

Herbgenomics: Decipher molecular genetics of medicinal plants

Wei Sun,^{1,2,4} Zhichao Xu,^{3,4} Chi Song,² and Shilin Chen^{1,2,*}

¹Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100070, China

²Institute of Herbgenomics, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China

³College of Life Science, Northeast Forestry University, Harbin 150040, China

⁴These authors contributed equally

*Correspondence: slchen@icmm.ac.cn

Received: June 22, 2022; Accepted: September 9, 2022; Published Online: September 14, 2022; <https://doi.org/10.1016/j.xinn.2022.100322>

© 2022 The Authors. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Citation: Sun W., Xu Z., Song C., et al., (2022). Herbgenomics: Decipher molecular genetics of medicinal plants. *The Innovation* 3(6), 100322.

Medicinal plants synthesize a huge repertoire of specialized compounds as protective metabolites when they are confronted with complex abiotic and biotic conditions. Humans have used medicinal plants to treat ailments and maintain health throughout civilization. Paleontological studies have shown

that the application of medicinal plants, such as *Ephedra altissima* and *Centaurea solstitialis*, could be dated back to 60,000 years ago, since their fossils were found in the tomb of prehistoric Neanderthals. Historically, traditional herbal medicine systems, such as traditional Chinese medicine, Ayurveda,

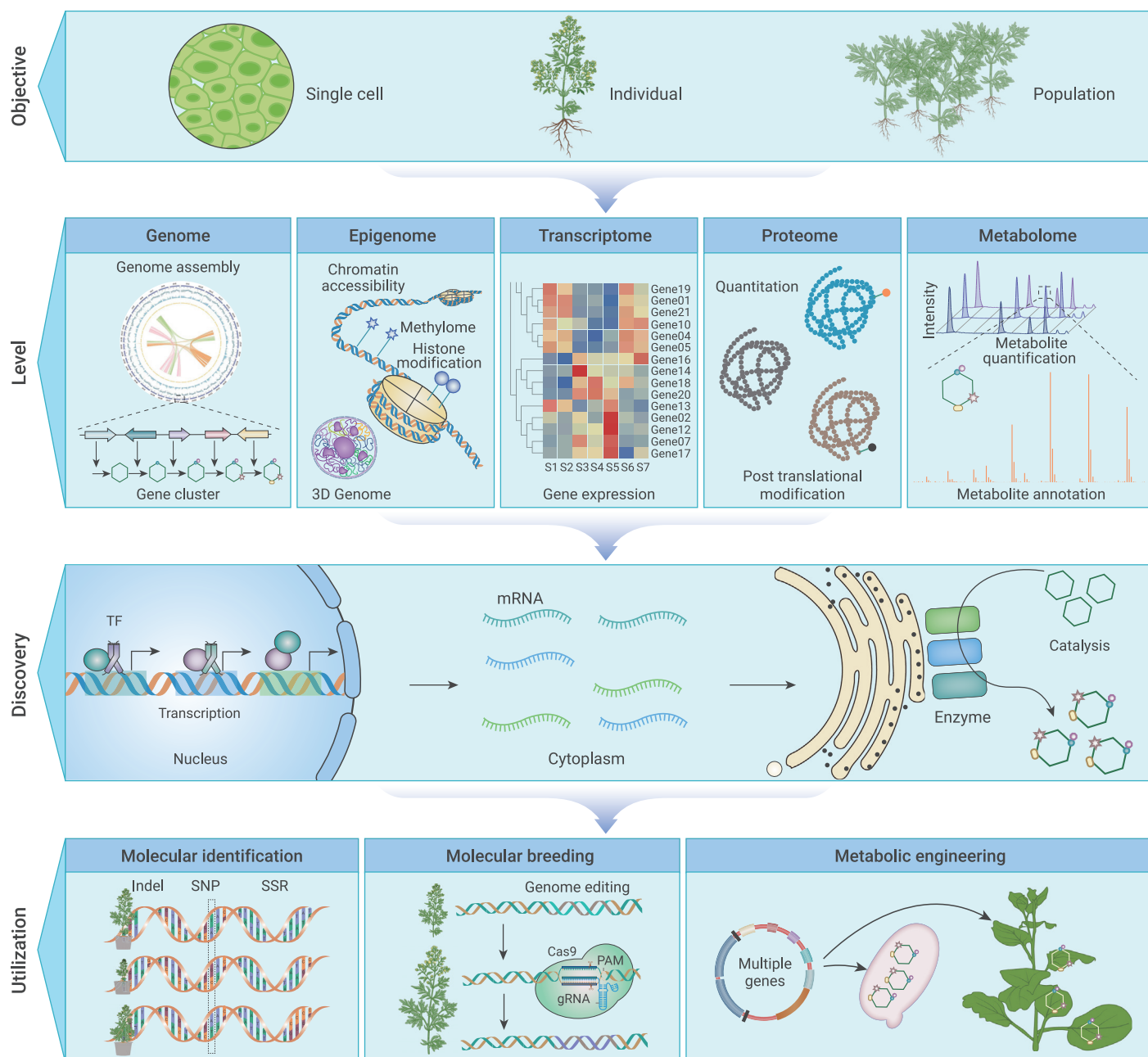


Figure 1. Systematic approaches of herbgenomics to investigate the molecular genetics of medicinal plants (A) Molecular genetics studies on medicinal plants focus primarily on single cells, individuals, and populations. (B) Integration of omics data, including genomic, transcriptomic, proteomic, metabolomic, and epigenomic data, may accelerate the discovery of the molecular genetic basis determining beneficial traits in medicinal plants. (C) Discovery of the molecular mechanism related to beneficial traits in medicinal plants, particularly the synthesis and regulation of phytopharmaceutical compounds. (D) Application of molecular genetic data on molecular identification, molecular breeding, and metabolic engineering.

traditional Arabic and Islamic Medicine, and traditional Malay medicine, are considered as the main natural healthcare systems across the globe. As early as in the ninth century, advanced technologies of chemical isolation and pharmacological tests contributed to the development of modern phytochemical compounds. Numerous plant natural products discovered by western and eastern scientists, including artemisinin, aspirin, atropine, ephedrine, morphine, podophyllotoxin, vinblastine, and taxol, have been clinically applied as therapeutic agents. However, intractable concentrations as well as limited or uneven quality of herb resources often restrict the exploitation of their pharmaceutical potentials and clinical applications.

To secure a high-quality and sustainable supply of herbal materials, some factors must be taken into account, such as accurate identification, superior varieties, environmental conditions, cultivation management, and cultural practices. Development of genetically elite varieties with desired traits under appropriate environments ought to be firstly guaranteed for ensuring high-quality herb cultivation. With the aid of omics tools, herbgenomics has recently become an emerging field, which investigates the molecular genetic basis of species identity and beneficial traits in medicinal plants, particular the synthesis and regulation of phytopharmaceutical compounds via integration of at least a part of the genomic, transcriptomic, and metabolomic data.¹ Most reverse genetics approaches supplemented with some cases from forward genetics have been applied to positioning of genes that determine bioactive specialized metabolites and qualify phenotypes from the individual, population, or mutant library of model herbs, such as *Artemisa annua*.² Herbgenomics sheds light on the two main aspects described in detail below (Figure 1).

ACCURATE IDENTIFICATION, EVOLUTION, AND MOLECULAR BREEDING

The increasing availability of reference genomes with a complete genomic variation repertoire has benefited the foundation of molecular genetics of medicinal plants in the last decade.³ The size and complexity of genomes do not hinder the *de novo* generation of genome assemblies, with complex genomes represented by *Ginkgo biloba*, *Panax ginseng*, and *Taxus chinensis* having been successfully deciphered. At an affordable cost of genome sequencing and re-sequencing, a burst of inter- or intragenomic comparisons, such as those conducted in genera *Cannabis*, *Erythroxylum*, *Panax*, *Papaver*, and *Perilla*, have uncovered the linkage of extensive gene content variations with pharmaceutical traits during domestication or the evolutionary process. A metabolome-based genome-wide association study, which has integrated growing genomic and metabolomic resources of wild and cultivated populations of *Cannabis* and *Perilla*, was performed to identify allelic variants associated with variations in pharmaceutical compounds, leading to the discovery of enzymes and regulators involved in specific biological processes. A genomic database of medicinal plants from global pharmacopoeia has been created to deposit all genomic sequences, which is useful for the molecular identification of raw materials. Nevertheless, one reference genome and short reads from re-sequencing cannot adequately represent the whole spectrum of sequence diversity within a species. The affordable whole-genome sequencing will enable more accomplishments of haplotype-resolved genomes, which will boost the use of pan-genome-wide association studies for discovering genes controlling pharmaceutical traits by coupling with significantly advanced metabolomics and phenomics technologies.

Medicinal plant cultivation has been developed systematically over the past few decades, leading to the release of multiple high-yield varieties, most of which originated directly from wild land or require long generation time during breeding. A major challenge for breeders is how to increase the effectiveness and efficiency of selection and accelerate breeding progress to meet the requirement of pharmaceutical manufacturing. Molecular marker-assisted breeding offers a new approach for developing medicinal plant cultivars, complementing conventional breeding selection and serving as an extremely powerful methodology. In addition, based on the comprehensive understanding of the molecular mechanism of desired traits, *de novo* domestication has been proposed as innovative tactics for medicinal plant breeding with the aid of breakthrough biotechnologies, especially genome editing and genetic transformation. To achieve our new breeding goals, domestication-related traits should be rapidly introduced into eminent wild materials by utilizing a combination of genetic and breeding tools to create new cultivars that harbor beneficial traits. Specifically, scientists have moved forward to utilize more advanced techniques such as clustered regularly interspaced short palindromic repeats-associated

protein-9 nuclease to edit genes encoding catalytic enzymes or transcription factors (TFs) that control pharmaceutical and toxic compound biosynthesis and regulation, thereby improving these major pharmaceutical traits.

PATHWAY ELUCIDATION, REGULATORY MECHANISM, AND METABOLIC BIOENGINEERING

The boom in transcriptome and genome sequencing data allows scientists to facilitate the interpretation of underlying genomic basis of metabolic pathways in a wide range of medicinal plant species. The transcriptome-based and above-mentioned genomic approaches yielded pioneering achievements of nearly complete biosynthesis of phytochemical drugs or important intermediates, including notable cannabidiol, colchicine, glycyrrhizic acid, diosgenin, and hyoscyamine. It is worth noting that genomic annotation of metabolic genes has become a ubiquitous tool to mine biosynthetic gene clusters, which encode a chain of enzymes catalyzing specialized metabolites.⁴ Based on the information of biosynthetic genes, metabolic engineering enables the reconstitution of a metabolic route for their mass production via heterologous biosynthesis using microorganisms or plants. The catalytic promiscuity and regioselectivity of enzymes involved in producing desirable but intractable compounds are supported with more evidence from molecular docking, site-directed mutagenesis experiments, and crystal structure, as well as molecular mechanics calculation. Efficiently identifying genes encoding enzymes involved in the core catalysis of the pharmaceutical compound biosynthesis process, performing high-throughput screening of high-yielding strains, and uncovering catalytic mechanisms of core enzymes will be the challenges in future research.

Pharmaceutical compounds are often constitutively synthesized in specific tissues or even distributed in specialized cells; they could be generated by external stimulation in response to the changing and variable environments. Most of the studies regarding pharmaceutical compounds so far focus on control at the transcriptional level through integrating developmental and environmental cues. TFs recognized as the best elucidated regulators play a pivotal role in controlling the spatiotemporal regulation of metabolic pathways.⁵ The best-characterized TFs from three model medicinal plants, *Artemisa annua*, *Catharanthus roseus*, and *Salvia miltiorrhiza*, include the R2-R3 MYB, basic helix-loop-helix, APETALA2/ETHYLENE RESPONSE FACTOR (AP2/ERF), WRKY, and bZIP families, activating or repressing phytopharmaceutical compounds. The epigenomic regulation of pharmaceutical compounds in specialized tissues or at the single-cell level, including DNA methylation, non-coding RNA, histone modification, chromatin accessibility, and three-dimensional genome, will be extensively studied in medicinal plants in the future. Likewise, proteomics and mass spectrometry evidence has documented that those enzymes and regulators undergo extensive and dynamic post-transcriptional modifications (PTMs) to regulate metabolic processes. Proteins undergo various PTMs through the addition of small molecules, including acetylation, methylation, phosphorylation, ubiquitination, sumoylation, and glycosylation, which can alter their stability, localization, conformation, and interacting partners. The significance of PTMs for dynamically modulating the catalytic power of enzymes and activity of regulators in secondary metabolism will be validated in medicinal plants.

REFERENCES

- Hu, H., Shen, X., Liao, B., et al. (2019). Herbgenomics: a stepping stone for research into herbal medicine. *Sci. China Life Sci.* **62**, 913–920.
- Liao, B., Shen, X., Xiang, L., et al. (2022). Allele-aware chromosome-level genome assembly of *Artemisa annua* reveals the correlation between ADS expansion and artemisinin yield. *Mol. Plant* **15**, 1310–1328.
- Rai, A., Yamazaki, M., and Saito, K. (2019). A new era in plant functional genomics. *Curr. Opin. Syst. Biol.* **15**, 58–67.
- Polturak, G., and Osbourn, A. (2021). The emerging role of biosynthetic gene clusters in plant defense and plant interactions. *PLoS Pathog.* **17**, e1009698.
- Lacchini, E., and Goossens, A. (2020). Combinatorial control of plant specialized metabolism: mechanisms, functions, and consequences. *Annu. Rev. Cell Dev. Biol.* **36**, 291–313.

ACKNOWLEDGMENTS

The Herbgenomics Project was supported by the Scientific and Technological Innovation Project of the China Academy of Chinese Medical Sciences (C12021A03710).

DECLARATION OF INTERESTS

The authors declare no competing interests.