



Contents lists available at ScienceDirect

Engineering

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

Views & Comments

Natural Diverse Components, Herbgonomics, and Artificial Intelligence: A Multi-Dimensional Paradigm for Polyherbal Formulation Research

Liang Leng^{a,b,*}, Jing Mo^{a,b,c}, Hongguo Chen^d, Shenjie You^{a,b}, Can Wang^{a,b}, Jun Lu^e, Zhaotong Cong^{a,b}, Wei Chen^{a,b}, Chi Song^{a,b}, Sanyin Zhang^{a,b,*}

^aInnovative Institute of Chinese Medicine and Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China

^bInstitute of Herbgonomics, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China

^cHubei University of Chinese Medicine, Wuhan 430065, China

^dState Key Laboratory of Quality Research in Chinese Medicine, Institute of Chinese Medical Sciences, University of Macau, Macao 999078, China

^eSchool of Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China

1. Research on polyherbal formulations (PHFs) lags far behind research on its constituent herbs and compounds

A PHF is characterized by the inclusion of multiple herbs for treatment purposes and is a significant aspect of traditional medicine systems, including traditional Chinese medicine (TCM), Ayurveda, and Kampo [1,2]. In general, a PHF must include multiple herbs, with each herb comprising a variety of compounds. The hierarchical organization of PHF introduces two key elements: firstly, there may be a synergistic interaction among the various compounds or herbs, which can enhance the effectiveness of a PHF compared to individual components; secondly, the intricate nature of PHF complicates the process of studying it.

Notably, a PHF may encompass various kinds of components, including secondary metabolites, natural macromolecules (such as DNA, RNA, and peptides), and other types of metabolites [3,4]. From the perspective of modern science, the medicinal properties of secondary metabolites in these components are the most studied and best understood [5,6]. The isolation of morphine, the first known alkaloid, was accomplished in 1803 [7]. This landmark discovery revolutionized the research paradigm of natural medicine study. Morphine is derived from opium, a plant which has a long-standing history of medicinal application that spans several millennia, and addresses issues such as toothaches, bruises, coughs, and diarrhea [8,9]. The introduction of morphine marked a significant milestone, as it was the first instance where individuals acknowledged that a plant could effectively alleviate health problems due to its distinctive compounds, thereby paving the way for a more thorough exploration of the mechanism that enable the herb to combat disease. The isolation of secondary metabolites, combined with subsequent pharmaceutical assessments, has resulted in a series of significant successes including the identification of acetylsalicylic acid (aspirin), Penicillin, Paclitaxel, and

Artemisinin. Even after two centuries since morphine's discovery and despite considerable hurdles such as high-throughput screening (HTS) difficulties, natural substances continue to produce compounds for drug discovery [5,10].

The fact that secondary metabolites come from herbs, predominantly medicinal plants, highlights the importance of studying these plants in the field of natural medicine. One of the primary objectives of researching medicinal plants is to guarantee a consistent availability of specialized metabolites. Paclitaxel has been validated as an effective chemotherapy medication for a variety of cancers, such as lung and ovarian cancers, among others. Paclitaxel was first sourced from the bark of the Pacific yew tree, but the supply is significantly limited due to two factors: the Pacific yew tree's slow growth rate and the restricted concentration of Paclitaxel found in its bark. The "Taxol crisis" illustrates a similar issue regarding the availability of specialized metabolites derived from medicinal plants, thus, prompting a broad spectrum of studies focused on the ecological and evolutionary dynamics of these medicinal plant species [11]. Another primary objective is the biosynthesis of secondary metabolites [12]. Investigating and isolating secondary metabolites like morphine shows that the health benefits of herbs result from specific compounds found within them. Due to their specialized metabolism, different herbs produce unique compounds, which allows them to be effective in treating a variety of diseases. The advancements in genomics during the early part of the 21st century have particularly enhanced the collaboration of medicinal plant research, thus leading to the establishment of a new field termed herbgonomics [13]. Herbgonomics primarily supports research on medicinal plants through two key areas of focus. First, it contributes to the accurate classification of herbs, revealing their evolutionary features like population structures and phylogenetic relationships, which consequently promotes breeding initiatives. Second, it provides insight into the specialized metabolism, detailing regulatory mechanisms, and ultimately paves the way for innovations in metabolic bioengineering [14–16].

Despite significant progress in the fields of secondary metabolites and herbal research, the research trajectory for PHFs

* Corresponding authors.

E-mail addresses: lling@cdutcm.edu.cn (L. Leng), tcmzsy@cdutcm.edu.cn (S. Zhang).

<https://doi.org/10.1016/j.eng.2026.01.021>

2095-8099/© 2026 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

is somewhat constrained primarily due to the difficulty of achieving a comprehensive understanding of this complicated topic. For example, the Huo Xiang Zhen Qi formulation, which contains ten herbs, has been reported to be beneficial in the treatment of severe acute respiratory syndrome (SARS) [17], colorectal cancer [18], and gastrointestinal diseases [19]. Nevertheless, research dedicated to identifying efficacious herbs or compounds for diverse disease treatments remains restricted. These challenges cannot be effectively addressed through the former research paradigm, thus, necessitating the establishment of a new one.

2. Herbgenomics-based investigation of all natural diverse components of PHFs

Here, we propose a new paradigm for PHF research with herbgenomics-based investigation of all natural diverse components research at its core. We first explain why we need to focus on all the natural diverse components and then explain why this relies on herbgenomics research.

First, the herbs included in a PHF are limited in quantity, usually numbering less than 100 and often fewer than 10, whereas the components are present in significantly larger quantities—often by several orders of magnitude. As a library of metabolites for a single species, the Human Metabolome Database currently has more than 200 thousand compounds [20]. Thus, the overwhelming number of compounds is an inescapable challenge in PHF research, and marks it as the most crucial issue within this field. Second, PHF not only contains secondary metabolites but also encompasses other types of compounds including peptides, nucleic acids, and carbohydrates. Since all these types of compounds have reported therapeutic effects on specific diseases, none of them should be omitted when studying PHF. Third, an attractive but hard-to-study property of PHF is the synergistic effect between its components. Without a comprehensive view of all natural diverse components within a PHF, it is always possible to ignore the true component or synergistic relationship that treats a disease. Given that the numerous and diverse components of PHF can undergo synergistic interactions, all its natural components must be studied.

Since obtaining the diverse components of PHFs is the core issue of PHFs research, the new paradigm of PHF research should aim to overcome this issue. Given that the chemical nature of these diverse components varies, the approaches to isolating them are also quite complicated. Thus, it can be highly challenging, or even unattainable, to collect all these diverse components when relying on traditional methods of chemical investigation. A set of methods that can be used to study diverse components has been developed based on biological principles [21]. The fundamental concept stems from the fact that every natural diverse component of any herb is either directly or indirectly encoded by the genomic sequence of that herb. For instance, all forms of RNA are generated from the DNA sequence, while most peptides are created by translating messenger RNA (mRNA) sequences. Thus, these components could be defined as direct gene-encoded components. Meanwhile, the genomic sequence plays a central role in the indirect determination of certain secondary metabolites. To synthesize any secondary metabolite, the necessary enzymes must be present, and these enzymes are specified by the organism's genetic sequence. Consequently, one can argue that the genomic sequence establishes the upper limit of diverse natural substances that any herb can produce. Herbgenomics provides a range of tools and methods to study the natural diverse components of herbs. The genomic sequence can facilitate the identification of components through genome mining, followed by verification using transcriptomics, proteomics, and metabolomics techniques. This strategy has proven effective when applied to a multitude of herbs and has resulted

in the discovery of hundreds of millions of natural diverse components. Recently, Chen and colleagues [21,22] developed the Gene-encoded Natural Diverse Components Repository (GNDC), which utilizes custom pipelines and artificial intelligence (AI) algorithms for the genomic analysis of herbs. This methodology facilitated the identification of 2.32 million secondary metabolites, 229 million small peptides, 2.38 million small RNAs, and 0.26 million carbohydrates. However, it is believed that these numbers underestimate the total number of components in these herbs as current omics data might overlook certain developmental stages or tissues where many secondary metabolites are specifically produced. Although it remains unconfirmed whether herbgenomics-based approaches can capture the natural chemodiversity of herbs, the results hint at the substantial potential of this approach [10,12].

3. AI-powered multi-dimensional paradigm for PHF research

Based on herbgenomics-based mining of natural diverse components, here we proposed a multi-dimensional paradigm of PHF research. This paradigm includes multiple layers of PHFs such as compounds, herbs, and PHFs themselves; multiple omics techniques such as genomics, transcriptomics, proteomics, and metabolomics; multiple experimental platforms such as HTS platforms, pharmacological platforms, omics, and platforms; and multi-dimensional data. Starting with a designated PHF, we first gathered the herbs along with their genomic sequence. Next, we extracted all the natural diverse components from the herbs, and carried out HTS and a range of pharmacological tests on some of these components. Multi-faceted data could be collected for every single component. Last, we leveraged statistical modeling and AI to unify data from all components, obtaining a complete overview of the focal PHF (Fig. 1).

To start, we created a comprehensive list of all individual herbs in a PHF and gathered their genome sequences. It should be highlighted that not all herbs found in a PHF possessed a fully assembled and annotated genome (Figs. 1(a) and (b)). Next, a series of analysis published along with GNDC enabled a herbgenomics-based study of secondary metabolites, small peptides, and small RNAs [21] (Fig. 1(c)). There are additional types of natural diverse components, like carbohydrates, which cannot be entirely covered by GNDC's methodology at this time. Nonetheless, it is the most cutting-edge technique, employing genomic information and AI to increase yield and allow for detailed analysis. After collecting as many components as possible, all the components can be used for virtual screening, such as molecular docking and flexible docking based on AI. Then, HTS experiments should be conducted since it is essential to discover the cellular target for individual component (Fig. 1(d)). For example, G protein-coupled receptors (GPCRs), which represent a superfamily of membrane proteins, are therapeutic targets for roughly one-third of all marketed pharmaceutical products. They have also been identified as potential targets for various compounds extracted from traditional medical treatments. Genome-wide pan-GPCR cell libraries represent a robust tool that significantly enhances the screening of GPCR ligands, enables detailed investigations into GPCR signaling pathways, and improves drug safety evaluations, thereby ultimately accelerating the process of GPCR-targeted drug discovery. HTS based on the pan-GPCR cell libraries should take full advantage of genome-wide pan-GPCR cell libraries that allow for the simultaneous exploration of all human GPCRs as targets [19,20]. Together with omics and pharmacological research, it is possible to compile extensive multi-faceted data for each component, encompassing its target, downstream signaling pathways, pharmacodynamics, and other pertinent details regarding its safety and therapeutic efficacy (Fig. 1(e)). Finally, by combining multi-dimensional data from each

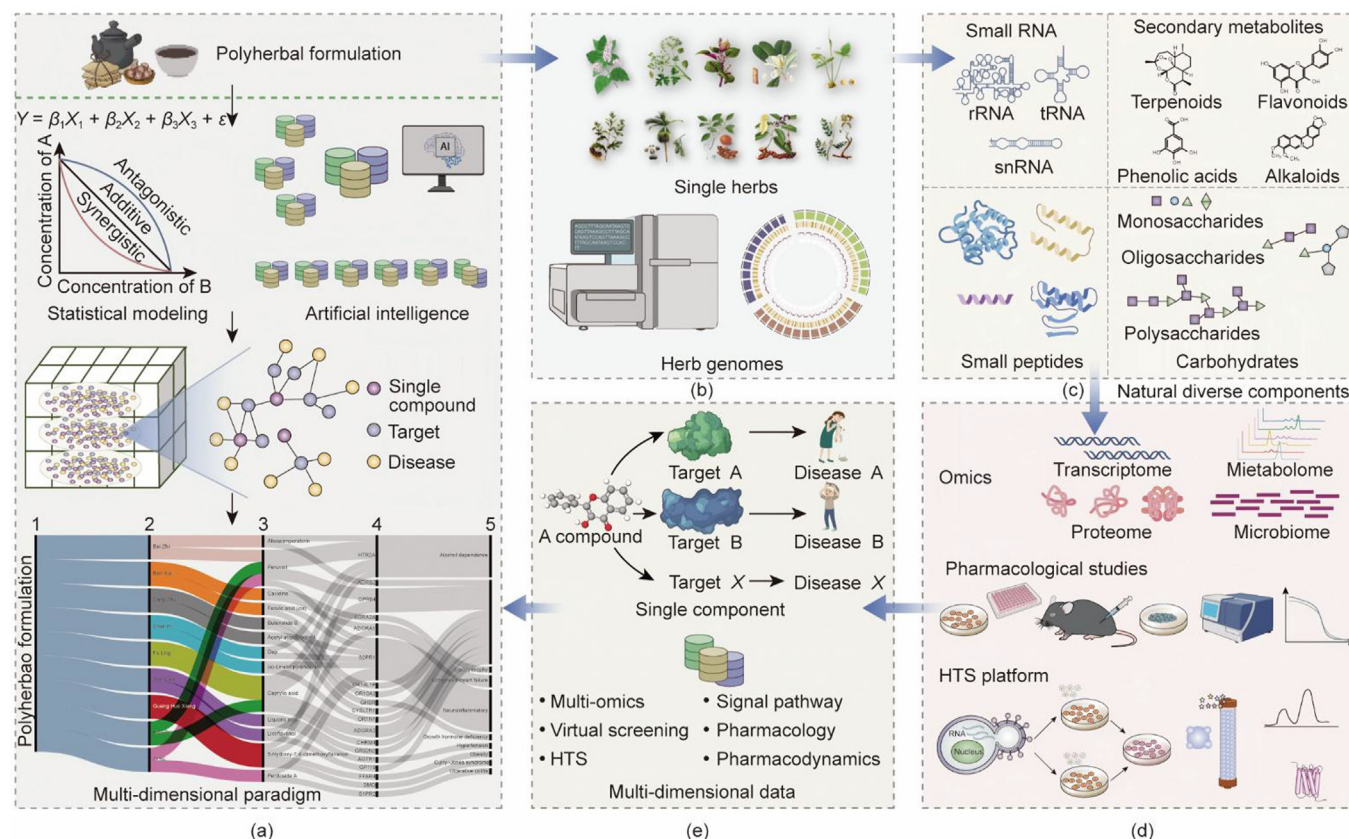


Fig. 1. Illustration of a multi-dimensional paradigm for investigating PHF. (a) Utilizing statistical modeling and AI methods within a multi-dimensional framework is crucial for unraveling the complex network involving PHFs, herbs, components, targets, and diseases; (b) it is essential for the genome of every herb to be sequenced and annotated; (c) the genomic sequences of herbs should be used to identify multiple types of natural diverse components; (d) experiments such as HTS are necessary to uncover the cellular targets of a component; (e) multi-dimensional datasets ought to be generated for each distinct component. rRNA: ribosomal RNA; tRNA: transfer RNA; snRNA: single-nucleus RNA.

element of a PHF, it becomes possible to uncover additive, synergistic, or antagonistic interactions among them, while utilizing statistical analysis, AI, and other computational techniques (Fig. 1(a)). The concluding step may provide answers to the final question: since PHFs are thought to tackle diseases in a multi-target and multi-pathway manner, the complex connections among PHFs, herbs, components, targets, and diseases should ultimately be clarified. Acknowledging that these elaborate relationships will produce a considerable amount of multi-dimensional data and lead to a complicated network, big data-driven AI models must become a vital part of the entire paradigm [22,23]. In other words, this paradigm should essentially be a multi-dimensional, big data-driven, and AI-powered one for drug discovery.

CRediT authorship contribution statement

Liang Leng: Writing – original draft, Conceptualization. **Jing Mo:** Writing – review & editing. **Hongguo Chen:** Writing – original draft. **Shenjie You:** Writing – original draft. **Can Wang:** Writing – original draft. **Jun Lu:** Writing – original draft. **Zhaotong Cong:** Writing – original draft. **Wei Chen:** Writing – original draft. **Chi Song:** Writing – original draft. **Sanyin Zhang:** Writing – review & editing, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] Chen H, Liu Z, Du Y, Meng X, Xue L, Sun X, et al. Comparative analysis of Chinese classical prescriptions and global traditional polyherbal formulations: insights from the database of global polyherbal formulation (GPFD). *Chin Med* 2025;20(1):169.
- [2] Parasuraman S, Thing GS, Dhanaraj SA. Polyherbal formulation: concept of ayurveda. *Pharmacogn Rev* 2014;8(16):73–80.
- [3] Broekaert WF, Cammue BPA, Bolle MFCD, Thevissen K, Samblanx GWD, Osborn RW, et al. Antimicrobial peptides from plants. *Crit Rev Plant Sci* 1997;16(3):297–323.
- [4] Dar RA, Shah Nawaz M, Ahanger MA, Majid I. Exploring the diverse bioactive compounds from medicinal plants: a review. *J Phytopharm* 2023;12(3):189–95.
- [5] Atanasov AG, Zotchev SB, Dirsch VM, Supuran CT. The International Natural Product Sciences Taskforce. Natural products in drug discovery: advances and opportunities. *Nat Rev Drug Discov* 2021;20(3):200–16.
- [6] Wink M. Modes of action of herbal medicines and plant secondary metabolites. *Medicines* 2015;2(3):251–86.
- [7] Wicks C, Hudlicky T, Rinner U. Morphine alkaloids: history, biology, and synthesis. *Alkaloids Chem Biol* 2021;86:145–342.
- [8] Bandyopadhyay S. An 8,000-year history of use and abuse of opium and opioids: how that matters for a successful control of the epidemic? (P4.9–055). *Neurology* 2019;92(15_supplement). P4.9–055.
- [9] Ng SV. Opium use in 19th-century Britain: the roots of moralism in shaping drug legislation. *AJP residents' Am J Psychiatry Resid J* 2016;11(6):14.
- [10] Harvey AL, Edrada-Ebel R, Quinn RJ. The re-emergence of natural products for drug discovery in the genomics era. *Nat Rev Drug Discov* 2015;14(2):111–29.
- [11] Cragg GM, Schepartz SA, Suffness M, Grever MR. The Taxol supply crisis. new NCI policies for handling the large-scale production of novel natural product anticancer and anti-HIV agents. *J Nat Prod* 1993;56(10):1657–68.
- [12] Lacchini E, Goossens A. Combinatorial control of plant specialized metabolism: mechanisms, functions, and consequences. *Annu Rev Cell Dev Biol* 2020;36(1):291–313.
- [13] Hu H, Shen X, Liao B, Luo L, Xu J, Chen S. Herbgonomics: a stepping stone for research into herbal medicine. *Sci China Life Sci* 2019;62(7):913–20.

- [14] Chae L, Kim T, Nilo-Poyanco R, Rhee SY. Genomic signatures of specialized metabolism in plants. *Science* 2014;344(6183):510–3.
- [15] Sun W, Xu Z, Song C, Chen S. Herbgonomics: decipher molecular genetics of medicinal plants. *Innovation* 2022;3(6):100322.
- [16] Pei Y, Leng L, Sun W, Liu B, Feng X, Li X, et al. Whole-genome sequencing in medicinal plants: current progress and prospect. *Sci China Life Sci* 2024;67(2):258–73.
- [17] Lin L, Xu YJ, He DP, Han Y, Tang GH, Yang ZM, et al. A retrospective study on clinical features of and treatment methods for 77 severe cases of SARS. *Am J Chin Med* 2003;31(6):821–39.
- [18] Yeh MH, Chiu HP, Wu MC, Koo M, Lin NW, Liao KK, et al. Integrated Chinese herbal medicine and western medicine on the survival in patients with colorectal cancer: a retrospective study of medical records. *Evid Based Complement Alternat Med* 2020;2020(1):4561040.
- [19] Zhao M, Chen Y, Wang C, Xiao W, Chen S, Zhang S, et al. Systems pharmacology dissection of multi-scale mechanisms of action of Huo-Xiang-Zheng-Qi formula for the treatment of gastrointestinal diseases. *Front Pharmacol* 2019;9:1448.
- [20] Wishart DS, Guo A, Oler E, Wang F, Anjum A, Peters H, et al. HMDB 5.0: the human metabolome database for 2022. *Nucleic Acids Res* 2022;50(D1):D622–31.
- [21] Chen W, Yu Z, Leng L, Sun D, Liu H, Gong RZ, et al. Artificial intelligence-curated repository of gene-encoded natural diverse components from herbal medicines. *Innovation* 2025;6(12):101011.
- [22] Chen W, Liu X, Zhang S, Chen S. Artificial intelligence for drug discovery: resources, methods, and applications. *Mol Ther Nucleic Acids* 2023;31:691–702.
- [23] Song Z, Chen G, Chen CYC. AI empowering traditional Chinese medicine? *Chem Sci* 2024;15(41):16844–86.