|| ISSN(online): 2589-8698 || ISSN(print): 2589-868X ||
International Journal of Medical and Biomedical Studies
Available Online at www.ijmbs.info
NLM (National Library of Medicine ID: 101738825)
Index Copernicus Value 2019: 79.34
Volume 5, Issue 2; February: 2021; Page No. 18-24

Original Research Article

SYNTHETIC VIEWS TO TRADITIONAL HERBAL MEDICINE DERIVED FROM TEXT MINING OF BIOMEDICAL BIBLIOGRAPHIES.

Sergey Feranchuk

Self-employed

Article Info: Received 08 January 2021; Accepted 29 January 2021
DOI: https://doi.org/10.32553/ijmbs.v5i2.1670
Corresponding author: Sergey Feranchuk
Conflict of interest: No conflict of interest.

Abstract

Text mining approach was applied to a subset of MEDLINE database in a conjunction with several chemoinformatics and bioinformatics databases. The connections between biomedical herbs, active compounds, traditional cultures and most important human diseases were collected. The integrated statistical properties of the obtained connections were estimated. The consistency of the obtained distributions with a conventional approaches of molecular biology and medicine was approved.

The generic statistical properties of the obtained connections allow to clarify and specify the generic relationships between herbal cultures and between diseases of humans. The averaged content of chemical compounds was in agreement with abundances of specific enzymes in plant genomes. The role of traditional Chinese medicine as most stable and conservative herbal culture was approved.

The clear separation of mental disorders from oncology and cardiovascular diseases was observed. A uniform activity of plants associated with a treatment of oncology diseases appears to be in contrast to a diverse activity of plants used to correct brain disorders. The obtained results could be meaningful to close the gap between traditional and modern approaches of medicine.

Introduction

All cultures, ancient or modern, have the traditions for the use of plants for as a food and for treating diseases. As a food, plants could be empirically classified as toxic and eatable. In herbal medicine, some of toxic plants are used and the choice of doze is important there.

And all cultures include some tradition for use of psychoactive plants, for recreation and in rituals [Alrashedy and Molina, 2016]. With the reservations about the complexity of mental processes, the action of these plants could be classified as sedative or stimulating. The most harmful adverse effect of psychoactive plants is a drug addiction, which is also developed depending on a doze and a way of using. The effect of addiction there is only indirectly related to the mode of action.

The modern culture of herbal medicine is enriched by standardized protocols for an assessment of medicinal plants, and by ability to identify chemical compounds which are responsible for a therapeutic effect of a plant. The both advantages allow connecting the experience of herbal medicine with a knowledge accumulated in natural sciences, mostly in molecular biology.

Among the recent studies, in the [Cazander at al., 2012] the plants used for wound healing are listed, and the mechanisms of action are separated to an involvement of classic and alternative pathways for an inhibition of inflammation. And in [Guzman and Molina, 2018] the plants used for a treating of cardiovascular diseases are listed, for which a classification to several specific mechanisms of action is proposed; but for some of plants mechanism of action is anyway unknown.

Text mining of biomedical literature was also used in another ways, to provide some synthesis and a generic views to properties of medicinal herbs [Xue et al., 2012; Choi et al., 2016;]. And, in these frames, the aim of the described study is to present a kind of binary classification for medicinal plants, together with an adjacent classification of human diseases.

The precision of the expected results would be obviously limited, and the provided classification is not supported by terms from molecular biology, which would specify a mechanism of action.

The extracts from medicinal plants are composed from many active compounds, and the effect of poly-specificity makes it often difficult to specify the exact biological targets for an active compound. This and another reasons makes it difficult to provide even the robust level of statistical significance for the presented results.

But as a kind of verification for the presented distribution of medicinal plants, the consistency of the approach is demonstrated on the two another aspects. The results are obtained as a reduction of the database with relations between plants and chemical compounds. The same database could be reduced by several other ways. So, the first kind of verification is based on a comparison between estimated classes of chemical compounds, and an abundance of specific enzymes in genomes, depending on a taxonomic clade of medicinal plants. This kind of analysis was performed using databanks of ontologies from ChEBI project [Degtyarenko et al., 2008] and GO project [Ashburner et al., 2000] togewer with genome annotations from PlantTFDB site [Jin et al., 2017] site and Ensembl [Herrero et al., 2016] site.

Second kind of verification based on an estimated relations between traditions of herbal medicine. The consistency of the obtained relations could be supported by and intuitive meaning, and it is compared with a relations between psychoactive cultures, estimated from the [Alrashedy and Molina., 2016].

Material and methods

A presented result are based on a sequential compilation of three databases, as it is shown on fig. 1, and several types of statistical analysis of the compiled databases and auxiliary data sources. A database of relations between medicinal plants and chemical compounds is central in a presented pipeline. It was than used to estimate associations between medicinal plants, terms related to human diseases and terms related to region of ethnic tradition. A straightforward text search in bibliographies from Medline database was used to compile these two databases.

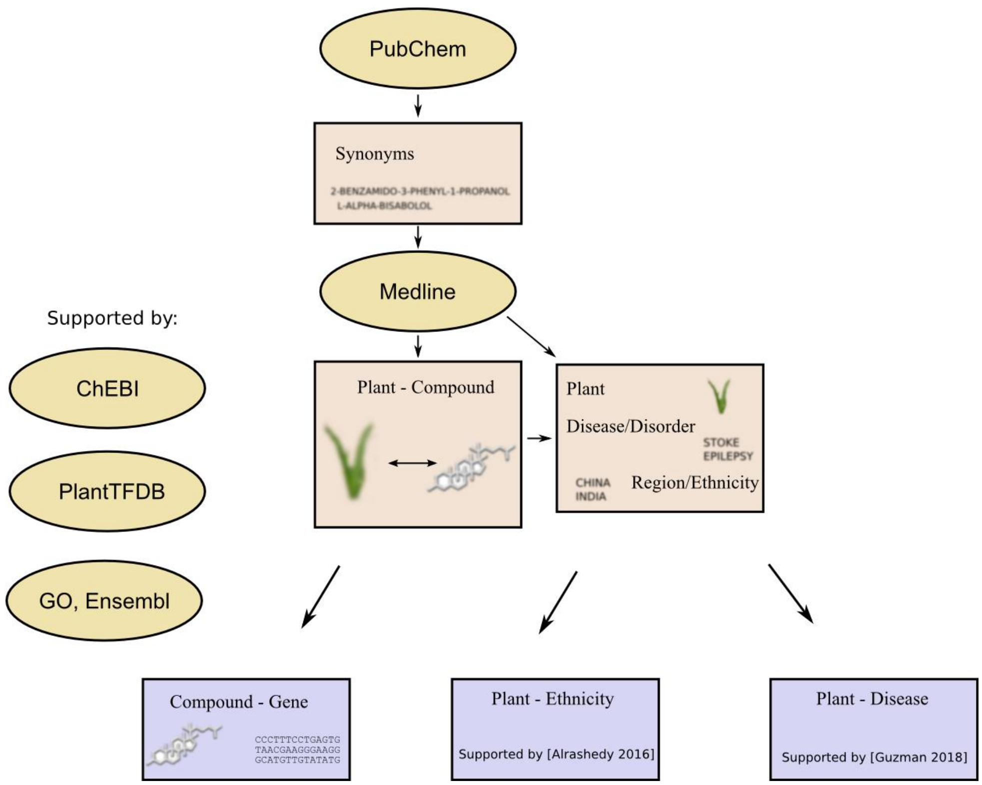

Figure 1: A pipeline used to derive the results presented in the study

A list of tokens with names of medicinal plants (411 entries) was prepared, which include mostly Latin generic names of species. The names of chemical compounds, about 50000 entries, were prepared from synonyms in PubChem project []. An abundant and abigious terms were filtered manually when the lists of medicinal herbs and chemical compounds were prepared.

Bibliographies from Medline were downloaded with the use of Entrez utility, as a result of text-based search for each of prepared tokens (generic names of the plants). Selected bibliographical records were screened to select entries where names of chemical compounds are mentioned in the title or abstract. This allows to compile the database of relations between medicinal herbs and chemical compound.

Also, the presence of geographic terms and names of diseases in the bibliographies allow to count number of records where these terms are met, together with reference to any of compounds.

The network of ontology terms in ChEBI database was used to separate four generic groups of compounds with references to that database. These groups (alkaloids, flavanoids, terpens and lactons), were used in the first type of downstream analysis, together with taxonomy of plants available at NCBI site. The corresponding groups of genes in plant genomes were separated using network of terms, like “flavanoid metabolic process” (GO:0009812) in 'Gene Ontology' project. Two approaches for access genome annotations were used for comparative processing of the plant genomes. First approach was based on annotations available at Plant TFDB site. Second approach was based on a reference genome annotation of Arabidopsis thaliana and pan-genome alignments available at EnsemblPlants site.

The constructed tables of counts of bibliographic records were used to present the generic relations between geographic terms, and between terms for diseases. The dendrogramms of relations were constructed using UPGMA clustering. Kendall correlation measure, correspondence analysis and PCA decomposition, used to process tables of counts, were used as it is implemented in scipy, scikit-bio and sklearn python packages. The charts were prepared with the use of matplotlib graphics library. The developed pieces of code on python and C++ are available at ..... . Also, a suite of tools available at d3b-charts.bri-shur.com site was used to present the tables of counts.

Results

1. Relations between medicinal plants and chemical compounds

The generic relations between evolution in plant kingdom and the use of plants in phytotherapy were systematically studied in [Sharma, V., & Sarkar, 2013; Hao & Xiao. 2015]. A dependencies between taxonomic group of plants and therapeutic use were confirmed in these and similar more specific studies. But these dependencies are uneven and some narrow groups of plants are featured with high diversity of active compounds.

In the way to connect the content of active compounds in phytotherapy and available annotations of plant genomes, the generic classes of ontologies for both compounds and gene families could be used. But due to the complexity of the subject, the misannotations in plant genomes are typical [Shnoes et al., 2008], besides another sources of errors. And in attempts to qualitatively estimate the abundance of gene families using alternative approaches of comparative genomics, some inconsistencies are observed.

In the fig. 2 the estimated relations between generic classes of compounds and generic clades of plants are shown, together with the relations for enzyme families related to the classes of compounds. As the distributions to be compared are highly uneven, rank-based transformations are used to for a color scheme in all three charts in the figure. The relative levels of abundances within each clade of plants are transformed to the non-parametric coefficient of correlation, as it is described in MethodsX supplement.

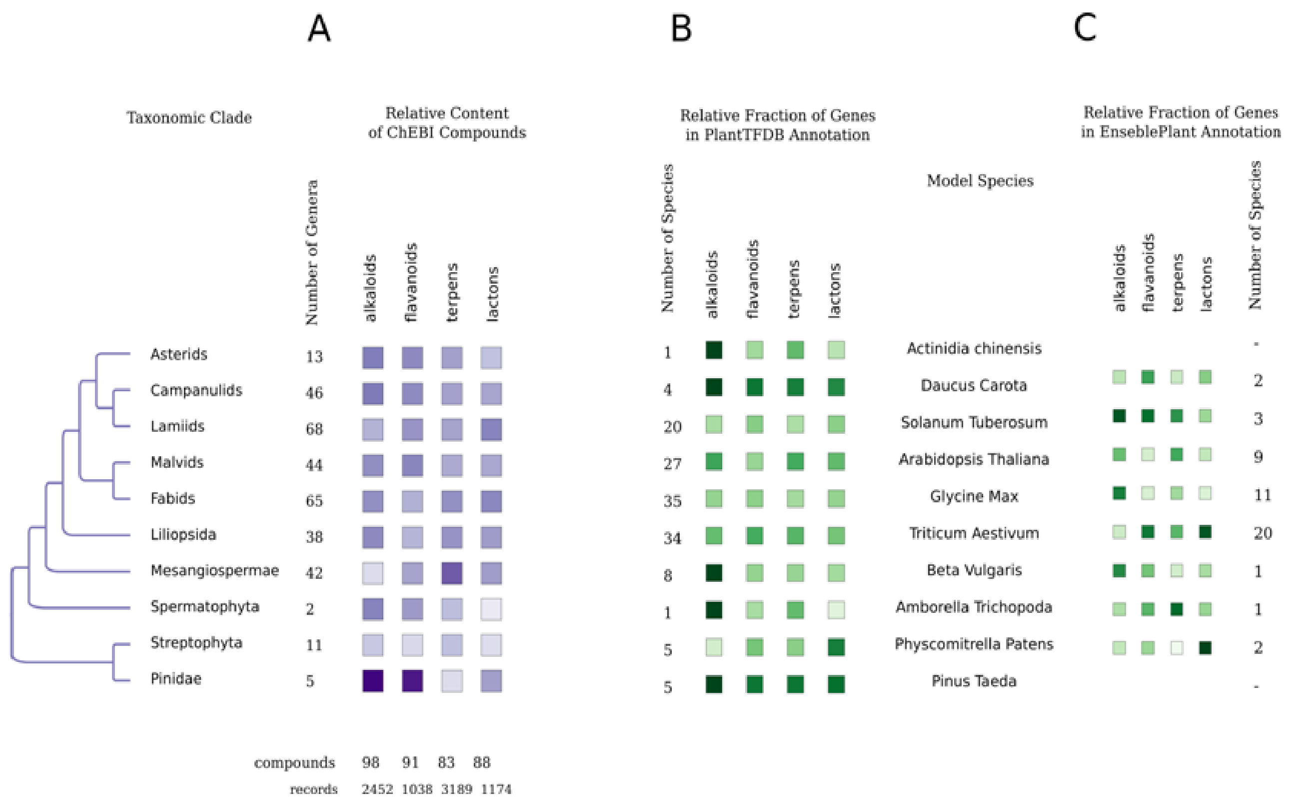

Figure 2: (A) Relations between content of compounds and a taxonomic clade of medicinal plants.
(B,C) Relations between abundance of specific enzymes in plant genomes and a taxonomic clade of the plants, estimated using annotations from PlantTFDB site (B) and pan-genome alignments from EnsemblePlants site (C).

A moderate correspondence was detected for a comparison between distribution of compounds and distribution of related enzymes following the annotations from PlantTFDB: Kendall correlation was 0.16 (p-value 0.06). An insignificant (k.c. 0.08/pv=0.2) correlation was detected for a comparison between the two approaches for a comparison of genomes: on annotations from PlantTFDB site and on pan-genome alignments from EnsemblePlants site. This discrepancy shows the limit of significance, available in the attempts to coincide the loosely related domains.

2. *Relations between traditions of herbal medicine*

The relations between herbal medicines reflect relations between cultures. In the the review of [Pan et al., 2014] where traditions of Chinese, Indian and Arabic medicine are systematically compared, unique features of each of the herbal traditions are described, as well as

But as the cultures are evolved, differentiating from the same origin, the question about relation between traditions could be asked in the same way. The references to herbal tradition are common in biomedical abstracts on herbal medicine. So, a post-processing of the database of chemical compounds allow to estimate relations between traditional medicines, using a pre-defined list of geographic terms. As the the abundances of the geographic terms are highly uneven, Kendall measure of correlation was used to estimate distance between cultures form a matrix of relations between names of plants and geographic terms.

In Fig. 3, the reconstructed relations are shown as dendrogramm and as bubble chart. In an extensive review of psychoactive plants provided in [Alrashedy and Molina, 2016], one or several indigenous cultures are specified for each of the plant. This allows to reconstruct dendrogramm of proximity between “drug” cultures, and to use it for a comparison and verification of estimated proximity between herbal cultures.

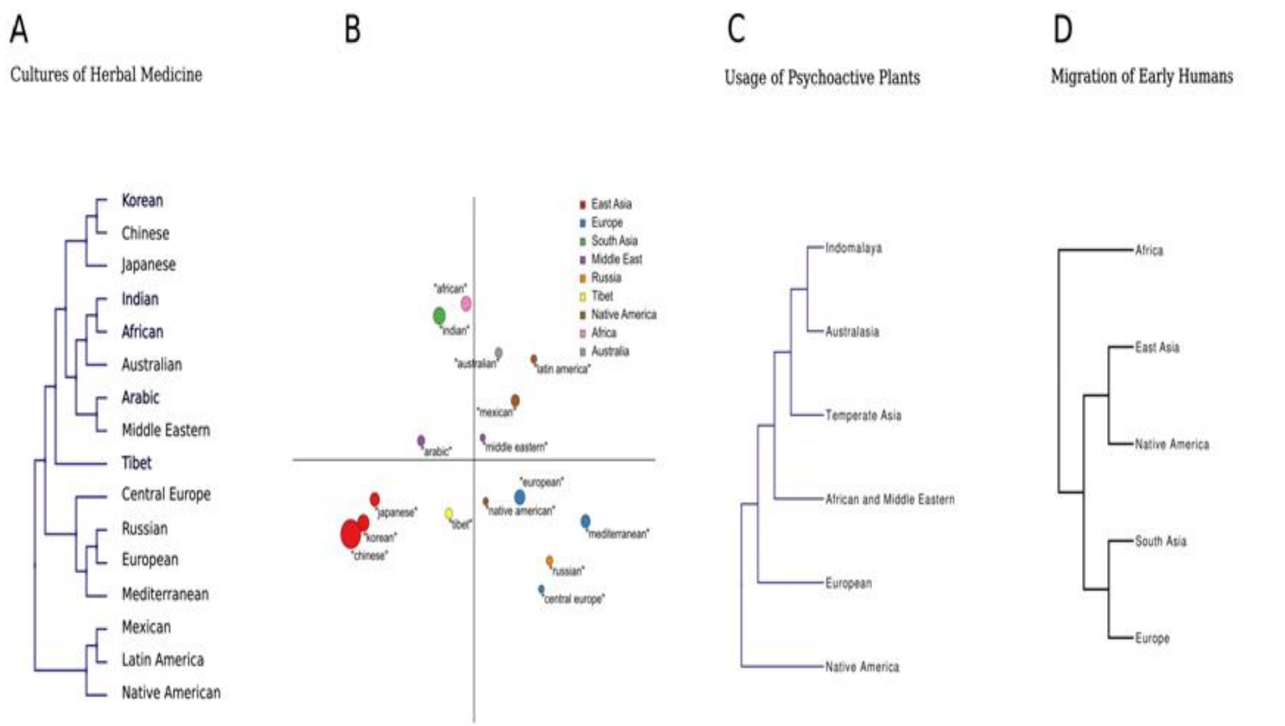

Figure 3: Dendrogramm (A) and correspondence analysis chart (B) for relations between traditions of herbal medicine. The points on a chart represent keywords, which are used to denote a specific herbal culture.
The size of circle shows the relative number of annotations in the subset, which include a corresponding keyword.
(C) Relations between traditions of drug culture, reconstructed from list of psychoactive drugs [Alrashedy & Molina, 2016].
(D) Relations between ethnic groups, as a well-established result in studies on evolution of early humans [Henn et al., 2008].

Also in Fig. 3 the well-established [Henn et al., 2012] relations between origins of ethnic groups are shown. The associations of culture with genotype of ethnic groups are ether known [Li et al 2011], or suggested [Lumsden & Wilson, 1980]. The dendrogramms of proximities between 'drug' cultures and generic herbal cultures are in consistency, and the its differences with genetic evolution of for ethnic groups could help to recognize features specific to an evolution of herbal medicine. At least in can be said that the settlement in the areas with new fauna was resulted in the discoveries of new medicinal plants and a divergence of herbal traditions. And, the authority of traditional Chinese medicine as most stable and conservative of herbal cultures is confirmed in the presented estimates.

3. Relations between chronic diseases / disorders and medicinal plants

A list of tokens which specify chronic diseases and health disorders was also used for a post-processing of the database of chemical compounds. This allowed to construct the matrix of relations between names of diseases and names of plants. Bubble chart on fig. 4 present the proximities between human diseases, constructed from that matrix using direct PCA decomposition. As a conjugated result, coordinates along the same principal axes were assigned to each of plants, as rows of the matrix.

On the chart, the first principal axis separates mental disorders from oncological and cardiovascular diseases. These latter diseases is a cause for most of the deaths in elder age, and the estimated position of metabolic disorders, like diabetes of second type, support the obtained separation. The distribution of diseases along second principal axis also has an intrinsic consistence. Abundances of the terms in the biomedical abstracts is highly uneven, but the dendrogramm of proximities between diseases based on Kendall rank correlation confirms the estimated separation of diseases on PCA chart.

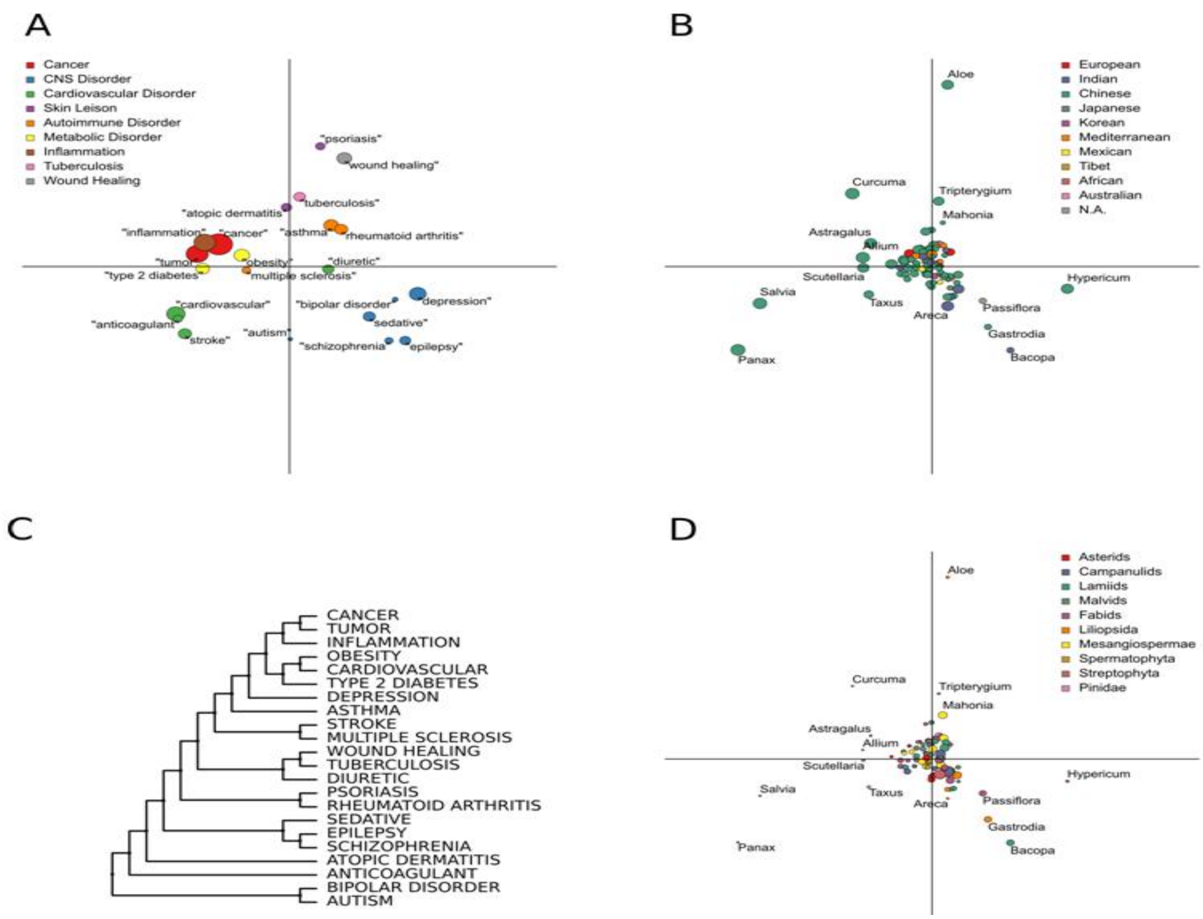

Figure 4: Relations between terms used for notation of diseases, and medicinal plants, decomposed using PCA (A,B,D) and compared using Kendall measure of correlation (C).
(A) - scatter chart for relations between diseases. Size of circle show the relative frequency of term used for notation of a disease.
(B) - scatter chart for relations between medicinal plants. Size of circle show the relative frequency of term used for notation of a medicinal plant. Color show the traditional culture associated with a medicinal plant.
(C) - scatter chart for relations between medicinal plants, positions of circles are the same as on section B. Size of circle show the proportion between relative frequency of term and a number of diseases associated with the term. Color show the taxonomic clade of the medicinal plant.
(D) Relations between diseases, presented as UPGMA dendrogramm.
In sections B and C, the list of plants is limited to 80 terms, selected as 40 + 40 terms with maximal values on first two principal component.

Separation of herbs along conjugated principal axes (fig 4 B,D) follow the separation of diseases, so it could be used for verification of the obtained distribution. The number of herbs shown as circles on fig. 4 B,D was limited to 40 names with maximal absolute coordinates, along each of the axes. 26 plants from 40 selected for a first principal component were mentioned in [Guzman and Molina, 018], where plants with cardiovascular activity were systematically described; among them, Panax(Ginseng), Hypeicum (St. John's wort), Salvia, Scuttelaria, Allium, Astragalus. Form these herbs, a bias was toward a negative direction of the axis, a scale of bias was estimated as 0.26 (Kendall correlation) . An unsignuficant (k.c 0.06) bias toward opposite direction was detected in 16 herbs from the list of 40, mentioned as psycohoactive in [Alrashedy and Molina, 2016]. Three herbs mentioned as effective for wound healing in [Cazander et al., 2012], from the list of 40, Melissa officinalis, Uncaria tomentosa and Aloe vera, are also grouped in the direction towards psychic disorders.

In the matrix of relation between herbs and names of diseases, the number of diseases mentioned for a herb could be evaluated. These values are shown on fig. 4D as sizes of the circles, and the trend is observed that psychoactive plants have more diverse spectrum of prescriptions. This is confirmed by Kendall measure of correlation (0.41). Also, a number of compounds detected in a herb, relatively to a number of bibliographies, is also correlated with value of first principal component (k.c. 0.26).

Discussion

The presented chart for a distribution of human diseases is of insufficient significance to be used for a deeper discoveries. But it provides, in an acceptable level of verification, an expressively simple relations for the most of terrible sufferings, the major problems in a modern medicine. Only uncertain hints could be proposed to interpret the obtained distributions. Among them is the observed correlation between diversity of effects of a plant to a health and diseases of psychic. In another side of a distribution are the plants with uniformly directed activity, with an association to Chinese tradition of medicine.

The balance between benefits and risks to a health is very delicate, for any plant. The example from the history, for psychoactive plants, is the narrative from [Hamarneh 1972]. It tells how the verse in Qur'an, that the harm of the wine is greater than it's benefit, was transformed to expansion of drug addiction in medieval Arabic culture. But, in another side of risks related to brain disorders, there are terrible diseases which almost inevitably lead to death. This suggests that the conservativity and uniform mode of action is just another side of the problem, a slow but stable path towards nowhere. The solution should be sought somewhere in-between. In a terms of proverb from Bible, the path to the Tree of Life may exist, but as a precise balance between denial and misuse, between diversity and stability. But it is impossible to estimate for any of humans...

The cases of mistreatment [Jargin 2019] are the excuse to strict requirements of rigorous prove in legal regulation in modern herbal medicine. But despite the lack of strict verification for most of the medicinal plants, the use of herbal preparations is expanded in recent times [Pan et al, 2014]. And the results of present study could support that no contradictions are between modern scientific approaches and generic traditions of herbal medicine. But it also reveal a presence of inconsistencies and even nonsenses within each of both mentalities.

References

1. Alrashedy, N. A., & Molina, J. (2016). The ethnobotany of psychoactive plant use: a phylogenetic perspective. *PeerJ*, *4*, e2546. doi:10.7717/peerj.2546
2. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., … Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics*, *25*(1), 25–29. doi:10.1038/75556
3. Cazander, G., Jukema, G. N., & Nibbering, P. H. (2012). Complement activation and inhibition in wound healing. *Clinical & developmental immunology*, *2012*, 534291. doi:10.1155/2012/534291
4. Choi, W., Kim, B., Cho, H., Lee, D., & Lee, H. (2016). A corpus for plant-chemical relationships in the biomedical domain. *BMC bioinformatics*, *17*, 386. doi:10.1186/s12859-016-1249-5
5. Degtyarenko, K., de Matos, P., Ennis, M., Hastings, J., Zbinden, M., McNaught, A., … Ashburner, M. (2007). ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic acids research*, *36*(Database issue), D344–D350. doi:10.1093/nar/gkm791
6. Guzman, E., & Molina, J. (2018). The predictive utility of the plant phylogeny in identifying sources of cardiovascular drugs. *Pharmaceutical biology*, *56*(1), 154–164. doi:10.1080/13880209.2018.1444642
7. Hao, D. C., & Xiao, P. G. (2015). Genomics and Evolution in Traditional Medicinal Plants: Road to a Healthier Life. *Evolutionary bioinformatics online*, *11*, 197–212. doi:10.4137/EBO.S31326
8. Hamarneh S. (1972). Pharmacy in medieval Islam and the history of drug addiction. *Medical history*, *16*(3), 226–237.
9. Henn, B. M., Cavalli-Sforza, L. L., & Feldman, M. W. (2012). The great human expansion. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(44), 17758–17764. doi:10.1073/pnas.1212380109
10. Herrero, J., Muffato, M., Beal, K., Fitzgerald, S., Gordon, L., Pignatelli, M., … Flicek, P. (2016). Ensembl comparative genomics resources. *Database : the journal of biological databases and curation*, *2016*, baw053. doi:10.1093/database/baw053

11. Sergei V. Jargin Drugs and dietary supplements with unproven effects in research and practice JOURNAL OF COMPLEMENTARY MEDICINE RESEARCH, 2019 VOL 10, NO. 1, PAGE 27–37
12. Jin JP, Tian F, Yang DC, Meng YQ, Kong L, Luo JC and Gao G. (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Research, 45(D1):D1040-D1045.
13. Li, D., Zhao, H., & Gelernter, J. (2011). Strong association of the alcohol dehydrogenase 1B gene (ADH1B) with alcohol dependence and alcohol-induced medical diseases. *Biological psychiatry*, *70*(6), 504–512. doi:10.1016/j.biopsych.2011.02.024
14. Lumsden, C. J., & Wilson, E. O. (1980). Translation of epigenetic rules of individual behavior into ethnographic patterns. Proceedings of the National Academy of Sciences of the United States of America, 77(7), 4382–4386.
15. Pan, S. Y., Litscher, G., Gao, S. H., Zhou, S. F., Yu, Z. L., Chen, H. Q., … Ko, K. M. (2014). Historical perspective of traditional indigenous medical practices: the current renaissance and conservation of herbal resources. Evidence-based complementary and alternative medicine : eCAM, 2014, 525340. doi:10.1155/2014/525340
16. Sharma, V., & Sarkar, I. N. (2013). Leveraging biodiversity knowledge for potential phyto-therapeutic applications. *Journal of the American Medical Informatics Association : JAMIA*, *20*(4), 668–679. doi:10.1136/amiajnl-2012-001445
17. Schnoes, A. M., Brown, S. D., Dodevski, I., & Babbitt, P. C. (2009). Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS computational biology*, *5*(12), e1000605. doi:10.1371/journal.pcbi.1000605
18. Xue, R., Fang, Z., Zhang, M., Yi, Z., Wen, C., & Shi, T. (2012). TCMID: Traditional Chinese Medicine integrative database for herb molecular mechanism analysis. *Nucleic acids research*, *41*(Database issue), D1089–D1095. doi:10.1093/nar/gks1100