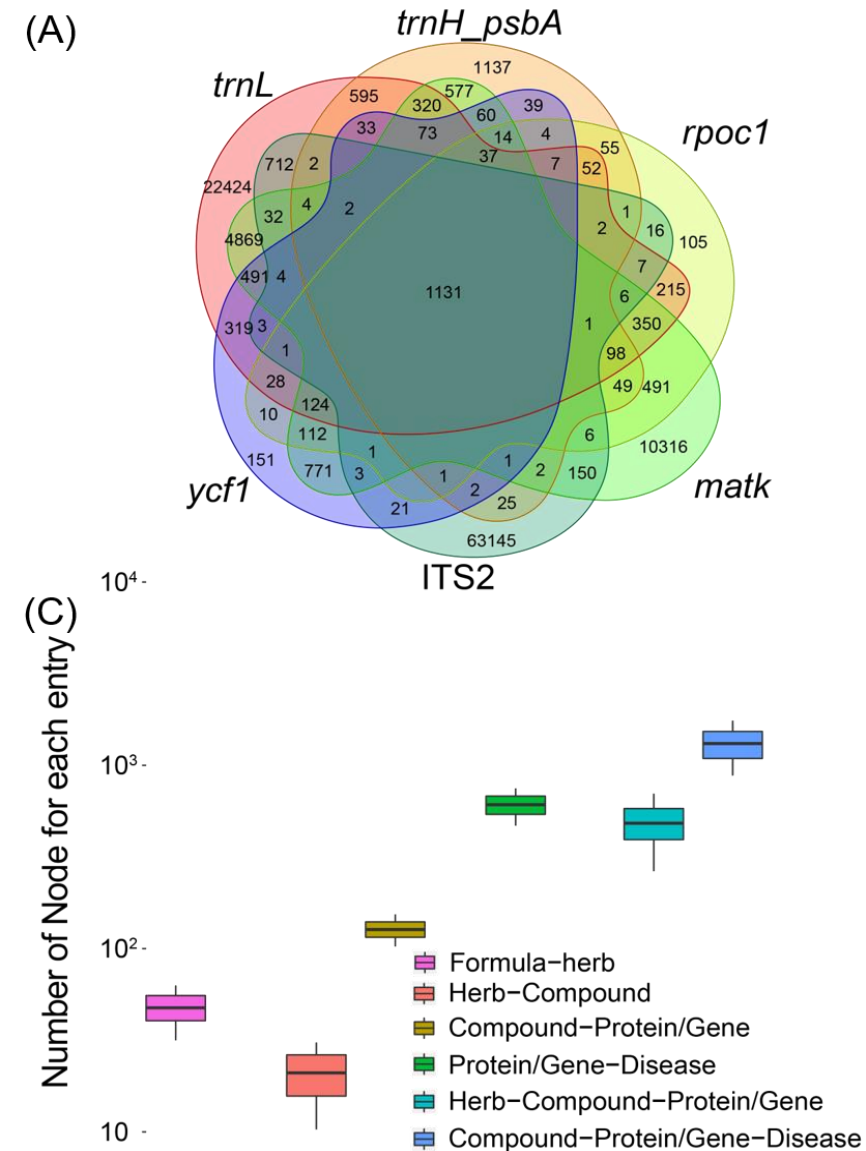
Data entity	Data source	# of entries	Description	Websites
Barcode genes(Holmes-Suite)	NCBI	157,937	marker gene ITS2	https://www.ncbi.nlm.nih.gov/nucleotide/
	NCBI	24,267	marker gene matK	https://www.ncbi.nlm.nih.gov/nucleotide/
	NCBI	5,418	marker gene trnH-psbA	https://www.ncbi.nlm.nih.gov/nucleotide/
	NCBI	39,643	marker gene trnL	https://www.ncbi.nlm.nih.gov/nucleotide/
	NCBI	4,086	marker gene rpoc1	https://www.ncbi.nlm.nih.gov/nucleotide/
	NCBI	4,119	marker gene ycf1	https://www.ncbi.nlm.nih.gov/nucleotide/
Formula(Watson-Suite)	TCM-ID	6,692	Formula and corresponding Herb information	http://bidd.group/TCMID/
Herb-ingredients (Watson-Suite)	TCMID	57,249	Herbs and corresponding ingredients	http://www.megabionet.org/tcmid/
	TCMSP	29,384	Herbs and corresponding ingredients	http://tcmspw.com/tcmsp.php
Compounds (Watson-Suite)	STITCH	82,841,024	Compound-CID number links	http://stitch.embl.de/
	TOXNET	103	Compound's toxicity	https://toxnet.nlm.nih.gov/
	TCMSP	12,144	Compound	http://tcmspw.com/tcmsp.php
	SIDER	163,221	Compound's side effect	http://sideeffects.embl.de/
Proteins (Watson-Suite)	STITCH	4,523,609	CID number-protein links	http://stitch.embl.de/
	STRING	4,274,001	Protein interactions	http://www.string-db.org/
	OMIM	2,449,433	Proteins and corresponding aliases	http://omim.org/
	OMIM	15,591	Approved gene and MIM number	http://omim.org/
Diseases (Watson-Suite)	OMIM	7,086	MIM number-disease links	http://omim.org/
	TCMSP	837	Gene-disease associations	http://tcmspw.com/tcmsp.php
	GAD	167,130	Gene-disease associations	https://geneticassociationdb.nih.gov/

Content



- Introduction
- Material
- **Result**

Data resources and database construction



(A) The species distribution for six marker genes.

(B) The length distribution for six marker gene database.

(C) Node distribution for six kinds of pharmacology networks in Watson database.

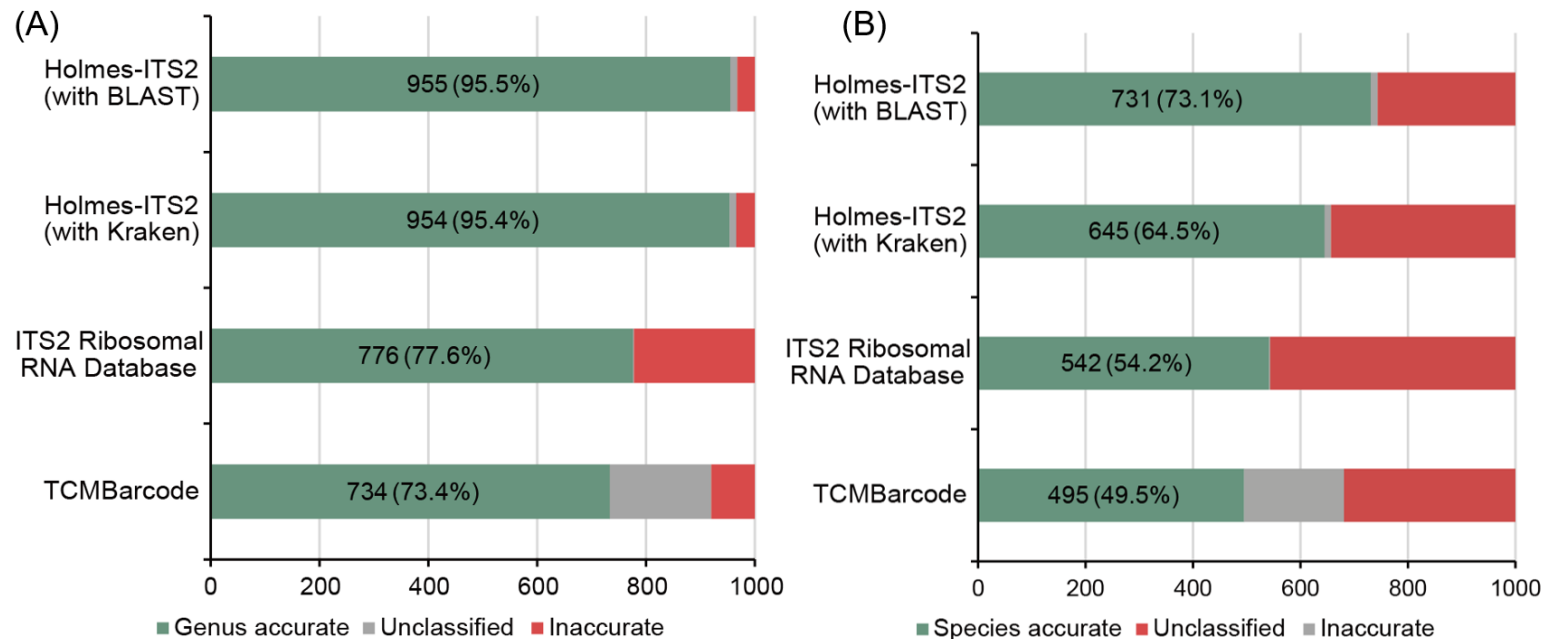
(D) Degree distribution for six kinds of pharmacology networks in Watson database.



Accuracy and efficiency analyses for TCM biological composition identification of Holmes-Suite

Table 2. Databases used for comparison of biological composition of TCM.

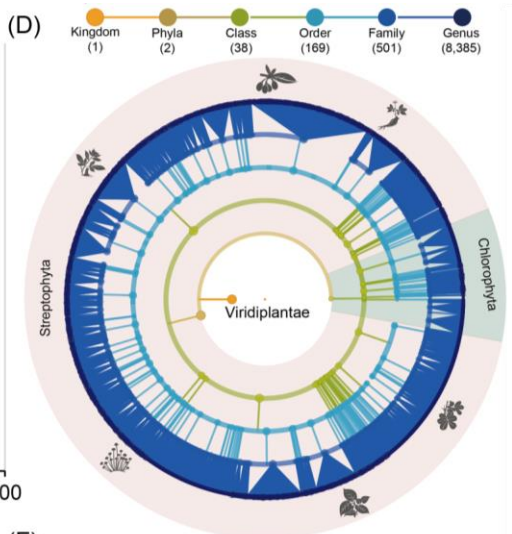
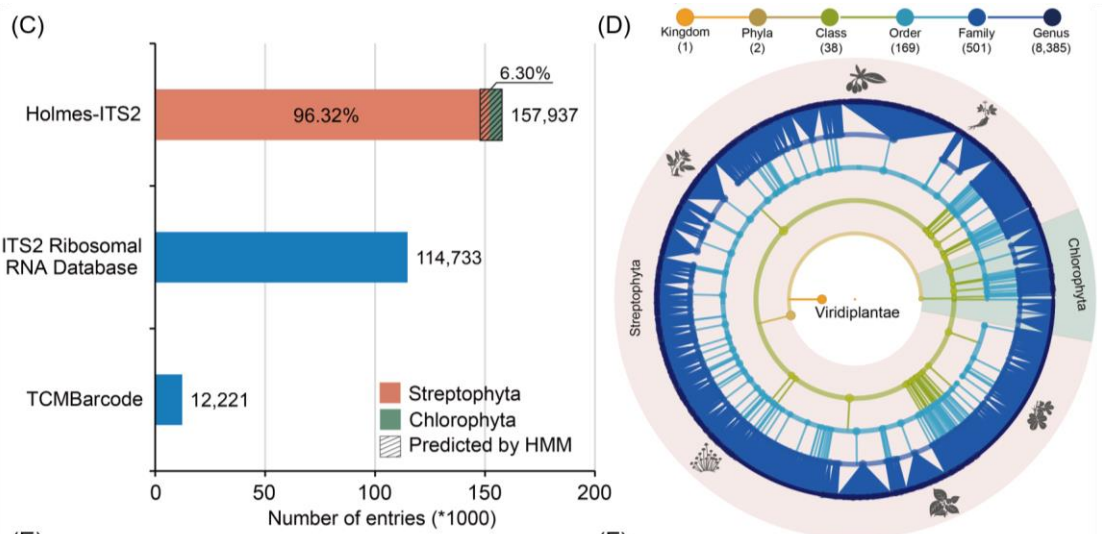
Existing Database	Number of sequences	Latest update time	Website
ITS2 Ribosomal RNA Database	113,364	2015-08-07	http://its2.bioapps.biozentrum.uni-wuerzburg.de/
TCMBarcode Database	452,517	2018-05-06	http://www.tcmbarcode.cn/en/
Holmes-Suite Database	1,251,548	2021-07-27	http://TCM-Suite.AImicrobiome.cn



(A) Genus-level sensitivity and (B) species-level sensitivity were shown for three databases.

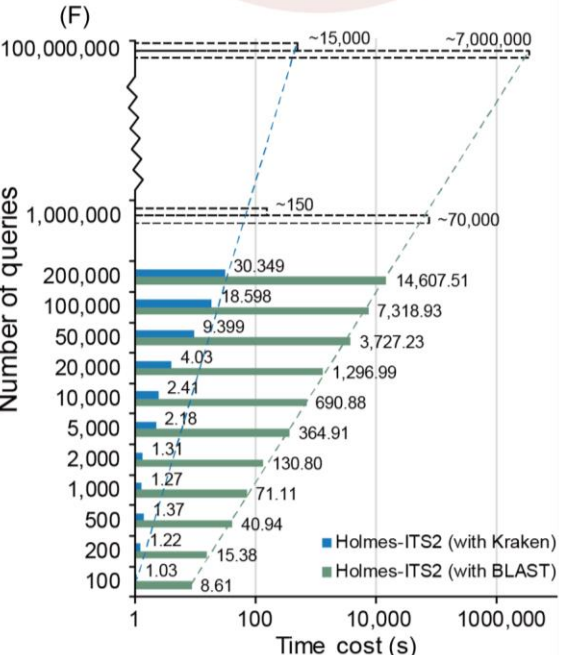
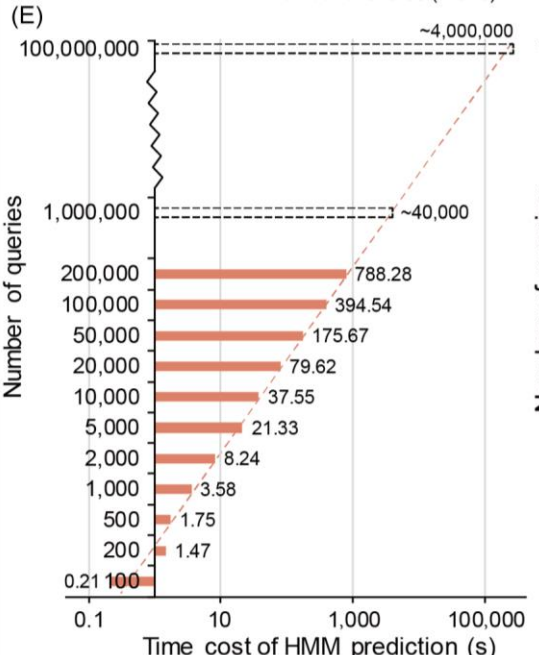


Accuracy and efficiency analyses for TCM biological composition identification of Holmes-Suite



(C) The number of entries (DNA-based marker sequences) contained in each database (blue) and data component of Holmes-ITS2 (ITS2 part of the Holmes-Suite database).

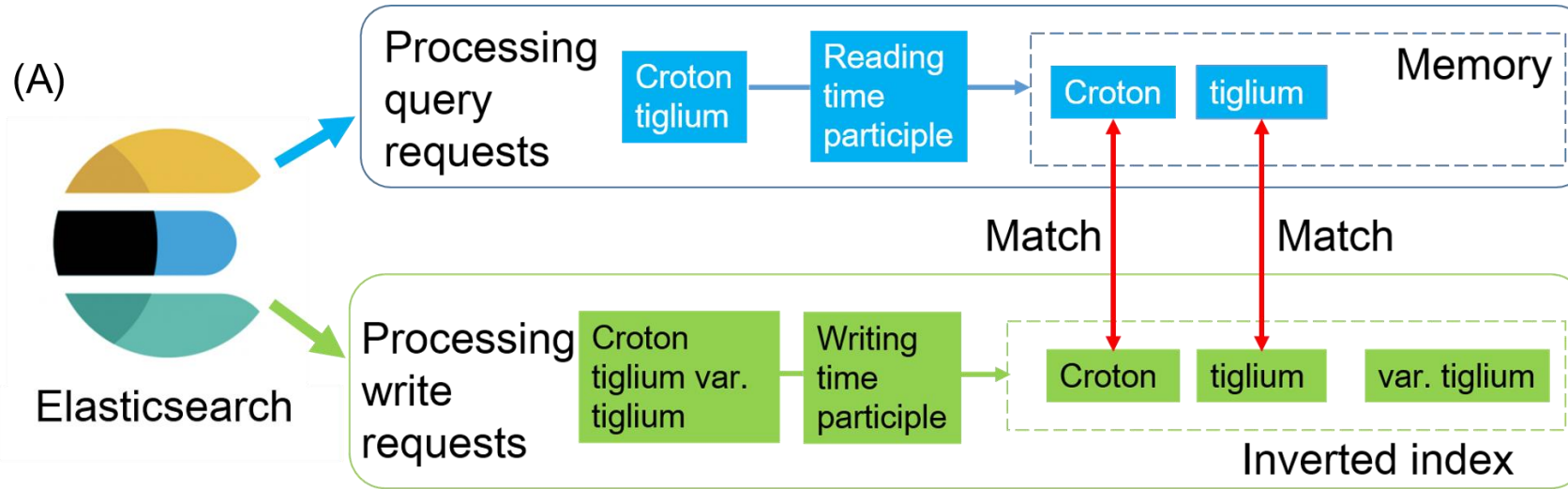
(D) From inside to outside of the taxonomy hierarchical tree, each node representing a taxon with a different level of taxonomy such as kingdom, phylum, class, etc.



(E) HMM prediction for training the ITS2 sequence edge model and (F) identification with BLAST and Kraken search engines.



Database features and web interface



The interconnection of Holmes-Suite and Watson-Suite databases.

(A) Fuzzy matching of biological composition for interconnection of Holmes-Suite and Watson-Suite.

(B) Summary of Species (Holmes database)

Species	Croton tiglium	Domain	Eukaryota
Kingdom	Viridiplantae	Phylum	Streptophyta
Class	Gunneridae	Order	Malpighiales
Family	Euphorbiaceae	Genus	Croton

(C) Summary of Herb (Watson database)

PinYin Name	BA DOU
Latin Name	Croton tiglium
Merdians	Large Intestine, Stomach

(B) The detailed species information page for Holmes-Suite database.

(C) The detailed herbal information page for Watson-Suite database.

Species properties

Link to Watson

Croton tiglium	Croton oblongifolius [Syn. Croton laevigatus]	Croton macrostachys
Croton flavens	Croton draconoide	Croton tonkinensis
Croton schiedeianus	Croton sublyratus	Croton cumingii
Croton cajucara	Croton eluteria	Croton lechleri

Records in Watson database

Marker Gene

Database	Marked	Sequence
ITS2	⊕	CGCTCCCAACGCATTGGGGTGCAGGAATATGGCTCCCGTGCATCTCTCCCGGTTGGCCCTAAAGAATGGTCTCGGCTGCAAATGCCGGATAATCGGTGGTTCAAGGCCCTCGGACACAGTCGCGCGCCAGCATGCTTTGGATAACGAGACCCCTCTGCGTCCGCTCTTCGCGCAAGCTCACATCGCGACCCAGGTCAAGGGATTACCCGCTGAGTTTAAGCATATCA

Marker gene sequences

Link to Holmes

Croton tiglium (its2)
Croton schultzei (its2)
chloroplast Croton roxanae (matk)
chloroplast Croton poecilanthus (matk)

Show More ...

Linked to Holmes database



Database features and web interface

(A) **TCMSuite** **Module interface** [Home](#) [Holmes](#) [Watson](#) [Help](#) [Contact](#) [Download](#)

TCM-Suite
A comprehensive and holistic platform for TCM composition identification (Holmes-Suite) and network pharmacology analysis (Watson-Suite)

Formula Search Formula by English Name, PinYin Name, or Chinese Name ...

Example Formula: **Hu Gan Pills 护肝丸**

[BROWSE PHARMACOLOGY DATABASE](#) [SEARCH TCM MARKER GENE](#)

(C) **Marker gene search**

Database: ITS2 Domain: Domain Kingdom: Kingdom
Phylum: Phylum Class: Class Order: Order
Family: Family Genus: Genus Species: Species

[Intro](#) [Search](#) [Reset](#)

Domain	Kingdom	Phylum	Class	Order	Family	Genus	Species	Detail
Eukaryota	Viridiplantae	Streptophyta	Gunneridae	Asterales	Asteraceae	Picnomon	Picnomon acarna	Detail
Eukaryota	Viridiplantae	Streptophyta	Gunneridae	Asterales	Asteraceae	Warionia	Warionia saharae	Detail

(D) **Pharmacological pathway selection**

searchType: Formula Search: [Intro](#) [Search](#) [Reset](#) [Advanced](#)

English Name	Chinese Name	PinYin Name	Detail
San Niu Tablets	三拗片	San Niu Pian	Detail
Dan Xiang Qing Zhi Granules	丹香清脂颗粒	Dan Xiang Qing Zhi Ke Li	Detail
Dan Yi Tablets	丹益片	Dan Yi Pian	Detail
Yi Gan Yang Yin Huo Xue Granules	乙肝养阴活血颗粒	Yi Gan Yang Yin Huo Xue Ke Li	Detail
Jiu Xiang Zhi Tong Pills	九香止痛丸	Jiu Xiang Zhi Tong Wan	Detail

(B) **Marker gene database** **Holmes Suite**

ITS2:ATCGATA..... Matk:ACGTATA..... trnH-psbA:TCGAATA..... trnL:ATCGATC..... rpoC1:CGATATA..... YCF1:TAATCTA.....

Species → DNA → Barcode(Marker gene)

Formulas → **Herbs** → **Compounds** → **Protein/genes** → **Disease**

Radix Ginseng → Radix Ginseng

Nicotinic acid hydrazide, Acetylenedicarboxylic acid, NSC69863, Ginkgolic acid, Ginsenoside Re → HMOX1, PPP1R3A, AKT1, HEBP1, NFE2L2, RB1 → Atherosclerosis, Bone Mineral Density, Gastritis, Atherosclerosis, Kidney Failure, Chronic Cardiovascular Diseases

Watson Suite

(E) **Summary of Herb**

PinYin Name	DANG GUI	English Name	Chinese Angelica Equivalent plant: Phlojodicarpus
Latin Name	Angelica sinensis	Properties	Warm,Pungent,Sweet
Meridians	Spleen,Liver,Heart	Use Part	root

Effect: To nourish blood and regulate menstruation, quicken blood, relieve pain, moisten intestines and relieve constipation.

Compound-Protein Combined Score: 0 150 1000

(F) **Connection for two databases**

chloroplast **Angelica sinensis** (matk)

Link to chloroplast **Angelica archangelica** (rpoC1)

Holmes **Angelica grayi** (its2)

Angelica amurensis (its2)

[Show More ...](#)

(E) **Summary of Species**

Species chloroplast **Angelica sinensis**

Angelica sinensis

Angelica pubescens

Link to Watson **Camellia sinensis** [Sy]

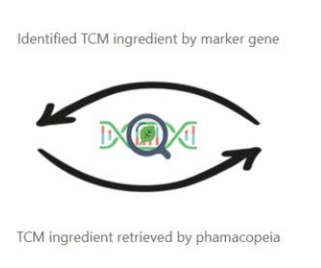
Spiranthes sinensis

[Show More ...](#)

Holmes-Suite

A standardized short sequence of DNA from a small region of an organism's genome that can differentiate the species from others.

For plants, the chosen and validated sequences include ITS2, matK, trnH-psbA etc. for the identification of plant species.



Watson-Suite

Watson-Suite is designed as an integration of database and a datamining system for network pharmacology analysis of TCM preparations. We have collected information on all respects of TCM including herbs, herbal ingredients, targets, related disease, side effect and toxicity



Case study: the investigation of COVID-19

Symptom	# Gene target	# Compound	# Biological ingredient
SARS-coronavirus	3	54	156
Pneumonia	10	67	138
Typhoid fever	8	60	215
Cough	5	46	215
Dyspnea	1	13	100
Diarrhea	12	25	125

Gene	Compounds	biological ingredients
ACE2	12	58
SLC6A20	23	128
SIT1	1	2
CCR9	2	8
CXCR6	14	54

A total of 15 drug targets, 52 compounds and 86 biological ingredients.

86 biological ingredients were identified as important components of clinically effective TCM to cure COVID-19.

Lonicera japonica, Scutellaria baicalensis, and Forsythia suspensa are the active ingredients of **Shuanghuanglian**; Forsythia suspensa, Ephedra sinica, Lonicera japonica, Isatis indigotica, Mentha haplocalyx, Dryopteris crassirhizoma, are the main component of **Lianhua Qingwen capsule**.



Conclusion

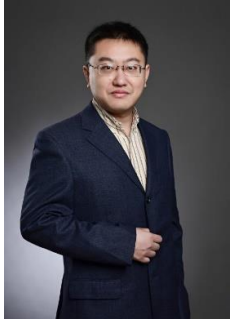
- For TCM biological ingredient identification, we developed a highly accurate, comprehensive, and efficient identification system Holmes-Suite.
- For network pharmacology analysis, Watson-Suite enables deeper mining of the pipeline for “herb-compound-protein-disease” interpretation
- TCM-Suite achieved a smooth interconnection between Holmes-Suite and Watson-Suite, illustrating its great potential for TCM-based drug discovery and repurposing.



致谢



Group Members



K Ning



H Bai



YM Zhao



Q Yao



PS Yang



MY Cheng



YG Zha



X Zhu



N Wang



L Ji



GZ Xiong



YX Li



D Zhao

Our Team: Ning lab





“iMeta”是由威立、肠菌分会和本领域数百位华人科学家合作出版的开放获取期刊，主编由中科院微生物所刘双江研究员和荷兰格罗宁根大学傅静远教授担任。目的是发表原创研究、方法和综述以促进宏基因组学、微生物组和生物信息学发展。目标是发表前10%(IF > 15)的高影响力论文。期刊特色包括视频投稿、可重复分析、图片打磨、青年编委、前3年免出版费、50万用户的社交媒体宣传等。2022年2月正式创刊发行!



主页: <http://www.imeta.science>

出版社: <https://onlinelibrary.wiley.com/journal/2770596x>



投稿: <https://mc.manuscriptcentral.com/imeta>



office@imeta.science



[iMeta](#)

