



OPEN ACCESS

EDITED BY

Rajesh Kumar Pathak,
Chung-Ang University, Republic of Korea

REVIEWED BY

Sutanu Nandi,
University of Colorado Anschutz Medical
Campus, United States

*CORRESPONDENCE

Cheng Song

✉ lanniao812329218@163.com

Yunpeng Cao

✉ xfcypeng@126.com

RECEIVED 02 August 2025

ACCEPTED 19 September 2025

PUBLISHED 01 October 2025

CITATION

Song C, Sabir IA, Zhao W and Cao Y (2025)
Decoding the mystery: AI-assisted
bioinformatics and functional genomics
technologies in medicinal plants.
Front. Plant Sci. 16:1678483.
doi: 10.3389/fpls.2025.1678483

COPYRIGHT

© 2025 Song, Sabir, Zhao and Cao. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License](#)
(CC BY). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Decoding the mystery: AI-assisted bioinformatics and functional genomics technologies in medicinal plants

Cheng Song^{1*}, Irfan Ali Sabir², Wanli Zhao³ and Yunpeng Cao^{4*}

¹College of Biological and Pharmaceutical Engineering, West Anhui University, Luan, China, ²Key Laboratory of Biology and Genetic Improvement of Horticultural Crops, Ministry of Agriculture and Rural Affairs, College of Horticulture, South China Agricultural University, Guangzhou, China, ³Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, China, ⁴State Key Laboratory of Plant Diversity and Specialty Crops, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, China

KEYWORDS

bioinformatics, functional genomics, artificial intelligence - AI, machine learning, large language model

Introduction

For millennia, medicinal plants have been a cornerstone of human healthcare, providing a rich source of bioactive compounds used in both traditional and modern medicine. A diverse array of therapeutic molecules is offered by these plants, from the antimalarial artemisinin in *Artemisia annua* to the anticancer alkaloids in *Catharanthus roseus*. The integration of artificial intelligence (AI) with bioinformatics and functional genomics has revolutionized the study of these medicinal plants, enabling researchers to explore their genetic and molecular underpinnings with unprecedented accuracy. These integrated technologies are transforming the study of medicinal plants, including drug discovery, responses to abiotic stresses, and the therapeutic potential of sustainable healthcare. However, the complexity and volume of genomic data pose significant challenges, necessitating advanced computational tools. AI, incorporating machine learning (ML) and deep learning (DL) techniques, has emerged as a powerful solution, capable of processing large volumes of data, identifying patterns and making predictions that traditional methods cannot match. This opinion explores several areas in which AI models in bioinformatics and functional genomics analysis are transforming medicinal plant research. Through detailed discussions and an exploration of future trends, we highlight how AI is reshaping our approach to medicinal plants, offering new possibilities for drug development and sustainable agriculture.

ML is considered a core technology in AI. Standard ML methods are overly narrow in their application to complex, natural, and high-dimensional raw data like genomic data. In contrast, DL methods are a promising and exciting area currently being widely applied in genomics, with successful applications in image recognition, audio classification, natural language processing, online web tools, chatbots, and robotics (Alharbi and Rashid, 2022). In this regard, DL as a genomics method is well-suited for analyzing large amounts of data. Although DL is still in its infancy in genomics, it holds the potential to transform fields such

as clinical genetics and functional genomics. Multiple genomic fields are leveraging the generation of high-throughput data and harnessing the power of deep learning algorithms to make complex predictions. Modern advances in DNA/RNA sequencing technologies and machine learning algorithms, particularly deep learning, have opened up a new chapter in research, enabling the translation of large biological datasets into new knowledge and discoveries across various subfields of genomics (Lee, 2023). In the field of next-generation sequencing, modern deep learning tools have been proposed to overcome the limitations of traditional interpretation pipelines (Alharbi and Rashid, 2022). It has demonstrated that combining the deep learning-based variant caller DeepVariant with traditional variant callers (such as SAMtools and GATK) can improve the accuracy scores of single-nucleotide variant and indel detection (Kumaran et al., 2019). DeepVariant relies on graphical differences in input images to perform the classification task of genetic variant calling from NGS short reads (Hall et al., 2024). It treats mapped sequencing datasets as images and transforms variant calling into an image classification task.

Functional genomics aims to reveal the roles of genes and their interactions in biological systems (Zhang Y, et al., 2025). Traditional methods, such as gene set enrichment analysis, rely on existing genomic databases and are relatively cumbersome and time-consuming. However, many intriguing biological questions often exceed the limitations of these databases, and the introduction of AI offers new possibilities for filling these gaps. AI is reshaping the traditional way genomics research is conducted. By utilizing large language models (LLMs), scientists can significantly reduce manual analysis time and rapidly identify gene functions and interactions (Lotter et al., 2024). AI systems can quickly examine vast volumes of genomic data in drug discovery to find biomarkers and gene mutations linked to disease. This accelerates the development of new drugs and increases the success rate of drug discovery. For example, AI can screen thousands of compounds within hours to identify the most likely effective drug candidates. PDGrapher can identify the multiple factors that contribute to disease in cells and predict treatment options that can restore healthy cell function. Focusing on multiple pathogenic drivers, PDGrapher can identify the genes most likely to transform diseased cells into a healthy state and recommend the best single or combination therapeutic targets. Results indicated that the tool not only accurately predicted known effective drug targets but also discovered several new potential candidates (Gonzalez et al., 2025). Compared to similar models, PDGrapher achieved 35% higher predictive accuracy and operated up to 25 times faster.

Genome annotation involves identifying genes and their functions within a genome. It is a critical step in understanding the genetic basis of the therapeutic properties of medicinal plants. Traditional annotation methods, which rely on sequence similarity to known genes, can be labor-intensive and ineffective when dealing with novel or divergent paralogs, which are prevalent in plant genomes. However, AI has introduced innovative solutions that use machine learning algorithms, such as support vector machines (SVMs) and Bayesian methods, to predict gene functions based

on sequence features and expression patterns. For instance, SVMs have been employed to identify drought-resistance genes in *Arabidopsis thaliana*, establishing a model for analogous applications in medicinal plants (Murmu et al., 2024). A significant advancement in this field is the application of deep learning to predict protein structures. Developed by DeepMind, AlphaFold 2 has achieved remarkable accuracy in predicting protein structures from amino acid sequences, thereby transforming functional genomics (Jumper et al., 2021; McCall and Almudevar, 2012). In *Salvia miltiorrhiza*, the structures of key enzymes involved in tanshinone biosynthesis were predicted, which helped the rational design of enzymes to enhance the production of these cardiovascular disease-protecting compounds (Chang et al., 2019; Zhou et al., 2017). Similarly, the homology-based gene prediction has been used to identify genes involved in withanolide biosynthesis, which are key adaptogenic compounds (Agarwal et al., 2017; Hakim et al., 2025). AI is also advancing single-cell genomics, enabling the study of gene expression at the cellular level. Tools like SIMLR (Single-cell Interpretation via Multi-kernel Learning) address challenges such as low-coverage single-cell RNA sequencing data, facilitating the clustering and annotation of rare cell types (Wang et al., 2018). In *C. roseus*, some bioinformatic tools were applied to annotated genes involved in terpenoid indole alkaloid (TIA) biosynthesis, thereby enhancing our understanding of tissue-specific expression (Rai et al., 2022). Despite these advancements, challenges persist. Many medicinal plants have large, complex genomes, and comprehensive genomic data for rare species is often lacking. The interpretability of DL models also poses a hurdle, as understanding their predictions is crucial for gaining biological insights. Ongoing efforts to develop standardized datasets and understandable DL models are addressing these issues, and these efforts are promising to expand the application in genome annotation for medicinal plants.

Metabolic pathways are central to the production of secondary metabolites in medicinal plants. These are often responsible for their therapeutic properties. Reconstructing these pathways is essential for understanding biosynthesis and for engineering plants to produce more compounds (Song et al., 2022a). Bioinformatics and genomics have transformed this process by combining metabolomics data with sophisticated computational methods. Machine learning algorithms predict metabolic pathways by analyzing metabolite concentrations and gene expression patterns. For instance, metabolic engineering helps to reconstruct the artemisinin biosynthetic pathway in *A. annua*, identifying key genes and enzymes, thereby informing strategies to increase artemisinin yields (Costello and Martin, 2018). Gene mining is the process of identifying genes of interest from genomic data. This is another area where AI excels. ML models classify genes based on sequence and expression data, pinpointing those involved in metabolite production. In *Panax ginseng*, the glycosyltransferases (UGTs) and CYP450 family genes responsible for ginsenoside production, paving the way for genetic engineering to boost ginsenoside content (Hou et al., 2021; Xu et al., 2017). Similarly, large-scale gene mining in *C. roseus* genome has shed light on the biosynthesis of TIAs, which are vital anti-cancer agents (McCall and

Almudevar, 2012). AI facilitates the discovery of novel pathways. By analyzing multi-omics datasets, we can predict pathways that are not apparent through traditional methods, particularly in understudied plants. In *Ophiorrhiza pumila*, some key genes involved in camptothecin biosynthesis were identified by integrating transcriptomic and metabolomic data (Yang et al., 2021). Tools such as ClusterFinder and DeepBGC use hidden Markov models (HMMs) and DL method to identify biosynthetic gene clusters (BGCs), which are essential for producing secondary metabolites (Liu et al., 2022; Hannigan et al., 2019). Genome-wide identification of WRKY members from *Myrica rubra* revealed that the WRKY14 significantly activates the promoter region of the SWEET1 gene, suggesting its positive regulatory role in sugar synthesis (Fan et al., 2025). These advancements would have a lasting effect on drug discovery and agricultural biotechnology by enabling targeted genetic modifications to optimize the production of therapeutic compounds. However, challenges such as data scarcity for rare plants.

Integrated multi-omics data — including genomics, transcriptomics, proteomics, and metabolomics — provides a comprehensive view of plant biology (Song et al., 2022b; Zhang et al., 2023). Large language model facilitates this process by managing the complexity and volume of the data. The orthogonal projections to latent structures (OPLS) method can integrate transcriptomic and metabolomic data, and tools such as iDREM can construct integrated networks from temporal data (Kumar et al., 2024). In *S. lycopersicum*, multi-omics integration has optimized metabolic networks to improve fruit quality (Cembrowska-Lech et al., 2023). The optimization of metabolic networks involves predicting and manipulating pathways to increase the yield of therapeutic compounds. Challenges such as data noise, sparsity and scaling issues are being overcome. This is because of its ability to handle high-dimensional data. Predicting gene regulatory networks (GRNs) is essential for understanding how genes are regulated in response to environmental and developmental cues. AI, particularly neural network-based methods, predicts transcription factor binding sites and regulatory relationships. In *C. roseus*, AI has been used to predict networks involved in TIA biosynthesis and identify key regulators (Pan et al., 2016). Transformer-based models are used by tools like Enformer and RNABERT to predict genome interactions and RNA clustering, respectively (Avsec et al., 2021). These advancements facilitate the identification of new therapeutic targets and pathways, enhancing the potential for genetic engineering in medicinal plants. However, genetic modification raises ethical concerns, requiring careful assessment of ecological impacts.

GRNs govern gene expression in response to environmental and developmental signals. Advances in AI have led to the development of tools such as iDREM and GRNBoost2, which can construct temporal and cell-specific GRNs from multi-omics data (Sharma et al., 2024). These tools have been used to study stress responses in *Arabidopsis*, revealing complex regulatory mechanisms. In medicinal plants, predicting GRNs is crucial for understanding how therapeutic compounds are produced. For example, AI has been employed to predict the regulatory

networks involved in TIA biosynthesis in *C. roseus*, identifying the transcription factors that control alkaloid production. Transfer learning has also enabled cross-species predictions, such as the identification of metabolism-related genes in *S. lycopersicum* (Badia-i-Mompel et al., 2023). In *Withania somnifera*, genome-wide identification has identified stress-responsive genes involved in withanolide biosynthesis, thereby enhancing plant resilience and compound yield (Nicolis et al., 2024; Tripathi et al., 2020). Gene co-expression network analysis is particularly valuable for identifying stress-related genes, as many secondary metabolites are produced in response to environmental stresses. By analyzing gene expression under various conditions, large language models can classify genes based on their stress responsiveness. This provides targets for breeding stress-tolerant medicinal plants. The complexity of GRNs and the need for comprehensive multi-omics data are just two of the challenges that must be overcome (Badia-i-Mompel et al., 2023; Otal et al., 2025). Using more sophisticated bioinformatics and data integration techniques is helping to resolve these issues and make predictions more accurate (Song et al., 2023; Zhang J, et al., 2025).

Discussion

Despite the immense success of these tools in genomics and bioinformatics, the adoption of different DL solutions and models remains limited. One reason is the lack of published DL-based protocols that can adapt to new, heterogeneous datasets that require extensive data engineering. In genomics, high-throughput data are used to train neural networks and have become a typical approach for disease prediction or understanding regulatory genomics (Schmidt and Hildebrandt, 2021). Similarly, developing new DL models and testing existing models on new datasets are significant challenges due to the lack of comprehensive, generalizable, and practical biology-oriented deep learning libraries (Munappy et al., 2022). In this regard, software frameworks and genomic packages are crucial for quickly adopting new research questions or hypotheses, integrating raw data, or conducting research using different neural network architectures. Recently, advances in NLP and LLMs have improved data integration and analysis. GRNs can address data scarcity by generating synthetic datasets, and attention mechanisms can enhance model interpretability. Future breakthroughs will depend on interdisciplinary collaboration between biologists, computer scientists, and data scientists. Despite significant advancements, challenges remain in applying AI to medicinal plant research. Standardized datasets that include genomic, transcriptomic, proteomic, metabolomic and phenotypic data are essential for training robust AI models. A range of resources on spice genomics have been developed to help identify the most promising future directions. These resources include genome assemblies, sequencing and re-sequencing projects, as well as studies based on the transcriptome, non-coding RNA-mediated regulation, organelles-based resources, developed molecular markers, web resources, databases and AI-directed resources (Das et al., 2023). All of these are focused on enhancing

the breeding potential of specific spices. While there are extensive datasets for model plants, many medicinal plants still lack sufficient genomic resources, which limits AI applications. Although deep learning models are highly accurate, they often operate like black boxes, hindering the translation of predictions into biological insights. Therefore, developing explainable AI models is crucial for gaining trust and extracting actionable biological insights. Additionally, the substantial computational resources required for genome-wide identification analyses present a challenge for researchers in settings with limited resources. One solution is to develop lightweight AI models for use in such environments, as well as using GRNs to create gene expression data for model training and integrating attention mechanisms to focus on biologically relevant features.

Author contributions

CS: Conceptualization, Funding acquisition, Project administration, Supervision, Writing – original draft, Writing – review & editing. IS: Supervision, Writing – original draft, Writing – review & editing. WZ: Supervision, Writing – original draft, Writing – review & editing. YC: Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. This work was supported by Quality Engineering Project of Anhui Province (2024zybj032), Quality Engineering Project of West Anhui University (wxxy2024011), and Development of Big Data Integration and

Analysis Platform for Traditional Chinese Medicine Genomics (0045025050).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Agarwal, A. V., Gupta, P., Singh, D., Dhar, Y. V., Chandra, D., and Trivedi, P. K. (2017). Comprehensive assessment of the genes involved in withanolide biosynthesis from *Withania somnifera*: chemotype-specific and elicitor-responsive expression. *Funct. Integr. Genomics* 17, 477–490. doi: 10.1007/s10142-017-0548-x
- Alharbi, W. S., and Rashid, M. (2022). A review of deep learning applications in human genomics using next-generation sequencing data. *Hum. Genomics* 16, 26. doi: 10.1186/s40246-022-00396-x
- Avsec, Ž., Agarwal, V., Visentin, D., Ledsam, J. R., Grabska-Barwinska, A., Taylor, K. R., et al. (2021). Effective gene expression prediction from sequence by integrating long-range interactions. *Nat. Methods* 18, 1196–1203. doi: 10.1038/s41592-021-01252-x
- Badia-i-Mompel, P., Wessels, L., Müller-Dott, S., Trimbou, R., Ramirez Flores, R. O., Argelaguet, R., et al. (2023). Gene regulatory network inference in the era of single-cell multi-omics. *Nat. Rev. Genet.* 24, 739–754. doi: 10.1038/s41576-023-00618-5
- Cembrowska-Lech, D., Krzemińska, A., Miller, T., Nowakowska, A., Adamski, C., Radaczynska, M., et al. (2023). An integrated multi-omics and artificial intelligence framework for advance plant phenotyping in horticulture. *Biology* 12, 1298. doi: 10.3390/biology12101298
- Chang, Y., Wang, M., Li, J., and Lu, S. (2019). Transcriptomic analysis reveals potential genes involved in tanshinone biosynthesis in *Salvia miltiorrhiza*. *Sci. Rep.* 9, 14929. doi: 10.1038/s41598-019-51535-9
- Costello, Z., and Martin, H. G. (2018). A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data. *NPJ Syst. Biol. Appl.* 4, 19. doi: 10.1038/s41540-018-0054-3
- Das, P., Chandra, T., Negi, A., Jaiswal, S., Iquebal, M. A., Rai, A., et al. (2023). A comprehensive review on genomic resources in medicinally and industrially important major spices for future breeding programs: Status, utility and challenges. *Curr. Res. Food Sci* 7, 100579. doi: 10.1016/j.crfs.2023.100579
- Fan, X., Chen, M., Zhang, H., Liu, Y., Yang, M., Ye, C., et al. (2025). Systematic identification and analysis of WRKY transcription factors reveals the role of MrWRKY14 in *Myrica rubra*. *Front. Plant Sci.* 16. doi: 10.3389/fpls.2025.1602750
- Gonzalez, G., Lin, X., Herath, I., Veselkov, K., Bronstein, M., and Zitnik, M. (2025). Combinatorial prediction of therapeutic perturbations using causally inspired neural networks. *Nat. Biomed. Eng.* doi: 10.1038/s41551-025-01481-x
- Hakim, S. E., Choudhary, N., Malhotra, K., Peng, J., Bültmeier, A., Arafa, A., et al. (2025). Phylogenomics and metabolic engineering reveal a conserved gene cluster in Solanaceae plants for withanolide biosynthesis. *Nat. Commun.* 16, 6367. doi: 10.1038/s41467-025-61686-1
- Hall, M. B., Wick, R. R., Judd, L. M., Nguyen, A. N., Steinig, E. J., Xie, O., et al. (2024). Benchmarking reveals superiority of deep learning variant callers on bacterial nanopore sequence data. *Elife* 13, RP98300. doi: 10.7554/eLife.98300
- Hannigan, G. D., Prihoda, D., Palicka, A., Soukup, J., Klempir, O., Rampula, L., et al. (2019). A deep learning genome-mining strategy for biosynthetic gene cluster prediction. *Nucleic Acids Res.* 47, e110–e110. doi: 10.1093/nar/gkz654
- Hou, M., Wang, R., Zhao, S., and Wang, Z. (2021). Ginsenosides in *Panax* genus and their biosynthesis. *Acta Pharm. Sin. B* 11, 1813–1834. doi: 10.1016/j.apsb.2020.12.017

- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi: 10.1038/s41586-021-03819-2
- Kumar, Y., Marchena, J., Awla, A. H., Li, J. J., and Abdalla, H. B. (2024). The AI-powered evolution of big data. *Appl. Sci.* 14, 10176. doi: 10.3390/app142210176
- Kumaran, M., Subramanian, U., and Devarajan, B. (2019). Performance assessment of variant calling pipelines using human whole exome sequencing and simulated data. *BMC Bioinf.* 20, 342. doi: 10.1186/s12859-019-2928-9
- Lee, J.-Y. (2023). The principles and applications of high-throughput sequencing technologies. *Dev. Reprod.* 27, 9–24. doi: 10.12717/DR.2023.27.1.9
- Liu, M., Li, Y., and Li, H. (2022). Deep learning to predict the biosynthetic gene clusters in bacterial genomes. *J. Mol. Biol.* 434, 167597. doi: 10.1016/j.jmb.2022.167597
- Lotter, W., Hassett, M. J., Schultz, N., Kehl, K. L., Van Allen, E. M., and Cerami, E. (2024). Artificial intelligence in oncology: current landscape, challenges, and future directions. *Cancer Discov.* 14, 711–726. doi: 10.1158/2159-8290.CD-23-1199
- McCall, M. N., and Almudevar, A. (2012). Affymetrix GeneChip microarray preprocessing for multivariate analyses. *Briefings Bioinf.* 13, 536–546. doi: 10.1093/bib/bbr072
- Munappy, A. R., Bosch, J., Olsson, H. H., Arpteg, A., and Brinne, B. (2022). Data management for production quality deep learning models: Challenges and solutions. *J. Syst. Software* 191, 111359. doi: 10.1016/j.jss.2022.111359
- Murmu, S., Sinha, D., Chaurasia, H., Sharma, S., Das, R., Jha, G. K., et al. (2024). A review of artificial intelligence-assisted omics techniques in plant defense: current trends and future directions. *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1292054
- Nicolis, V. F., Burger, N. F. V., Parshoham, R., Bierman, A., Rai, P. S., Muthusamy, A., et al. (2024). Differential gene expression analysis of *Withania somnifera* in response to salinity stress. *Research Square* 1–24. doi: 10.21203/rs.3.rs-4521592/v1
- Otal, H. T., Subasi, A., Kurt, F., Canbaz, M. A., and Uzun, Y. (2025). “Analysis of gene regulatory networks from gene expression using graph neural networks,” in *Digital Healthcare, Digital Transformation and Citizen Empowerment in Asia-Pacific and Europe for a Healthier Society* (Los Alamos, USA: Elsevier), 249–270. doi: 10.1016/B978-0-443-30168-1.00011-6
- Pan, Q., Mustafa, N. R., Tang, K., Choi, Y. H., and Verpoorte, R. (2016). Monoterpenoid indole alkaloids biosynthesis and its regulation in *Catharanthus roseus*: a literature review from genes to metabolites. *Phytochem. Rev.* 15, 221–250. doi: 10.1007/s11101-015-9406-4
- Rai, S. K., Rai, K. K., Apoorva, Kumar, S., and Rai, S. P. (2022). “Functional Genomics Approaches for Gene Discovery Related to Terpenoid Indole Alkaloid Biosynthetic Pathway in *Catharanthus roseus*,” in *The Catharanthus Genome*. Ed. C. Kole (Springer International Publishing, Cham), 155–173. doi: 10.1007/978-3-030-89269-2_9
- Schmidt, B., and Hildebrandt, A. (2021). Deep learning in next-generation sequencing. *Drug Discov. Today* 26, 173–180. doi: 10.1016/j.drudis.2020.10.002
- Sharma, A., Lysenko, A., Jia, S., Boroevich, K. A., and Tsunoda, T. (2024). Advances in AI and machine learning for predictive medicine. *J. Hum. Genet.* 69, 487–497. doi: 10.1038/s10038-024-01231-y
- Song, C., Ma, J., Li, G., Pan, H., Zhu, Y., Jin, Q., et al. (2022a). Natural composition and biosynthetic pathways of alkaloids in medicinal dendrobium species. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.850949
- Song, C., Wang, Y., Manzoor, M. A., Mao, D., Wei, P., Cao, Y., et al. (2022b). In-depth analysis of genomes and functional genomics of orchid using cutting-edge high-throughput sequencing. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1018029
- Song, C., Zhang, Y., Zhang, Y., Yi, S., Pan, H., Liao, R., et al. (2023). Genome sequencing-based transcriptomic analysis reveals novel genes in *Peucedanum praeruptorum*. *BMC Genom. Data* 24, 19. doi: 10.1186/s12863-023-01157-y
- Tripathi, S., Srivastava, Y., Sangwan, R. S., and Sangwan, N. S. (2020). In silico mining and functional analysis of AP2/ERF gene in *Withania somnifera*. *Sci. Rep.* 10, 4877. doi: 10.1038/s41598-020-60090-7
- Wang, B., Ramazzotti, D., De Sano, L., Zhu, J., Pierson, E., and Batzoglu, S. (2018). SIMLR: A tool for large-scale genomic analyses by multi-kernel learning. *Proteomics* 18, 1700232. doi: 10.1002/pmic.201700232
- Xu, J., Chu, Y., Liao, B., Xiao, S., Yin, Q., Bai, R., et al. (2017). Panax ginseng genome examination for ginsenoside biosynthesis. *Gigascience* 6, 1–15. doi: 10.1093/gigascience/gix093
- Yang, M., Wang, Q., Liu, Y., Hao, X., Wang, C., Liang, Y., et al. (2021). Divergent camptothecin biosynthetic pathway in *Ophiorrhiza pumila*. *BMC Biol.* 19, 122. doi: 10.1186/s12915-021-01051-y
- Zhang, J., Yang, Y., Si, J., Chen, D., Dong, C., and Han, Z. (2025). Artificial intelligence in the discovery and modification of biological elements in medicinal plants. *M. P. B.* 4, e012. doi: 10.48130/mpb-0025-0010
- Zhang, Y., Zhang, W., Manzoor, M. A., Sabir, I. A., Zhang, P., Cao, Y., et al. (2023). Differential involvement of WRKY genes in abiotic stress tolerance of *Dendrobium huoshanense*. *Ind. Crops Products* 204, 117295. doi: 10.1016/j.indcrop.2023.117295
- Zhang, Y., Zhou, Z., Wang, M., Mao, X., Wang, Z., and Zou, Q. (2025). A multi-omics data integration framework for gene regulatory network inference based on contrastive learning. *IEEE Trans. Comput. Biol. Bioinform.* 22, 1095–1106. doi: 10.1109/TCBBIO.2025.3548953
- Zhou, W., Huang, Q., Wu, X., Zhou, Z., Ding, M., Shi, M., et al. (2017). Comprehensive transcriptome profiling of *Salvia miltiorrhiza* for discovery of genes associated with the biosynthesis of tanshinones and phenolic acids. *Sci. Rep.* 7, 10554. doi: 10.1038/s41598-017-10215-2