INTERNATIONAL JOURNAL OF
EPIDEMIOLOGY AND TROPICAL MEDICINE
( I J E T M )

# Genomic Selection for Combined Disease Resistance and Secondary Metabolite Quality in Medicinal Plants: Frameworks, Challenges, and a Strategic Roadmap

**Muhammad Tahsin[1], Maryam Alim[2], Muhammad Saqlan[3], Ali Murtaza*[3]**

*[1]Department of Biochemistry & Biotechnology, University of Gujrat, Pakistan*
*[2]Institute of Horticulture, University of Agriculture Faisalabad, Pakistan*
*[3]Department of Plant Breeding and Genetics, University of Agriculture Faisalabad, Pakistan*

*Corresponding author: Ali Murtaza, Department of Plant Breeding and Genetics, University of Agriculture Faisalabad, Pakistan. Email: mralimurtazarai@gmail.com*

**Article History**
Accepted: 05 September 2026
Published: 06 September 2026

**Keywords**
*disease resistance; genomic estimated breeding values; genomic selection; medicinal plants; metabolomics; multi-trait models; plant breeding; secondary metabolites;*



***Abstract***

***Background:*** *Medicinal plants represent an irreplaceable source of pharmacologically active secondary metabolites, yet breeding programs targeting these species face a structurally distinctive dual challenge: achieving simultaneous genetic improvement of secondary metabolite content while enhancing disease resistance and agronomic stability. These objectives are biologically non-independent, since terpenoids, alkaloids, and phenolics function both as pharmaceutical active ingredients and as components of the plant's chemical defense system.*

***Methods:*** *Genomic selection (GS), which uses genome-wide markers to estimate genomic estimated breeding values for all quantitative traits concurrently, offers a transformative framework for addressing this challenge.*

***Results:*** *Multi-trait GS models leverage genetic correlations between resistance and metabolite traits to improve prediction accuracy for both objectives simultaneously, and integrating metabolomic and transcriptomic data as covariates substantially enhances accuracy for complex, environmentally sensitive metabolite traits. Despite rapid advances in genomic resources for numerous medicinal species, no published GS program currently targets secondary metabolite quality in any major medicinal plant.*

***Conclusions:*** *Critical barriers include the absence of genotyped training populations, narrow genetic bases, prohibitive metabolite phenotyping costs, and substantial genotype-by-environment interactions.*

## INTRODUCTION

Medicinal plants are the biochemical foundation of traditional and modern pharmacology, producing terpenoids, alkaloids, and phenolics that underlie their pharmaceutical value; more than 1,300 species are documented in Europe alone, most still wild-harvested rather than cultivated [1]. Because pharmaceutical value depends directly on metabolite content, artemisinin in Artemisia annua, ginsenosides in Panax ginseng, tanshinones in Salvia miltiorrhiza [2,3,] breeding programs face a distinctive dual challenge: improving metabolite

quality while enhancing disease resistance and agronomic stability. These objectives can be positively correlated, since many defense compounds are also the pharmacologically active ingredients [5,6], or antagonistic, since metabolite production imposes metabolic costs that reduce growth and stress tolerance.

Conventional breeding and marker-assisted selection are poorly suited to this polygenic, quantitative genetic architecture [7,8]. Genomic selection (GS), first conceptualized by. Genomic resources for economically important medicinal species have expanded rapidly [10,11], yet no medicinal species has integrated these resources into a functioning GS program for either disease resistance or metabolite quality. This review synthesizes the theoretical and applied foundations of GS for this dual improvement challenge and proposes a strategic research roadmap for the field.

### *Genomic Selection Framework*

Prediction accuracy, the correlation between predicted GEBVs and true breeding values, determines genetic gain per cycle and depends on trait heritability, linkage disequilibrium (LD) structure, marker density, training population size, and genotype-by-environment (G\u00d7E) interactions. Genomic Best Linear Unbiased Prediction (GBLUP) estimates breeding values from a genomic relationship matrix and performs well for highly polygenic traits; Bayesian methods (BayesB, BayesC) accommodate variable marker-effect distributions, suiting traits with mixed large- and small-effect loci, a pattern plausible for metabolite pathways where structural genes may have large effects against a polygenic regulatory background [12]. Machine learning approaches capture non-additive and epistatic effects and have outperformed linear models for complex traits in crop species, though they require larger training populations to avoid overfitting [12,13].

### *Model Selection for Dual-Objective, High-Dimensional Prediction*

Metabolite traits in medicinal plants often show low-to-moderate heritability due to strong environmental modulation of pathway enzyme activity [3,14], meaning single-environment phenotyping poorly estimates genetic potential. Marker density requirements depend on LD decay, and genotyping-by-sequencing (GBS) offers a cost-effective option for species lacking reference genomes [15,7]. Training population size has a particularly strong effect on accuracy for low-heritability traits and distantly related selection candidates, conditions typical of medicinal plant programs [8]. Cross-validation should reflect the intended application: forward-time validation, using historical generations to predict later ones, best simulates practical breeding use [16]. A further constraint is that most medicinal species still lack a high-

quality reference genome, and even where a draft assembly exists, the heterozygosity and structural variation typical of outcrossing or polyploid taxa make single-reference alignment an unreliable basis for marker calling (Figure 1). For medicinal plant programs beginning without an assembled genome, GBS-derived or k-mer-based marker sets, revisited as pan-genome resources mature, represent a more realistic near-term entry point than waiting for a chromosome-scale reference.

**Table 1. Genomic Selection Prediction Accuracy and Model Characteristics for Disease Resistance and Metabolite Quality Traits in Crop Species Applicable to Medicinal Plant Breeding**

| Crop Species | Target Trait(s) | GS Model | Outcome / Relevance | Applicability to Medicinal Plants | Reference |
|---|---|---|---|---|---|
| Triticum aestivum (wheat) | Fusarium resistance + grain protein content | MT-GBLUP | Successful simultaneous dual-objective selection; moderate-high accuracy | Direct template for combined resistance-metabolite GS; demonstrates MT model viability | [17] |
| Camellia sinensis (tea) | Catechin/EGCG content + disease resistance | GBLUP | Closest published analogue to medicinal plant GS; perennial species with complex metabolism | Direct template for phenolic and secondary metabolite GS in medicinal species | [18] |
| Cannabis sativa | Cannabinoid profiles (THC, CBD) + disease resistance | GBLUP, RF | Emerging program; regulatory constraints limit progress | Models applicable to terpene and alkaloid quality traits in other species | [2] |
| Hordeum vulgare (barley) / Solanum lycopersicum (tomato) | Disease resistance + β-glucan content / quality metabolite traits | Metabolomics-augmented GBLUP / MT-GBLUP | Metabolomics covariates substantially increased prediction accuracy for quality traits | Demonstrates metabolomics-augmented GS feasibility for non-primary and complex metabolite targets | [19] |

Note. GS = genomic selection; MT-GBLUP = multi-trait genomic best linear unbiased prediction; RF = random forest; EGCG = epigallocatechin gallate; THC = tetrahydrocannabinol; CBD = cannabidiol.

### *Multi-Trait Genomic Selection for Combined Disease Resistance and Metabolite Quality*

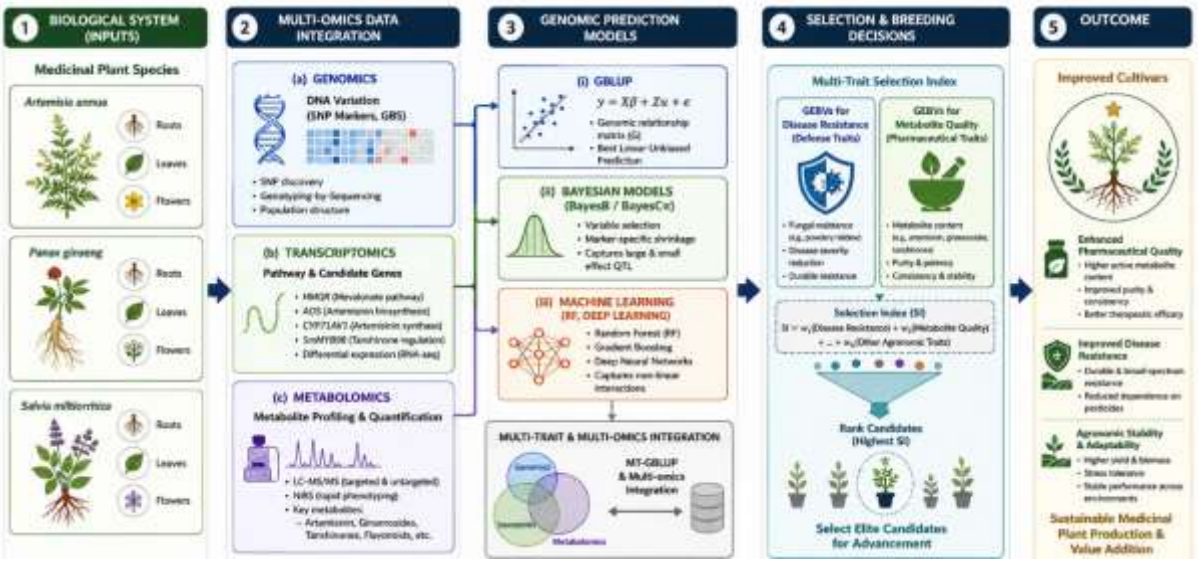
Multi-trait GBLUP (MT-GBLUP) models genetic and residual covariances across traits simultaneously, sharing information in proportion to genetic correlation. When disease resistance and metabolite content are positively correlated, MT-GBLUP improves prediction accuracy for both relative to single-trait models, particularly valuable when metabolite phenotyping is more costly than resistance screening [10]. Bayesian multi-trait models accommodate correlation uncertainty in the small training populations typical of medicinal plant programs, while selection indices weight trait GEBVs by economic importance [3]. No dual-objective GS program yet exists in medicinal plants, but instructive analogues exist in crop species. Wheat programs targeting Fusarium resistance and grain protein content confirm that MT-GBLUP outperforms single-trait prediction when traits are genetically correlated [17]. Tea (Camellia sinensis), where catechin/EGCG content parallels pharmaceutical metabolite

traits, offers the closest published analogue, demonstrating GS feasibility in a perennial species with complex secondary metabolism [18]. Emerging efforts in Cannabis sativa targeting cannabinoid profiles alongside disease resistance illustrate pharmaceutical-quality-focused breeding, constrained thus far by regulatory limits on phenotypic data [2].

### *Resolving the Growth-Defense Trade-off: Multi-Trait Indices and Pareto Optimization*

Disease resistance and secondary metabolite accumulation do not always align: induced and constitutive defense responses draw on shared carbon and nitrogen pools, and resistance mechanisms can impose a physiological growth-defense trade-off that suppresses constitutive metabolite synthesis even when the same jasmonate-responsive pathways nominally connect the two traits. Multi-trait selection indices that simply sum GEBVs weighted by economic value implicitly assume a linear, unconstrained trade-off and can therefore select against one objective when correlations are negative or curvilinear.

**Figure 1. Genomic Selection Prediction Accuracy for Disease Resistance and Metabolite Quality Traits in Crop Species Applicable to Medicinal Plant Breeding. This figure has been created by AI-assisted tools for visualization only.**




Two complementary approaches are better suited to negative or uncertain correlations. First, MT-GBLUP or Bayesian multi-trait models can be reparameterized with explicit priors on the genetic covariance between resistance and metabolite traits, allowing the model to shrink the estimated correlation toward an informative prior when the training population is too small to estimate it precisely, rather than defaulting to a naive independence assumption; this is particularly relevant given that, as discussed in Section 9, no medicinal species yet has a direct genetic-correlation estimate between these trait classes. Second, multi-objective (Pareto) optimization treats resistance and metabolite content as separate objectives and identifies the

Pareto frontier of genotypes for which no other genotype improves one trait without worsening the other, rather than collapsing both traits into a single weighted index; breeders can then select along the frontier according to the acceptable trade-off for a given end use, for example prioritizing metabolite yield for compounds with a narrow safety margin, or prioritizing resistance where phytosanitary failure would eliminate a harvest entirely.

### *Omics-Augmented Genomic Prediction for Secondary Metabolite Traits*

Metabolite phenotypes integrate genetic, developmental, and environmental variation, so SNP-only prediction captures only part of this complexity. Metabolomics-augmented GS treats metabolite profiles (LC-MS, NMR) as intermediate phenotypes that add predictive information beyond marker data; incorporating metabolite covariates has improved accuracy for quality traits in tomato and barley [19]. Analytical infrastructure already used for medicinal plant quality control, NIRS, HPLC-DAD, LC-MS/MS, is progressively being standardized for compounds such as artemisinin, ginsenosides, and tanshinones, creating reusable data for GS model training [16].

Transcriptomic selection incorporates biosynthetic pathway gene expression as predictive covariates. In Artemisia annua, transcript abundance of genes including HMGR, ADS, CYP71AV1, and DBR2 correlates with artemisinin content [20], and the discovery that SmMYB98 coordinately regulates tanshinone and salvianolic acid biosynthesis in Salvia miltiorrhiza [21] identifies master regulators of particularly high predictive value. Multi-omics models integrating SNP, metabolome, and transcriptome data in machine learning architectures represent the most information-rich approach [13], though the added training population and computational demands must be weighed against realistic resources.

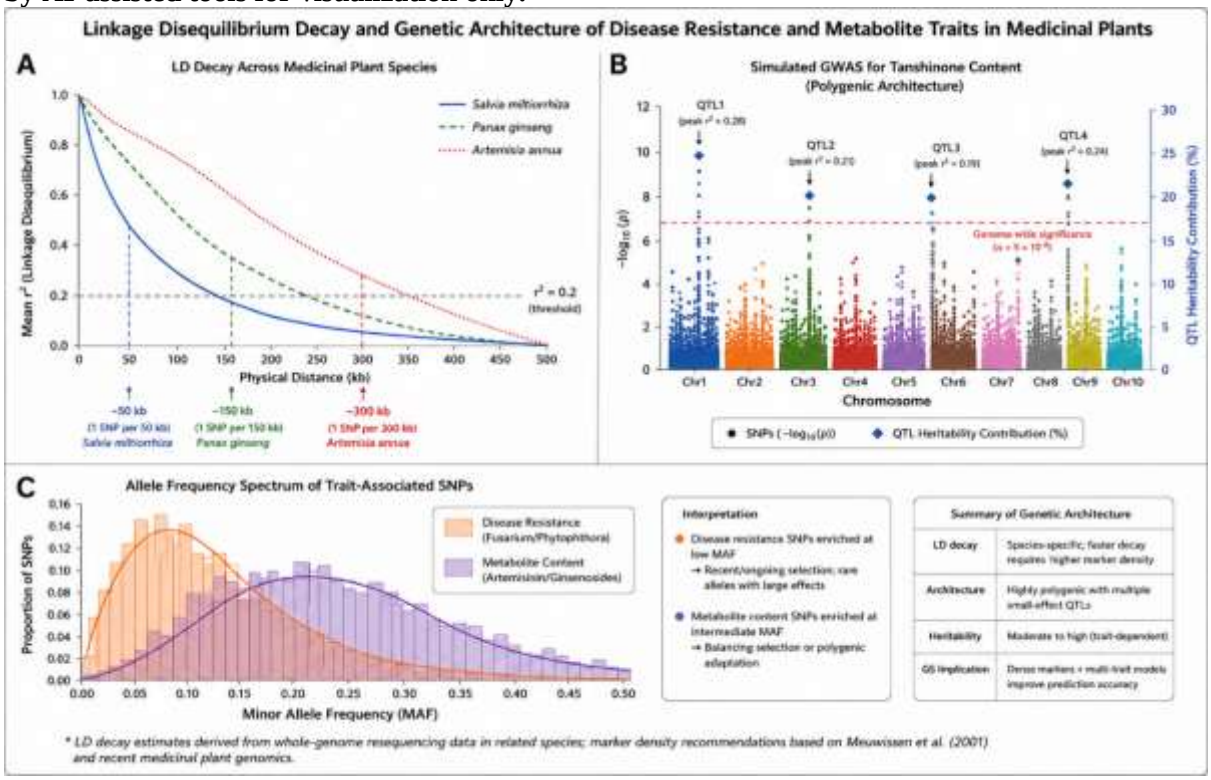
### *Genetic Architecture of Target Traits and Their Correlations*

Quantitative disease resistance is typically polygenic, involving many loci of small effect that together confer broader-spectrum, more durable protection than single major-gene resistance [22], an architecture GS is specifically designed to capture. Genetic control of metabolite accumulation varies widely: artemisinin content is a low-heritability, environment-sensitive trait, strongly induced by jasmonate signaling and light [20,3]; ginsenoside profiles show moderate heritability with identified major QTLs, and CRISPR knockout of protopanaxadiol 6-hydroxylase confirms the genetic tractability of specific pathway genes [23,24]; tanshinone content in Salvia miltiorrhiza reflects expansion of the CYP71D gene subfamily together with the SmABCG1 transporter that mediates root export [4,11], indicating that prediction models spanning the full pathway should outperform models focused on

structural genes alone (Figure 2). Terpene synthase gene family expansion is a major source of terpenoid diversity across species, and root-specific sesterterpene synthases in Arabidopsis directly influence root microbiota assembly, linking terpene diversity to disease-relevant plant-microbe interactions [25].

The biological basis (Table 2) for positive genetic correlations between metabolite content and resistance lies in shared defense-related biosynthesis, since jasmonate signaling that induces artemisinin biosynthesis also orchestrates broader defense responses [20]. Conversely, biosynthesis competes with primary metabolism for carbon, and constitutive terpene overexpression has caused growth retardation in Arabidopsis [26], showing that trade-offs are also possible. Correlation direction must therefore be empirically characterized in each target species before designing multi-trait selection indices.

**Figure 2. Conceptual Framework for Multi-Trait Genomic Selection Integrating Disease Resistance and Secondary Metabolite Quality in Medicinal Plants. This figure has been created by AI-assisted tools for visualization only.**




### *Training Population Design, Phenotyping, and Implementation Barriers*

Training population design, genotyped and comprehensively phenotyped individuals from which GS models are trained, is the single most consequential decision in program development and remains the field's principal barrier: no medicinal species yet has a training population combining genome-wide genotyping with multi-environment phenotypic data for both metabolite content and disease resistance [10,7]. Simulations suggest 500 to 1,000

individuals give moderate accuracy for moderate-heritability traits, while 1,500 or more may be needed for low-heritability traits such as artemisinin content [8]. Because most medicinal species lack germplasm at this scale, diverse accessions, landraces, and wild relatives must be strategically incorporated; iterative genomic-assisted phenotyping, prioritizing the most genetically informative individuals for costly deep phenotyping, helps allocate limited resources efficiently [27].

**Table 2. Priority Medicinal Plant Species for Genomic Selection Program Development: Genomic and Phenotyping Readiness Assessment**

| Species | Key Pharmaceutical Metabolite | Major Disease Threats | Genome Available? | SNP/Genotyping Resources | Training Pop. Feasibility | GS Priority |
|---|---|---|---|---|---|---|
| Artemisia annua | Artemisinin (antimalarial) | Fungal root rot, Alternaria | Yes (draft) | Developing; GBS feasible | Moderate; diverse accessions available | High |
| Panax ginseng / P. quinquefolius | Ginsenosides (adaptogenic, anti-tumor) | Phytophthora, Cylindrocarpon | Yes (P. ginseng) | Expanding; major QTLs identified | Moderate; long generation time | High |
| Salvia miltiorrhiza | Tanshinones, salvianolic acids (cardiovascular) | Root rot, viral diseases | Yes | Well-developed; CYP and TPS genes characterized | Moderate-high; transformation systems established | High |
| Catharanthus roseus | Vinblastine, vincristine (anti-cancer) | Phytoplasma, Fusarium | Yes | Developing; alkaloid pathway mapped | Moderate; existing diversity collections | High |
| Camellia sinensis (medicinal use) | Catechins, EGCG (antioxidant, anticancer) | Blister blight, grey blight | Yes | Well-developed; GS models under development | High; breeding programs active | Very High (analogue model) |
| Cannabis sativa (medicinal) | CBD, THC, terpene profiles (neurology, pain) | Botrytis, powdery mildew | Yes (high-quality) | High; genomics advancing rapidly | High; significant industry investment | High |

Note. GBS = genotyping-by-sequencing; SNP = single nucleotide polymorphism; QTL = quantitative trait locus; CYP = cytochrome P450; TPS = terpene synthase; EGCG = epigallocatechin gallate; CBD = cannabidiol; THC = tetrahydrocannabinol. GS priority reflects the combination of pharmaceutical importance, available genomic resources, and feasibility of training population development.

Disease phenotyping requires standardized inoculation and multi-environment trials, since single-environment GEBVs generalize poorly given strong G\u00d7E interactions [7,28]. Metabolite phenotyping must further specify tissue, developmental stage, and harvest conditions, since compounds can be transported between organs or exported by transporters such as SmABCG1 in Salvia miltiorrhiza [11]. Speed breeding, using extended photoperiods and optimized environmental conditions, can compress annual crop generation intervals

several-fold [7]; early generation selection using GEBVs from seedling genotypes allows advancement decisions before full phenotypic expression, particularly valuable where metabolite accumulation occurs late in development [29]. Doubled haploid technology can further compress inbreeding, and the establishment of protoplast regeneration systems in Salvia miltiorrhiza [30] provides a foundation for future development in medicinal species.

Beyond training population absence, remaining barriers include narrow genetic bases that limit allelic diversity and effective population size, though wild relatives may harbor useful diversity [25]; pronounced G\u00d7E interactions requiring resource-intensive multi-environment trials [14]; and the high cost and low throughput of LC-MS/MS metabolite quantification relative to GS training population scale. NIRS offers higher throughput once calibrated, but calibration itself requires large reference datasets, creating a circular resource challenge for species with no prior data [16]; drone-based hyperspectral phenotyping remains largely experimental but could substantially reduce costs if calibration challenges are resolved.

***Future Directions and Emerging Technologies***

The highest near-term priority is coordinated development of reference training populations for five to ten priority species [23]: approximately 500 to 1,000 diverse individuals per species, genotyped by GBS or affordable SNP arrays and phenotyped for key metabolites and disease resistance across two to three diverse environments, supported by open shared data infrastructure that allows models trained in one program to inform related species (Table 3). As training populations expand, machine learning and deep learning models incorporating multi-omic covariates will become increasingly viable for capturing epistatic interactions among pathway genes [12,13]. GS and CRISPR-based genome editing are strategically complementary: GS identifies optimal alleles and elite genetic backgrounds, while editing enables precise correction of suboptimal alleles without extended backcrossing [18,6]. Efficient transformation systems already established in Salvia miltiorrhiza [30,31] and demonstrated pathway edits in Panax ginseng [24,32] show that pathway-level metabolite engineering is achievable and that GS-informed targets could guide similar interventions. Drone-mounted near-infrared and hyperspectral sensors could reduce metabolite phenotyping costs by an order of magnitude if spectral signatures reliably correlate with compound concentration across germplasm and environments [27,16].

**Table 3. Comparison of High-Throughput Phenotyping Platforms for Secondary Metabolite Quantification in Large Training Populations**

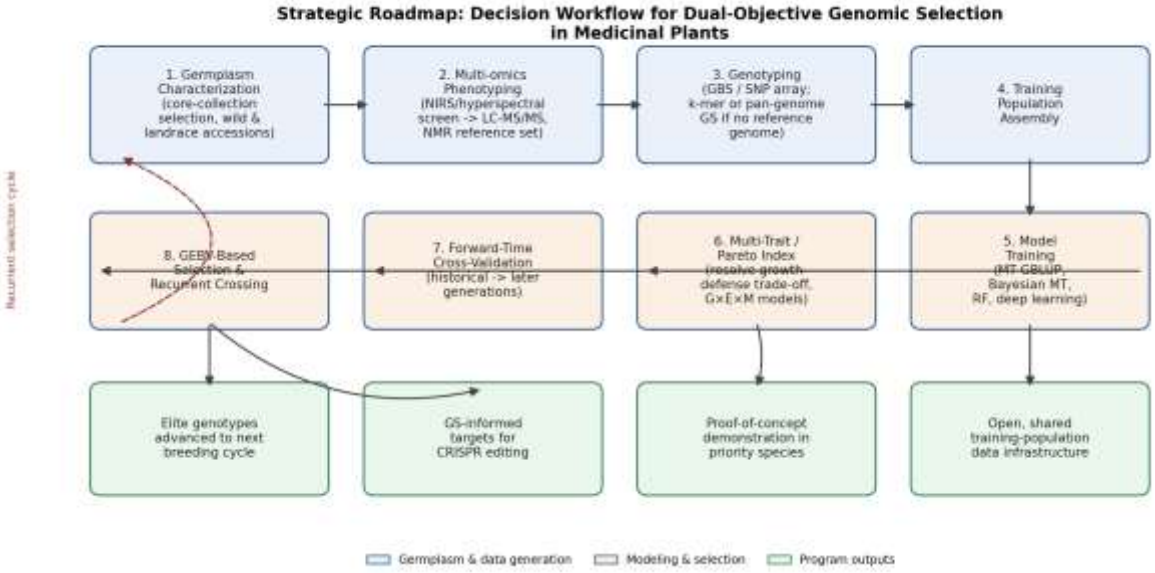
| Platform | Compounds Detectable | Throughput | Cost per Sample | Accuracy / Quantitative Range | Key Limitation | Reference |
|---|---|---|---|---|---|---|
| Near-Infrared Spectroscopy (NIRS) | Multiple metabolites simultaneously (after calibration) | Very high (100s–1000s/day) | Very low once calibrated | Moderate; requires reference calibration against HPLC or LC-MS | Calibration costly; accuracy limited for trace compounds | [16] |
| HPLC-DAD | Known chromophoric metabolites; targeted classes | Moderate (10s–100s/day) | Low-moderate | High for UV-active compounds; quantitative | Limited to UV-absorbing compounds; moderate throughput | [16] |
| LC-MS/MS | Targeted or untargeted; broad metabolome coverage | Low-moderate (10–50/day per system) | High | Very high; trace-level quantification; reference standard required | Expensive; requires reference standards; low throughput relative to cost | [16] |
| Drone-based NIR / Hyperspectral imaging | Canopy-level estimates of metabolite-correlated reflectance traits | Very high (field-scale, 1000s simultaneously) | Very low per plant once system established | Moderate; requires ground-truth calibration; non-destructive | Currently developmental; calibration requires extensive validation | [16,27] |

Note. HPLC-DAD = high-performance liquid chromatography with diode array detection; LC-MS/MS = liquid chromatography-tandem mass spectrometry; NIR = near-infrared; UV = ultraviolet. Platform selection should be guided by specific compound targets, required throughput, and available resources. Combined platforms (NIRS for screening; LC-MS/MS for validation) are recommended for large-scale GS training populations.

## Strategic Roadmap: A Structured Decision Workflow

The theoretical case for GS (Figure 3) in medicinal plant breeding is strong: the polygenic architectures of both quantitative disease resistance and secondary metabolite content are precisely the trait classes for which GS delivers the greatest advantage over marker-assisted selection, and the biological connectivity between defense and pharmaceutical metabolism provides a structural basis for multi-trait models to exploit positive genetic correlations. This said, no study directly quantifies genetic correlations between pharmaceutical metabolite content and disease resistance in any medicinal species; current claims rest on shared regulatory pathways rather than direct genetic estimates [20], and multi-trait GS will not deliver expected benefits if true correlations prove weak. Translating theoretical potential into practice also faces challenges more severe than in established crop systems: absent training populations, narrow genetic bases, expensive phenotyping, and limited breeding infrastructure together constitute a barrier that methodological refinement alone cannot resolve.

**Figure 3. Decision workflow from germplasm characterization through recurrent selection for dual-objective genomic selection programs in medicinal plants. Blue: germplasm and data-generation steps; orange: modeling and selection steps; green: program outputs feeding the next breeding cycle.**




## CONCLUSION

The central conclusion of this review is that this barrier is infrastructural rather than methodological: no GS program exists for medicinal plant metabolite quality not because models are unavailable, but because no training population with the necessary genotypic and phenotypic data has been assembled. Closing this gap requires coordinated investment in diverse, genotyped germplasm collections for priority species; multi-environment metabolite and disease phenotyping trials; high-throughput phenotyping platforms including NIRS calibration and remote sensing; and open-access data infrastructure enabling cross-species model sharing and validation. The parallel expansion of genome editing tools creates useful synergy, since species where editing is efficient (Salvia miltiorrhiza, Panax ginseng) are also those where GS is most tractable. A focused initial investment in three to five genomically prepared species could deliver the proof-of-concept demonstrations needed to catalyze broader adoption, ultimately transforming the consistency, sustainability, and quality of botanical pharmaceutical raw materials for global healthcare.

## DECLARATIONS

**Ethical Approval:** Not applicable; this is a narrative literature review involving no human or animal subjects.

**Competing Interests:** The authors declare no conflict of interest.

**Funding:** This research received no external funding.

**Author Contributions:** Muhammad Tahsin, Ali Murtaza, Maryam Alim : conceptualization, literature synthesis, critical analysis, and writing – original draft and revision. All authors approved the final manuscript.

**Data Availability:** Not applicable; no new data were generated. All sources discussed are cited in the reference list.

**AI Usage Disclosure:** During the preparation of this manuscript, AI-assisted tools (specifically, large language models) were used exclusively for language refinement, grammar correction, and formatting assistance. No AI tools were used for literature review, data analysis, conceptualization, interpretation, or generation of scientific content. All substantive intellectual contributions, including the synthesis of literature, critical analysis, and formulation of conclusions, were performed by the human authors. The authors take full responsibility for the accuracy, originality, and integrity of all content presented in this manuscript.

## REFERENCES

1. Davis CC, Choisy P. Medicinal plants meet modern biodiversity science. Current Biology. 2024;34(4):R158–R173.
2. Wang Q, Zhao X, Jiang Y, Jin B, Wang L. Functions of representative terpenoids and their biosynthesis mechanisms in medicinal plants. Biomolecules. 2023;13(12):1725. https://doi.org/10.3390/biom13121725
3. Wani KI, Choudhary S, Zehra A, Naeem M, Weathers P, Aftab T. Enhancing artemisinin content in and delivery from Artemisia annua: A review of alternative, classical, and transgenic approaches. Planta. 2021;254(2):29. https://doi.org/10.1007/s00425-021-03684-3
4. Ma Y, Cui G, Chen T, Ma X, Wang R, Jin B, Yang J, Kang L, Tang J, Lai C, et al. Expansion within the CYP71D subfamily drives the heterocyclization of tanshinones synthesis in Salvia miltiorrhiza. Nature Communications. 2021;12(1):685.
5. Erb M, Kliebenstein DJ. Plant secondary metabolites as defenses, regulators, and primary metabolites: The blurred functional trichotomy. Plant Physiology. 2020;184:39–52.
6. Gao H, Pei X, Song X, Wang S, Yang Z, Zhu J, Lin Q, Zhu Q, Yang X. Application and development of CRISPR technology in the secondary metabolic pathway of the active ingredients of phytopharmaceuticals. Frontiers in Plant Science. 2025;15:1477894.
7. Das S, Kwon M, Kim JY. Enhancement of specialized metabolites using CRISPR/Cas gene editing technology in medicinal plants. Frontiers in Plant Science. 2024;15:1279738. https://doi.org/10.3389/fpls.2024.1279738
8. Mani V, Park S, Kim JA, Lee SI, Lee K. Metabolic perturbation and synthetic biology strategies for plant terpenoid production an updated overview. Plants. 2021;10(10):2179.
9. Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001;157(4):1819-1829. https://doi.org/10.1093/genetics/157.4.1819
10. Guo C, Xu S, Guo X. Genome-wide analysis of oxidosqualene cyclase genes in Artemisia annua: Evolution, expression, and potential roles in triterpenoid biosynthesis. Current Issues in Molecular Biology. 2025;47(8):545. https://doi.org/10.3390/cimb47080545

11. Li Y, Chen J, Zhi J, Huang D, Zhang Y, Zhang L, Duan X, Zhang P, Qiu S, Geng J. The ABC transporter SmABCG1 mediates tanshinones export from the peridermic cells of Salvia miltiorrhiza root. Journal of Integrative Plant Biology. 2025;67(1):135–149. https://doi.org/10.1111/jipb.13812

12. Ndochinwa OG, Wang QY, Amadi OC, Nwagu TN, Nnamchi CI, Okeke ES, Moneke AN. Current status and emerging frontiers in enzyme engineering: An industrial perspective. Heliyon. 2024;10(6):e32673. https://doi.org/10.1016/j.heliyon.2024.e32673

13. Lawson CE, Martí JM, Radivojevic T, Jonnalagadda SVR, Gentz R, Hillson NJ, Peisert S, Kim J, Simmons BA, Petzold CJ, et al. Machine learning for metabolic engineering: A review. Metabolic Engineering. 2021;63:34–60. https://doi.org/10.1016/j.ymben.2020.10.005

14. Zhang Y, Cao X, Wang J, Tang F. Enhancement of linalool production in Saccharomyces cerevisiae by utilizing isopentenol utilization pathway. Microbial Cell Factories. 2023;21(1):212. https://doi.org/10.1186/s12934-022-01953-6

15. Sirirungruang S, Markel K, Shih PM. Plant-based engineering for production of high-valued natural products. Natural Product Reports. 2022;39(7):1492–1509. https://doi.org/10.1039/D2NP00010B

16. Oh SW, Imran M, Kim EH, Park SY, Lee SG, Park HM, Jung JW, Ryu TH. Approach strategies and application of metabolomics to biotechnology in plants. Frontiers in Plant Science. 2023;14:1192235. https://doi.org/10.3389/fpls.2023.1192235

17. Li L, Fu J, Liu N. Advances in the structures, pharmacological activities, and biosynthesis of plant diterpenoids. Journal of Microbiology and Biotechnology. 2024;34(8):1563–1579. https://doi.org/10.4014/jmb.2405.05012

18. Zhang F, Neik TX, Thomas WJW, Batley J. CRISPR-based genome editing tools: An accelerator in crop breeding for a changing future. International Journal of Molecular Sciences. 2023;24(10):8623. https://doi.org/10.3390/ijms24108623

19. Scossa F, Benina M, Alseekh S, Zhang Y, Fernie AR. The integration of metabolomics and next-generation sequencing data to elucidate the pathways of natural product metabolism in medicinal plants. Planta Medica. 2018;84(12–13):855–873.

20. Hassani D, Taheri A, Fu X, Qin W, Hang L, Ma Y, Tang K. Elevation of artemisinin content by co-transformation of artemisinin biosynthetic pathway genes and trichome-specific transcription factors in Artemisia annua. Frontiers in Plant Science. 2023;14:1118082. https://doi.org/10.3389/fpls.2023.1118082

21. Hao X, Pu Z, Cao G, You D, Zhou Y, Deng C, Shi M, Nile SH, Wang Y, Zhou W, Kai G. Tanshinone and salvianolic acid biosynthesis are regulated by SmMYB98 in Salvia miltiorrhiza hairy roots. Journal of Advanced Research. 2020;23:1–12. https://doi.org/10.1016/j.jare.2020.01.012

22. Köllner TG, Degenhardt J, Gershenzon J. The product specificities of maize terpene synthases TPS4 and TPS10 are determined both by active site amino acids and residues adjacent to the active site. Plants. 2020;9(5):552. https://doi.org/10.3390/plants9050552

23. Guo C, Xu S, Guo X. Metabolic engineering of terpenoid biosynthesis in medicinal plants: From genomic insights to biotechnological applications. Current Issues in Molecular Biology. 2025;47(9):723.

24. Choi HS, Koo HB, Jeon SW, Han JY, Kim JS, Jun KM, Choi YE. Modification of ginsenoside saponin composition via the CRISPR/Cas9-mediated knockout of protopanaxadiol 6-hydroxylase gene in Panax ginseng. Journal of Ginseng Research. 2022;46(3):505–514. https://doi.org/10.1016/j.jgr.2021.06.004

25. Chen Q, Jiang T, Liu Y, Liu H, Zhao T, Liu Z, Gan X, Hallab A, Wang X, He J, Ma Y, Zhang F, Jin Q, Wang G, Zhang Y, Zhang P. Recently duplicated sesterterpene (C25) gene clusters in Arabidopsis thaliana modulate root microbiota. Science China Life Sciences. 2019;62:947–958.

26. Bureau JA, Oliva ME, Dong Y, Ignea C. Engineering yeast for the production of plant terpenoids using synthetic biology approaches. Natural Product Reports. 2023;40(12):1822–1848. https://doi.org/10.1039/D3NP00005A

27. Li C, Yan X, Xu Z, Wang Y, Shen X, Zhang L, Zhou Z, Wang P. Pathway elucidation of bioactive rhamnosylated ginsenosides in Panax ginseng and their de novo high-level production by engineered Saccharomyces cerevisiae. Communications Biology. 2024;5(1):775.

28. Rinaldi MA, Ferraz CA, Scrutton NS. Alternative metabolic pathways and strategies to high-titre terpenoid production in Escherichia coli. Natural Product Reports. 2022;39(1):90–118. https://doi.org/10.1039/D1NP00025J

29. Traverse KKF, Breselge S, Trautman JG, Dee A, Wang J, Childs KL, Lee-Parsons CWT. Characterization of the ZCTs, a subgroup of Cys2-His2 zinc finger transcription factors regulating alkaloid biosynthesis in Catharanthus roseus. Plant Cell Reports. 2024;43(8):209. https://doi.org/10.1007/s00299-024-03306-8

30. Hsu CT, Chiu CC, Hsiao PY, Lin CY, Cheng S, Lin YC, Yeh KW. Transgene-free CRISPR/Cas9-mediated gene editing through protoplast-to-plant regeneration enhances active compounds in Salvia miltiorrhiza. Plant Biotechnology Journal. 2024;22(6):1549–1551. https://doi.org/10.1111/pbi.14285

31. Cao X, Xie H, Song M, Zhao L, Liu H, Li G, Zhu J. Simple method for transformation and gene editing in medicinal plants. Journal of Integrative Plant Biology. 2024;66(1):17–19. https://doi.org/10.1111/jipb.13593

32. Yao L, Wang J, Sun J, He J, Paek KY, Park SY. A strategy combining overexpression and CRISPR/Cas9-mediated gene knockdown to enhance the accumulation of ginsenoside Rg3 in Panax ginseng hairy roots. Journal of Ginseng Research. 2022;46(4):571–579.