Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Characterization of High-Artemisinin Yielding Artemisia annua Bioecotypes Using Gene-Specific STS Markers and HPLC

Mohammad Taher Hallajian
Nuclear Science and Technology Research Institute (NSTRI)
Sahar Tavakoli Mohammadi
Payam Noor University Alborz Branch
Mohammad Ali Ebrahimi
Payam Noor University Alborz Branch
Mohammad Reza Naghavi
University of Tehran
Mojtaba Kordrostami
mkordrostami@aeoi.org.ir

Nuclear Science and Technology Research Institute (NSTRI)

Article

Keywords: Artemisia annua, artemisinin, STS marker, artemisinin biosynthesis, molecular diversity, marker-assisted selection (MAS), HPLC, PCA, secondary metabolites

Posted Date: December 17th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7265007/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Artemisia annua is the sole natural source of artemisinin, an essential compound for malaria treatment. However, variable yields among bioecotypes hinder consistent production. This study aimed to characterize 17 *A. annua* biotypes using gene-specific STS markers targeting key enzymes in the artemisinin biosynthesis pathway and to evaluate their association with artemisinin content. Eleven STS primer pairs were designed for genes including *ADS*, *CYP71AV1*, *DBR2*, *ALDH1*, *HMGR*, *TTG1*, and others. PCR amplification, gel electrophoresis, and HPLC quantification of artemisinin were performed. Molecular marker efficiency indices were calculated. Hierarchical clustering, PCA, correlation heatmaps, and radar chart visualizations were employed to assess genotype–metabolite relationships. A total of 10 markers were polymorphic across biotypes, with significant correlation to artemisinin content ($R^2 > 0.85$ for key markers). Cluster and PCA analyses grouped biotypes into high-, medium-, and low-producing clusters. Biotypes 260 and 316 consistently displayed superior marker profiles and artemisinin content (> 2.7 mg/g DW). Marker indices such as PIC, MI, and EMR highlighted *TTG1*, *DXR*, and *ADS* as the most informative loci. Radar chart analysis confirmed these markers and genotypes as optimal targets for selection. Gene-targeted STS markers can effectively distinguish elite *A. annua* biotypes with high artemisinin content. This integrated molecular-metabolite framework supports marker-assisted breeding and transgenic improvement for enhanced artemisinin production.

1. Introduction

Medicinal plants have long been a cornerstone of both traditional and modern healthcare systems, offering a vast array of bioactive compounds essential to pharmaceutical innovation [1,2]. Among these, *Artemisia annua* L., commonly known as sweet wormwood, holds a special place due to its role in producing artemisinin—a sesquiterpene lactone with a unique endoperoxide bridge [3]. Artemisinin is widely recognized as the most effective treatment for malaria caused by *Plasmodium falciparum* [4]. Since its discovery in China during the 1970s and the subsequent awarding of the Nobel Prize in Physiology or Medicine in 2015, artemisinin and its derivatives (such as artesunate, artemether, and dihydroartemisinin) have become the cornerstone of artemisinin-based combination therapies (ACTs), which are currently endorsed by the World Health Organization as the frontline treatment against drug-resistant malaria [5]. Despite its proven efficacy, the commercial supply of artemisinin faces persistent challenges [6]. Natural production in the plant is low—often less than 0.5% of dry weight—and varies significantly between genotypes [7]. Additionally, chemical synthesis remains inefficient and costly. These constraints have spurred worldwide interest in strategies to improve artemisinin yields, including metabolic engineering, tissue culture, and—critically—the identification of high-producing natural genotypes [8].

Artemisinin biosynthesis is governed by a complex enzymatic cascade beginning with amorpha-4,11-diene synthase (*ADS*) [9] and proceeding through several redox and cyclization steps involving key enzymes like *CYP71AV1*, *DBR2*, *ALDH1*, *HMGR*, and *FPS* [10]. The regulation and expression of these

genes are central to the plant's capacity to synthesize artemisinin [2]. Therefore, molecular profiling of *A. annua* germplasm using gene-specific markers provides a powerful and targeted approach for identifying elite genotypes with superior biosynthetic potential [11].

Among available molecular tools, Sequence Tagged Site (STS) markers are especially useful for such applications [12]. These short, gene-targeted DNA fragments are highly specific, reproducible, and capable of detecting functional allelic variations [13]. When designed from genes involved in artemisinin biosynthesis, STS markers can be used as predictive tools to screen *A. annua* populations for high-yield potential. This molecular diagnostic approach is particularly valuable for countries like Iran, where the local germplasm remains genetically undercharacterized despite harboring a wealth of natural *A. annua* ecotypes adapted to diverse climates.

The novelty of this study lies in its integrated approach—merging gene-targeted STS marker analysis with phytochemical quantification via HPLC—to perform the first molecular and metabolite-level characterization of Iranian *A. annua* biotypes. While previous studies have largely focused on morphological traits or general genetic variability, our research zeroes in on key genes in the artemisinin pathway and directly correlates their amplification with actual artemisinin levels. This dual-layered methodology enables a more precise and functional selection of elite genotypes.

Among the 17 biotypes screened, two genotypes—biotype 260 and biotype 316—stood out for their combined high expression of multiple biosynthetic genes and elevated artemisinin content. These findings not only underscore the genetic basis of artemisinin accumulation but also provide concrete molecular targets for breeding programs. The identified genotypes represent promising candidates for commercial cultivation and genetic improvement initiatives. In essence, this study pioneers the application of targeted molecular screening in *A. annua* within the Iranian context, demonstrating how combining modern molecular techniques with phytochemical analysis can accelerate the identification of high-yielding medicinal plant varieties. The outcomes of this work lay a foundation for more effective genotype selection, support the development of improved cultivars, and contribute to the long-term goal of sustainable and scalable artemisinin production.

2. Materials and Methods

2.1 Plant Material and Growth Conditions

Seventeen biotypes of *Artemisia annua* L. were collected from diverse ecological regions across northern Iran, particularly the Caspian provinces (Guilan, Mazandaran, and Golestan), where the species is naturally distributed. Biotypes of *Artemisia annua* were collected from Mazandaran, Guilan, and Golestan provinces under formal authorization from the National Botanical Gene Bank of Iran and the Iranian Biological Resource Center (IBRC), in compliance with institutional and national biodiversity regulations [14]. Formal taxonomic identification of all *Artemisia annua* L. biotypes was carried out by Dr. Mohammad Ali Ebrahimi, a plant taxonomist at the Payam Noor University of Karaj. Voucher specimens

of all 17 biotypes have been deposited in the Herbarium of the Payam Noor University of Karaj (Herbarium Code: PNKH), under the accession number PNKH-AAN2024-01 to PNKH-AAN2024-17. These specimens are publicly accessible for verification and future reference. All experimental protocols and plant sampling activities involving *Artemisia annua* L. were performed following institutional, national, and international guidelines and legislation. No endangered or protected species were involved. Biotypes were selected based on morphological variation and environmental origin to represent a broad genetic base. Biotypes were collected from the Caspian region of northern Iran, including: Mazandaran Province: 36.5651° N, 52.0494° E; Guilan Province: 37.1986° N, 49.6208° E and Golestan Province: 36.8471° N, 54.4342° E. Seeds were germinated in pots filled with a sterile soil-sand mixture (3:1), and the plants were grown under controlled greenhouse conditions (25 ± 2°C, 16/8 h light/dark photoperiod, 60–70% relative humidity). Plants were watered regularly, and no additional fertilizers were applied to avoid interference with secondary metabolite production. Young, healthy, fully expanded leaves were harvested from 8-week-old plants at the vegetative stage for DNA extraction and artemisinin quantification. Three biological replicates were collected for each biotype and stored at − 80°C until further use.

2.2 Genomic DNA Extraction

Genomic DNA was extracted using the CTAB-based method described by Doyle and Doyle (1987), with minor modifications to optimize yield and purity. Approximately 100 mg of frozen leaf tissue was ground in liquid nitrogen using a sterile mortar and pestle. The ground tissue was transferred to 2 mL microcentrifuge tubes containing 700 µL of preheated (65°C) CTAB extraction buffer (2% CTAB, 100 mM Tris-HCl pH 8.0, 20 mM EDTA, 1.4 M NaCl, 1% PVP, and 0.2% β-mercaptoethanol). Samples were incubated at 65°C for 30 minutes with occasional mixing by inversion. Following incubation, an equal volume of chloroform:isoamyl alcohol (24:1) was added, and the mixture was centrifuged at 12,000 rpm for 10 minutes at room temperature. The aqueous phase was transferred to a new tube, and DNA was precipitated by adding cold isopropanol (0.7 volume), followed by incubation at − 20°C for 1 hour. The DNA pellet was recovered by centrifugation at 13,000 rpm for 15 minutes, washed with 70% ethanol, air-dried, and resuspended in 50 µL of TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). DNA samples were stored at − 20°C.

2.3 DNA Quality and Quantification

DNA quality and concentration were assessed by two methods: Spectrophotometry using a NanoDrop (Thermo Scientific) at 260 and 280 nm to determine purity (A260/A280 ratio between 1.8–2.0 was considered acceptable). Agarose gel electrophoresis, where 5 µL of each DNA sample was loaded on a 0.8% agarose gel stained with ethidium bromide and visualized under UV transillumination to verify integrity.

2.4 Selection of STS Primers and PCR Amplification

Eleven gene-specific STS primer pairs were selected from published sequences targeting key enzymes involved in the artemisinin biosynthetic pathway, including *amorpha-4,11-diene synthase (ADS)*, *cytochrome P450 (CYP71AV1)*, *dihydroartemisinic aldehyde reductase (DBR2)*, *aldehyde dehydrogenase*

1 (ALDH1), *3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGR)*, *artemisinic aldehyde Δ11(13) reductase (Cdsaahdr)*, *farnesyl diphosphate synthase (FPS)*, *1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR)*, *artemisinin metabolite biosynthetic gene (AMDS)*, and *triterpene glycosyltransferase (TTG1)*. Primer sequences, expected amplicon sizes, and annealing temperatures are listed in Supplementary Table S1. PCR amplification reactions were performed in a total volume of 25 µL, containing 12.5 µL of 2× PCR Master Mix (Amplicon or equivalent, including *Taq* DNA polymerase, dNTPs, and $MgCl_2$), 0.5 µM of each forward and reverse primer, 100 ng of genomic DNA template, and nuclease-free water to adjust the final volume. PCR thermal cycling conditions were optimized individually for each primer pair but generally followed the protocol: initial denaturation at 94°C for 3 minutes, followed by 35 cycles of denaturation at 94°C for 30 seconds, annealing at 55–62°C (primer-dependent) for 30 seconds, and extension at 72°C for 60 seconds, with a final extension step at 72°C for 10 minutes. All reactions included negative controls lacking DNA template to monitor for potential contamination.

2.5 Gel Electrophoresis and Band Scoring

PCR products were separated on 1.5% agarose gels in 1× TAE buffer at 100 V for 60 minutes and stained with ethidium bromide. A 100 bp DNA ladder was used as a size marker. Gels were photographed under UV light using a gel documentation system (Bio-Rad Gel Doc or equivalent). Presence or absence of bands was scored manually in a binary format (1 for presence, 0 for absence) across all biotypes. Polymorphic information content (PIC) and gene diversity were calculated using PowerMarker v3.25.

2.6 Genetic Relationship and Cluster Analysis

Binary data matrices derived from PCR results were subjected to genetic similarity analysis using the R software package. Jaccard's similarity coefficient was calculated, and a dendrogram was constructed using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) to visualize genetic relationships among biotypes.

2.7 Artemisinin Quantification by HPLC

Artemisinin content in dried leaf tissue was quantified using High-Performance Liquid Chromatography (HPLC). For each biotype, 100 mg of powdered leaf material was extracted in 10 mL of methanol by sonication for 30 minutes, followed by centrifugation at 10,000 rpm for 10 minutes. The supernatant was filtered through a 0.22 µm syringe filter and injected into the HPLC system. The HPLC analysis was performed using a C18 reverse-phase column (250 mm × 4.6 mm, 5 µm) with the mobile phase consisting of methanol:water (70:30 v/v) at a flow rate of 1.0 mL/min. Detection was done at 260 nm using a UV–Vis detector. Artemisinin standard (Sigma-Aldrich) was used to generate the calibration curve, and results were expressed as mg artemisinin per g dry weight.

2.8 Statistical Analysis

Data from molecular (binary) and biochemical (HPLC) assessments were analyzed using Microsoft Excel and SPSS v26.0. Correlation analysis was conducted between gene amplification profiles and

artemisinin content. Analysis of variance (ANOVA) was used to test for significant differences among biotypes in artemisinin concentration. Principal component analysis (PCA) was also performed to visualize patterns in marker–trait association.

3. Results

3.1 DNA Extraction and Quality Assessment

High-quality genomic DNA was successfully extracted from the young leaf tissues of all 17 *Artemisia annua* biotypes using a modified cetyltrimethylammonium bromide (CTAB) protocol, optimized for medicinal plant tissues rich in secondary metabolites (Fig. 1).

The protocol involved pre-treatment with PVP and RNase A to minimize polysaccharide and RNA contamination, respectively, and yielded robust quantities of DNA suitable for PCR amplification. The concentration of extracted DNA ranged from 80 to 150 ng/µL, as quantified using a NanoDrop™ 2000 spectrophotometer. The A260/A280 absorbance ratios fell between 1.80 and 1.92, indicating minimal protein contamination and confirming the suitability of the DNA for downstream enzymatic reactions. DNA purity was further supported by A260/A230 ratios above 2.0 in most samples, reflecting low levels of polyphenolic and carbohydrate impurities, which are common in *A. annua* leaf tissues. Integrity and quality of the extracted genomic DNA were assessed by electrophoresis on 1% agarose gels stained with ethidium bromide. All samples exhibited clear, high-molecular-weight bands with no visible smearing, degradation, or RNA contamination, thereby validating the effectiveness of the extraction protocol. The absence of shearing artifacts further ensured that the DNA was of sufficient quality for sequence-tagged site (STS) marker amplification. This high-quality genomic DNA served as the template for subsequent PCR-based genotyping of ten key genes involved in the artemisinin biosynthetic pathway. The consistency and integrity of the DNA samples across all 17 biotypes contributed significantly to the reproducibility and reliability of the STS profiling, correlation analyses, and downstream genetic diversity assessments.

3.2 Allelic Amplification Pattern of Artemisinin Biosynthetic Genes Across Biotypes

To investigate the presence of key genes involved in the artemisinin biosynthetic pathway among different *Artemisia annua* biotypes, a set of ten gene-specific STS primers was employed. The amplification profiles revealed distinct allelic patterns across the 17 tested biotypes, reflecting molecular diversity and differential gene distribution.

The genes analyzed included *FPS*, *TTG1*, *Cdsaahdr*, *ADS*, *CYP71AV1*, *DXR*, *HMGR*, *AMDS*, *DBR2*, and *ALDH1*. Overall, the amplification frequency varied substantially between genes and among biotypes. Some genes, such as *TTG1* and *DXR*, exhibited a broader amplification pattern, being present in up to 7 and 6 biotypes respectively, while others like *HMGR*, *AMDS*, *DBR2*, and *ALDH1* amplified in only two biotypes, notably *260* and *316*, which consistently showed strong amplification across the entire gene

panel. The biotype *260* demonstrated amplification of 9 out of 10 targeted genes, missing only *HMGR*, while biotype *316* showed a similarly complete pattern, suggesting that these two genotypes carry the most functionally intact artemisinin biosynthetic gene sets. In contrast, several biotypes—including *306*, *301*, *305*, *327*, *331*, *237*, *273*, *94*, *264*, and *243*—exhibited minimal or no amplification across most primer pairs, indicating the potential absence or reduced expression of the corresponding alleles. A heatmap representation of the banding data visually illustrates the presence/absence patterns of each gene across all biotypes, highlighting the clear genetic distinction of biotypes *260* and *316*. The allelic variability observed for genes such as *ADS*, *CYP71AV1*, and *Cdsaahdr*—all of which are critical to the conversion of farnesyl diphosphate to artemisinin precursors—suggests that differential gene presence could underlie variation in artemisinin production among ecotypes. This allelic profiling approach using gene-specific STS markers offers a rapid and informative strategy to identify elite genotypes with the full complement of biosynthetic genes. The results support the classification of *260* and *316* as promising candidates for high-yield artemisinin production, potentially due to their possession of a complete biosynthetic gene network.

3.3 Gene-wise Amplification Frequency and Correlation with Artemisinin Yield

To elucidate the potential role of specific artemisinin biosynthetic genes in determining metabolite yield, we analyzed both the amplification frequency of each gene and its statistical correlation with artemisinin concentration across the 17 *Artemisia annua* biotypes (Table 1). Among the STS markers used, the gene *TTG1* exhibited the highest amplification frequency, being detected in 7 out of 17 biotypes, followed by *ADS*, *CYP71AV1*, and *DXR*, each amplified in 4 biotypes. Conversely, *FPS*, *Cdsaahdr*, and *ALDH1* were present in only 3 biotypes, while *HMGR*, *AMDS*, and *DBR2* were amplified in just 2 biotypes. Notably, *FDS* failed to amplify in any biotype, likely due to absence or significant sequence divergence at primer binding sites. To investigate the relationship between gene presence and metabolite accumulation, Pearson correlation coefficients were calculated between gene amplification (binary: 1 = present, 0 = absent) and quantified artemisinin content across biotypes. Remarkably, *Cdsaahdr* showed the strongest positive correlation with artemisinin yield ($r = 0.83$), suggesting a significant role in downstream aldehyde reduction and artemisinin precursor stabilization. *FPS* ($r = 0.75$) and *ADS* ($r = 0.71$) also exhibited strong positive correlations, consistent with their central roles in the early biosynthetic steps. Interestingly, although *TTG1* was the most frequently amplified gene, its correlation with artemisinin production was low ($r = 0.10$), indicating that presence alone may not be predictive of functional contribution in this context. Other genes such as *DXR* and *CYP71AV1* showed moderate positive correlations ($r \approx 0.71$), reinforcing their involvement in precursor formation and oxidative transformation, respectively. These findings underscore the differential influence of individual pathway genes on artemisinin biosynthesis and suggest that not all genes equally contribute to yield variation among biotypes. Genes such as *Cdsaahdr*, *FPS*, and *ADS* emerge as strong molecular markers for selecting high-yielding genotypes. Their consistent association with artemisinin levels highlights the potential of integrating marker-assisted selection with targeted metabolic engineering for artemisinin enhancement.

Table 1
Gene-wise amplification frequency and its correlation with artemisinin yield among 17 *Artemisia annua* biotypes.

Gene	Amplification frequency	Correlation with Artemisinin
FPS	3	0.745714
TTG1	7	0.103172
Cdsaahdr	3	0.830671
ADS	4	0.714318
CYP7AV1	4	0.714318
DXR	5	0.622636
HMGR	2	0.435257
AMDS	2	0.933502
DBR2	2	0.933502
ALDH1	2	0.933502
FDS	0	

Each gene-specific STS marker was evaluated for its amplification frequency (number of biotypes showing positive PCR bands) and statistically correlated with quantified artemisinin content (mg/g DW). Markers such as *TTG1*, *ADS*, and *CYP71AV1* exhibited higher amplification frequencies and strong positive correlations ($r > 0.85$) with artemisinin levels, suggesting their critical roles in regulating the biosynthetic pathway. This analysis highlights key molecular markers with potential utility for genotype selection and artemisinin yield prediction in breeding programs.

3.3 Genetic Diversity and Cluster Analysis

To assess the effectiveness of the STS primers in differentiating among *Artemisia annua* biotypes, four key marker efficiency indices were calculated (Table 2): Frequency of Presence, Polymorphism Information Content (PIC), Marker Index (MI), and Resolving Power (Rp). These indices provide a quantitative framework for comparing the discriminatory capacity and informativeness of each primer. Among the markers analyzed, *TTG1* exhibited the highest performance across multiple indices. It was amplified in 41.2% of the biotypes and recorded the highest PIC value (0.484), indicating a strong ability to distinguish between genotypes based on allele presence. Additionally, it showed the highest Marker Index (MI = 3.39), highlighting both its informativeness and consistent amplification across samples. Although its Resolving Power (Rp = 0.35) was moderate, its combined performance across indices establishes TTG1 as the most efficient primer in this study. Markers such as FPS, ADS, and CYP71AV1 displayed identical profiles, each being present in 23.5% of the biotypes and yielding a PIC value of 0.360. These primers showed a relatively strong Rp (1.06) and moderate MI values (1.44), suggesting they are moderately effective in capturing genetic variation. Their performance indicates that although they may not be the most informative individually, they can complement more powerful markers in multi-

locus genotyping approaches. The Cdsaahdr primer, despite having the lowest frequency of presence (17.6%), recorded a respectable PIC value (0.291) and the highest Resolving Power (Rp = 1.29) among all primers. This indicates that although Cdsaahdr is less commonly amplified, it is highly efficient in separating distinct biotypes when present, likely due to its unique allele profile in select genotypes. In general, PIC values ranged from 0.291 to 0.484, reflecting moderate levels of polymorphism among the biotypes. These results suggest that while all markers demonstrated some degree of informativeness, TTG1 is particularly well-suited for broad-scale diversity assessments and marker-assisted selection. Conversely, Cdsaahdr offers high resolution in discriminating specific genotypes despite limited amplification frequency. The combined use of markers with high PIC (e.g., TTG1), high Rp (e.g., Cdsaahdr), and moderate MI (e.g., FPS and ADS) allows for an optimized genotyping strategy that balances marker informativeness with discriminatory power. These findings underscore the importance of evaluating multiple efficiency parameters when selecting molecular markers for population genetics, diversity screening, and targeted breeding in *Artemisia annua*.

Table 2
Molecular marker indices for gene-specific STS primers in *Artemisia annua* biotypes.

Gene	Frequency (Presence)	PIC	MI	Rp
FPS	0.235294	0.359862	1.439446	1.058824
TTG1	0.411765	0.484429	3.391003	0.352941
Cdsaahdr	0.176471	0.290657	0.871972	1.294118
ADS	0.235294	0.359862	1.439446	1.058824
CYP71AV1	0.235294	0.359862	1.439446	1.058824
DXR	0.294118	0.415225	2.076125	0.823529
HMGR	0.117647	0.207612	0.415225	1.529412
AMDS	0.117647	0.207612	0.415225	1.529412
DBR2	0.117647	0.207612	0.415225	1.529412
ALDH1	0.117647	0.207612	0.415225	1.529412

Shown are the Polymorphic Information Content (PIC), Marker Index (MI), Resolving Power (RP), calculated for each STS primer, indicating each marker's informativeness and utility in assessing genetic diversity among the biotypes.

To identify genetic relationships among *Artemisia annua* biotypes and explore associations between artemisinin biosynthesis gene presence and metabolite yield, hierarchical clustering was performed using Ward's method with Euclidean distance as the similarity metric (Fig. 3). The dataset comprised the presence/absence matrix of ten STS-amplified biosynthetic genes and the quantitative artemisinin content across 17 biotypes. The resulting dendrogram revealed a clear separation of the biotypes into three major clusters, each corresponding to distinct genetic and biochemical profiles: Cluster I included

biotypes 260 and 316, which consistently amplified the largest number of biosynthetic genes (9 and 10, respectively) and exhibited the highest artemisinin contents (3.01 and 2.75 mg/g DW). These biotypes showed amplification of key genes such as *ADS*, *CYP71AV1*, *DXR*, *Cdsaahdr*, *FPS*, and *ALDH1*, all of which are critical to precursor formation and artemisinin yield. Their tight clustering at low Euclidean distances suggests a high degree of genetic similarity and biosynthetic efficiency. Biotypes such as 258, 313, 94, and 328 formed a second distinct group. These biotypes displayed moderate artemisinin content (2.14–2.76 mg/g DW) and variable amplification of 3–6 genes. For instance, biotype 313 showed strong amplification of *ADS*, *FPS*, *TTG1*, and *HMGR*, while biotype 258 amplified only *TTG1*. This cluster likely represents intermediate phenotypes with partial biosynthetic capacity. The remaining biotypes, including 237, 273, 243, 331, and 264, clustered together and were characterized by the absence or minimal amplification of biosynthetic genes. Correspondingly, their artemisinin contents were negligible (< 0.05 mg/g DW). These ecotypes appear to lack the functional gene architecture required for efficient artemisinin production and may serve as low-yielding or null genotypes. The clustering pattern demonstrates a strong correlation between genetic amplification profiles and artemisinin yield. Biotypes with more complete gene presence profiles tended to group together and exhibited higher metabolite content. These findings not only highlight the genetic diversity within *A. annua* but also validate the use of gene-specific STS markers for classifying biotypes with superior biosynthetic potential.

3.4 Artemisinin Quantification via HPLC

Artemisinin content was quantified in 17 *Artemisia annua* biotypes using high-performance liquid chromatography (HPLC), and the results were visualized as a heatmap to reveal concentration patterns across the genotypes. As shown in Fig. 4, artemisinin accumulation varied widely among the biotypes, ranging from non-detectable levels to over 3.00 mg/g dry weight (DW), reflecting substantial biochemical diversity within the studied population.

Biotypes 260 and 316 emerged as the top producers, with artemisinin concentrations of 3.01 mg/g DW and 2.75 mg/g DW, respectively. These biotypes were characterized by both high metabolite yield and extensive amplification of biosynthetic pathway genes, supporting the association between genetic architecture and metabolic output. Biotypes such as 328, 258, 306, and 313 formed a mid-range group with moderate artemisinin content (2.1–2.7 mg/g DW), suggesting partial activation of the biosynthetic pathway. In contrast, several biotypes—including 237, 243, 273, 331, and 264—exhibited negligible or undetectable levels of artemisinin (< 0.06 mg/g DW), consistent with their limited or absent gene amplification profiles as revealed by STS marker analysis. These low-yielding biotypes may serve as valuable negative controls in future functional genomics studies or breeding efforts. The heatmap effectively illustrates the quantitative spectrum of artemisinin production, with a smooth gradient from high to low values. This visual distribution aligns closely with the genetic clustering and marker-based analyses described earlier, validating the functional relevance of the identified biosynthetic genes. The data also emphasize the potential of high-yielding genotypes like biotype 260 for use in elite breeding programs and commercial artemisinin production.

3.5 Integration of Molecular and Phytochemical Data

To gain deeper insight into the multivariate relationships between biosynthetic gene presence and artemisinin production among *Artemisia annua* biotypes, Principal Component Analysis (PCA) was performed using the binary amplification data of 10 STS gene markers along with the corresponding artemisinin content. The PCA biplot (Fig. 5) effectively captured the major trends in the dataset, with PC1 and PC2 together explaining a substantial proportion of the total variance.

The PCA revealed clear separation among biotypes based on their molecular and biochemical profiles. Biotypes such as 260 and 316, which were previously identified as high artemisinin producers, clustered closely in the positive PC1 and PC2 quadrants. This region was also strongly influenced by positive loadings of key genes including *TTG1*, *FPS*, *CYP71AV1*, *ADS*, and *AMDS*, suggesting that these genes substantially contribute to the variability in artemisinin yield across genotypes. Conversely, biotypes with low or undetectable artemisinin levels (e.g., 237, 243, 331, 273) clustered near the origin or in the negative region of PC1, indicating minimal contribution from biosynthetic genes and poor metabolic productivity. These biotypes exhibited minimal or no amplification of STS markers, reflecting their limited genetic capacity for artemisinin biosynthesis. Among the features, TTG1 and FPS vectors showed the longest projections on the PCA plot, indicating their strong discriminating power among biotypes and their close correlation with artemisinin content. The orientation of these vectors along PC1 suggests that this principal component primarily represents variation associated with artemisinin biosynthesis potential. In contrast, vectors such as *DXR*, *ALDH1*, and *Cdsaahdr* contributed more modestly to biotype separation along PC2, likely representing more specialized or secondary roles in the pathway.

3.12 Multitrait Profiling of Top Five Biotypes Based on Artemisinin and STS Marker Contribution

To visualize the combinatorial behavior of artemisinin accumulation and the presence of key biosynthetic genes, a radar chart was generated for the five most productive *A. annua* biotypes, as determined by their quantified artemisinin content (Fig. 6). The selected biotypes—260, 316, 328, 258, and 306—were analyzed across 11 traits, including artemisinin concentration and 10 STS marker amplification scores. The radar plot revealed notable differences in molecular profiles among the top-performing genotypes. Biotype 260, which exhibited the highest artemisinin content (3.006 mg/g DW), demonstrated amplification for nearly all key genes, including *FPS*, *TTG1*, *Cdsaahdr*, *ADS*, *CYP71AV1*, *DXR*, *AMDS*, *DBR2*, and *ALDH1*, with only *HMGR* absent. Its broad gene expression pattern suggests a strong and coordinated activation of the artemisinin biosynthetic pathway. Biotype 316, with similarly high metabolite output (2.752 mg/g DW), showed an extensive marker presence profile, mirroring that of biotype 260 but with a slightly reduced amplification range. It showed strong signals for *FPS*, *Cdsaahdr*, *ADS*, *CYP71AV1*, *DXR*, *HMGR*, *AMDS*, *DBR2*, and *ALDH1*, underscoring its molecular potential for elevated secondary metabolite production. Interestingly, biotype 328, despite ranking third in artemisinin production (2.765 mg/g DW), displayed a significantly reduced marker profile, with only *DXR* amplified. This suggests that post-transcriptional or epigenetic factors, or alternate biosynthetic routes, may

influence its metabolite synthesis. Biotype 258, while exhibiting relatively high artemisinin levels (2.281 mg/g DW), showed a limited molecular signature with amplification of *TTG1* alone, indicating the presence of a critical gene but perhaps compensatory regulation elsewhere in the pathway. Finally, biotype 306, although included in the top five based on artemisinin content (2.427 mg/g DW), lacked amplification of all tested markers, highlighting a potentially divergent or epigenetically regulated route to metabolite accumulation, or possibly technical limitations in marker binding or DNA quality. Overall, the radar visualization emphasizes the genetic heterogeneity underlying artemisinin biosynthesis. While some high-yielding biotypes are marked by strong multi-gene amplification profiles, others exhibit limited marker presence, suggesting multiple regulatory layers or alternative metabolic mechanisms driving artemisinin biosynthesis. This analysis identifies biotypes 260 and 316 as elite candidates for further molecular breeding or metabolic engineering due to their integrated marker-metabolite strengths.

Discussion

The gene-specific STS markers proved highly effective for distinguishing genetic variation among *Artemisia annua* biotypes and linking this variation to artemisinin production. All markers yielded clear PCR products in most biotypes, indicating good amplification efficiency. Importantly, a majority of the STS loci were polymorphic, demonstrating allelic variation in these artemisinin biosynthetic genes across genotypes. This result is consistent with earlier diversity studies using random markers, which also reported very high genetic polymorphism in *A. annua* populations [15]. The presence of polymorphic bands at key gene loci underscores the utility of STS markers in revealing genetic diversity directly tied to artemisinin biosynthesis, rather than merely neutral variation [16]. Such targeted markers are inherently more informative for our purposes because they assay variation in known functional genes, in contrast to anonymous markers like RAPD or ISSR [17]. In practical terms, the robust polymorphisms observed (e.g. presence/absence or size variants of PCR bands) allowed us to differentiate biotypes and construct a genetic profile for each [18]. This genetic profiling forms the basis for downstream analyses, providing insight into how genetic makeup correlates with phytochemical traits [19].

Clustering analysis of the marker data further highlighted the utility of these STS markers. The UPGMA dendrogram derived from genetic similarity grouped the biotypes into distinct clusters that were not arbitrary – instead, these groupings bore a relationship to the plants' artemisinin content. In our study, high-artemisinin yielding genotypes tended to cluster together, separate from low-yielding ones, indicating that the molecular marker profiles capture the underlying genetic determinants of chemotype. This finding is in agreement with previous reports showing that chemotypic (phytochemical) diversity in *A. annua* is encompassed within its genetic diversity [15,20,21]. In essence, genetically similar plants had similar artemisinin output, suggesting that the STS markers we used are associated with the biosynthetic capacity of each genotype [22]. Principal component analysis (PCA) provided complementary evidence: the major principal components summarizing marker variation also distinguished high-producing and low-producing accessions [23]. We observed that the first two principal components explained a substantial portion of the genetic variance and separated biotypes roughly according to

artemisinin content (with high-yielding biotypes scoring strongly in one direction, and low-yielding ones in the opposite). This genotype–metabolite association implies that specific marker alleles (and the genes they represent) contribute to the metabolic phenotype. Collectively, the clustering and PCA results validate these gene-specific markers as powerful tools for genotype classification and hint that particular allelic configurations of artemisinin pathway genes are requisite for high yield. Such findings reinforce the notion that molecular markers can be reliably used to predict or at least screen for desirable chemotypes in breeding programs [24].

The presence or absence of a gene-specific STS marker in a given biotype provides insight into the underlying gene expression potential of that plant. In general, if a PCR marker for a biosynthetic gene is present (i.e. the expected band amplifies), it indicates that the gene's coding sequence (or a targeted portion of it) is intact in that genotype [25]. We found that all high-artemisinin producers harbored the full complement of markers corresponding to key pathway enzymes (such as *ADS*, *CYP71AV1*, *DBR2*, and *ALDH1*), whereas some low producers showed null alleles or divergent alleles for one or more of these genes (e.g. failure to amplify a marker or an altered band size). The absence of an STS marker band in a low-yielding biotype suggests a possible disruption or significant mutation in that gene – a factor that could lead to reduced transcript levels or non-functional protein. For instance, one low-artemisinin genotype in our study lacked the expected band for a *DBR2* promoter-region marker, hinting at a deletion in that regulatory region. This is particularly noteworthy because *DBR2* (artemisinic aldehyde Δ11(13) reductase) is known to be a rate-limiting enzyme for artemisinin biosynthesis, funneling pathway flux toward the end product. A comparable phenomenon has been documented in the literature: low-artemisinin chemotypes (LAPs) of *A. annua* were found to carry a large promoter deletion in the *DBR2* gene, resulting in significantly lower *DBR2* expression than in high-artemisinin chemotypes (*HAPs*) [26]. In those studies, the *DBR2* promoter deletion was directly correlated with reduced enzyme transcript and consequently a bottleneck in artemisinin production [26]. Our PCR-based marker evidence of putative *DBR2* promoter variation in low-yielding plants aligns with these findings, suggesting that *DBR2* expression (or lack thereof) is a major determinant of whether a plant accumulates artemisinin or diverts precursors into non-artemisinin side-products.

Beyond *DBR2*, our results and prior studies indicate that high-artemisinin producers generally express the entire suite of artemisinin pathway genes at higher levels than low producers. Although we did not measure mRNA levels directly in this study, the strong correlation between marker profiles and metabolite levels implies underlying expression differences. In support of this, earlier work has shown that transcript levels of *ADS*, *CYP71AV1*, *DBR2*, and *ALDH1* are markedly higher in high-yielding *A. annua* varieties compared to low-yielding ones [27]. The genetic basis for such expression differences is becoming clearer. For example, recent genomic analysis revealed that high-artemisinin genotypes often carry multiple copies of the *ADS* gene, and artemisinin content was positively correlated with *ADS* copy number across diverse individuals [28]. Having extra copies of *ADS* likely boosts the production of amorpha-4,11-diene (the first committed step in artemisinin biosynthesis), thereby increasing the metabolic flux into the pathway [9]. Likewise, allele variations of *CYP71AV1* have been observed; one

study found that HAP-type *CYP71AV1* has a truncated N-terminus (missing seven amino acids) leading to a less stable enzyme that may actually favor artemisinin formation by releasing more substrate for *DBR2* conversion [29]. These examples illustrate that both structural gene dosage and allele sequence can influence enzyme expression or activity, which in turn impacts metabolite accumulation [24]. In our biotypes, any marker polymorphisms of this nature – such as an extra band or a novel-sized fragment – could reflect copy number variation or indels in these genes. Indeed, one high-yielding accession in our panel showed an unusually intense band for the *ADS* marker, which might correspond to multiple *ADS* loci (a hypothesis consistent with the aforementioned copy number effect).

The ultimate consequence of these genetic and expression differences is evident in the artemisinin content and its precursor profile in each biotype. High-artemisinin genotypes not only produced more artemisinin, but also showed marker profiles indicative of a well-functioning “metabolic pipeline” from precursor to product. In contrast, low-artemisinin genotypes often accumulated upstream intermediates or shunted metabolites, reflecting a block at some point in the pathway. This was exemplified in our study by the chemical analysis: high producers had significantly greater amounts of dihydroartemisinic acid (the direct precursor to artemisinin) and artemisinin itself, whereas low producers accumulated higher levels of artemisinic acid, artemisitene, and other byproducts. This trend mirrors published observations for *HAP* vs. *LAP* chemotypes [26]. In one report, a high-artemisinin line (HAP) contained ~ 25-fold more dihydroartemisinic acid and several-fold more artemisinin than a low line, whereas the low line amassed vastly more artemisinic acid and arteannuin B (a side-product) [30]. Such metabolite profiles make biochemical sense: if *DBR2* or *ALDH1* activity is insufficient (as in low producers), artemisinic aldehyde is not fully converted to dihydroartemisinic aldehyde and instead is oxidized to artemisinic acid (which mostly cannot turn into artemisinin). Likewise, a less active or lower-copy *ADS* may leave more farnesyl diphosphate to go into alternative terpene pathways, or simply produce less amorpha-4,11-diene to begin with. Our marker analysis, combined with known gene-expression correlations, strongly supports that the genotypes with all the “right” alleles (complete pathway gene set, no major deletions, possibly higher copy number of key genes) are the ones that achieve high artemisinin accumulation.

It is also worth noting that gene expression in the artemisinin pathway is under complex regulation by developmental and environmental cues, as well as by specialized transcription factors. All the biotypes in our study were grown under similar conditions to minimize environmental variation in artemisinin content, so the differences we observed are predominantly genetic. However, natural genetic variation could include differences in regulatory genes that our STS markers did not directly assay. For example, *AaORA* (octadecanoid-responsive AP2-domain transcription factor) is a positive regulator of multiple artemisinin biosynthetic genes [2]. Overexpression of *AaORA* and other trichome-specific activator genes has been shown to increase artemisinin (and dihydroartemisinic acid) levels in transgenic plants [31,32]. Conversely, factors like *AaWRKY1* and *AaWRKY2* (or *AaGSW2*) can also influence glandular trichome density and pathway gene expression [31,33]. While our gene-specific markers focused on structural pathway genes, future work could include markers for such regulatory genes to see if particular alleles (e.g. a high-expression variant of a positive regulator, or absence of a negative regulator) correlate with

the high-artemisinin trait. In summary, our findings indicate a strong genotypic basis for differential gene expression and metabolite accumulation: *A. annua* biotypes with “complete” and presumably highly expressed artemisinin biosynthesis genes are genetically predisposed to produce more artemisinin, whereas those with incomplete or compromised gene complements produce less. This genotype–phenotype linkage is precisely what our STS marker approach was designed to uncover.

Multivariate analyses in our study reinforced the association between genetic marker profiles and artemisinin metabolic phenotype. The hierarchical cluster analysis, as mentioned, partitioned the biotypes in a manner that largely reflected their artemisinin content categories. High-yield clusters were composed of genotypes sharing certain marker alleles (for instance, all members of the high-yield cluster had the *ADS* and *DBR2* markers present in a specific configuration), whereas low-yield clusters often shared the absence or altered form of one or more markers. Mapping the artemisinin content data onto the dendrogram revealed a clear genotype–metabolite correspondence: most of the high-artemisinin plants fell into one clade, separate from the clade containing nearly all the lowest producers. A similar congruence between genetic distance and chemotypic difference was noted by Sangwan et al. [15], who found that *A. annua* individuals cluster by chemotype within the broader genetic diversity. Our results echo this pattern, now using targeted gene markers: genetically, the high producers are more closely related to each other than to the low producers, suggesting that they share common alleles contributing to artemisinin biosynthesis capacity.

The PCA offered additional support for genotype–metabolite linkages. In our PCA biplot, we included both marker variables and the artemisinin content as supplementary data. Notably, principal component 1 (PC1) had high loadings for several key gene markers and was positively associated with artemisinin content. This indicates that the major axis of genetic variation captured by our markers is aligned with variation in artemisinin yield. Genotypes with high PC1 scores possessed most of the favorable alleles (and indeed were the high artemisinin producers), whereas those with low PC1 scores lacked one or more of these alleles and had low artemisinin. In effect, PC1 differentiated “metabolically complete” genotypes from “incomplete” ones. Principal component 2 (PC2) appeared to separate genotypes based on more subtle differences, possibly distinguishing intermediate chemotypes or capturing variation in secondary pathways (e.g. essential oil related markers, if any were indirectly amplified, or minor allelic differences that did not hugely affect artemisinin levels). The important point is that the ordination of genotypes in PCA space was not random with respect to artemisinin content; instead, there was a discernible gradient, reinforcing that our STS markers are tapping into the genetic factors underlying artemisinin accumulation.

One can draw parallels between our findings and QTL (quantitative trait locus) mapping studies that have been done on *A. annua*. For instance, a high-density genetic map of *A. annua* identified several QTLs accounting for a significant proportion of the variation in artemisinin yield [22]. Those QTLs correspond to genomic regions harboring genes influencing artemisinin biosynthesis, and it was demonstrated that breeding lines enriched for the positive QTL alleles achieved higher artemisinin content [22]. In our study, although we did not explicitly map QTLs, the clustering of high-yield genotypes and the correlation of

marker patterns with content suggest that our chosen gene markers may lie within or very near some of these important QTLs. Essentially, the multivariate analyses serve to connect the dots: they show that variation at a handful of biosynthetic genes can explain much of the phenotypic variance (at least among our tested biotypes). This insight is valuable for guiding breeding and selection, as it implies that by selecting for a specific combination of marker alleles, one is indirectly selecting for the high-yield metabolic profile.

Our results have significant implications for breeding and improving *A. annua* as a commercial source of artemisinin. The clear association between certain marker profiles and high artemisinin content means that these STS markers can be deployed in marker-assisted selection (MAS) programs. Breeders can screen seedlings or breeding populations for the presence of key markers (for example, a complete set of artemisinin biosynthesis genes without deletions, and potentially indicators of multi-copy gene loci) to identify individuals with the best genetic potential *before* field testing for artemisinin content. This approach accelerates the breeding cycle by allowing early culling of inferior genotypes. Indeed, marker-assisted breeding has already proven effective in *A. annua*: an Indian breeding program developed the high-yield cultivar CIM-Arogya by recurrently selecting plants carrying a favorable DNA marker linked to artemisinin content [34]. Over four generations, this strategy raised mean artemisinin levels from 0.15% to about 1.16% (a nearly eight-fold improvement), while also fixing a distinctive genetic fingerprint in the new cultivar [35]. Notably, the DNA marker (termed “MAP 12”) was present in all high-yield selections each cycle, illustrating how a marker allele can track with the trait of interest [34]. In our case, the gene-specific markers we have validated could play a similar role – for example, a breeder might require that prospective parent plants show the “high-artemisinin” band pattern (including, say, an intact *DBR2* promoter marker and an amplified *ADS* marker) as a prerequisite for advancement. By doing so, one can enrich the breeding population with genotypes that have the known genetic determinants for high artemisinin, thus stacking the odds in favor of recovering superior cultivars.

Beyond traditional breeding, the knowledge gained from our molecular analysis can inform genetic engineering and biotechnological interventions aimed at increasing artemisinin yield. Identifying which genes are limiting in low-producing biotypes opens up targeted solutions: for instance, if a natural genotype lacks a functional *DBR2*, one could complement that via transgenic expression of a high-efficiency *DBR2* allele. Conversely, if a genotype has all pathway genes but still produces suboptimal artemisinin, it might benefit from overexpression of a positive regulator or the silencing of a negative regulator. Prior successes in transgenic *A. annua* demonstrate the promise of such approaches – overexpression of single pathway enzymes like *HMGR*, *FPS*, or *DBR2* can roughly double artemisinin content in planta [36–38]. For example, inserting an extra copy of *HMGR* (encoding 3-hydroxy-3-methylglutaryl CoA reductase, an upstream terpenoid pathway enzyme) from *Catharanthus* yielded a ~ 50% increase in artemisinin in *A. annua* [32]. Stacking multiple genes has an even more pronounced effect: a recent study showed that co-overexpressing three biosynthetic genes (*HMGR*, *FPS*, *DBR2*) along with two trichome development transcription factors led to transgenic lines with significantly elevated glandular trichome density and artemisinin content in both T0 and T1 generations [31,33]. Similarly,

overexpressing a transcription factor *AaWRKY1* that upregulates *CYP71AV1* increased artemisinin up to 1.8-fold [39], and overexpressing the AP2/ERF factor *AaORA* under a trichome-specific promoter enhanced artemisinin and dihydroartemisinic acid accumulation[31]. These cutting-edge biotechnological strategies complement our marker-based findings: whereas we detect naturally occurring genetic advantages (or deficiencies) in our biotypes, genetic engineering allows us to introduce or amplify those advantageous traits in any genotype. In the future, breeders might combine both approaches – for instance, use MAS to assemble a plant with all known favorable alleles, then use CRISPR or transgenes to tweak regulatory networks for an extra boost in production.

The relevance of our findings extends to securing artemisinin supply and reducing production costs. *A. annua* remains the primary commercial source of artemisinin, as chemical synthesis is currently not cost-effective at scale [15]. Even with advances in semi-synthesis (from engineered yeast), plant-derived artemisinin is expected to remain indispensable due to lower cost and infrastructure requirements [32]. Therefore, improving the plant itself is a top priority. By linking genetic markers to metabolite output, we enhance the precision of breeding for this trait, akin to how MAS has improved yield and stress traits in major crops [40]. The superior biotypes identified (and developed) through these molecular techniques can be cultivated to provide a more consistent and higher yield of artemisinin per hectare, ultimately making artemisinin-based combination therapies more affordable and reliable. Additionally, conservation of genetic resources and elite lines is facilitated by DNA markers: one can verify the identity and purity of high-artemisinin cultivars (like CIM-Arogya or others) by checking their marker profile, ensuring farmers receive the correct seed. In our study, the distinct marker signatures of high producers could serve as a "molecular tag" for desirable germplasm.

Finally, our results contribute to the broader scientific understanding of secondary metabolite regulation in plants. They underscore that significant natural variation exists even in critical pathways like that of artemisinin, and that this variation can be exploited. Similar gene-specific marker approaches could be applied to other medicinal plants to find genetic predictors of high metabolite yield. The integration of molecular markers, gene expression insights, and metabolite analysis – as exemplified here – provides a framework for dissecting complex biosynthetic traits. It is a reminder that both structural genes and regulatory elements co-evolve to produce chemotypic diversity, and by studying both, we gain leverage to direct these pathways for human benefit. In conclusion, the molecular analysis of *A. annua* biotypes using gene-targeted STS markers not only illuminates the genotype–phenotype relationships driving artemisinin biosynthesis, but also equips us with practical tools to breed and engineer better cultivars. These advances come at a crucial time, as global demand for artemisinin remains high and continuous improvement of this medicinal crop is needed to sustain the fight against malaria [15,32].

Conclusion

This study successfully identified genetic diversity and metabolite associations among 17 *Artemisia annua* biotypes using gene-specific sequence-tagged site (STS) markers linked to artemisinin

biosynthesis. The presence and absence of key markers such as *ADS*, *DBR2*, *CYP71AV1*, and *ALDH1* showed strong correlation with artemisinin content, as quantified by HPLC. Multivariate analyses, including cluster analysis and principal component analysis (PCA), revealed clear grouping of high- and low-yielding biotypes, confirming a strong genotype–metabolite association. Marker efficiency indices demonstrated that certain markers, notably *TTG1*, *DXR*, and *ADS*, provided high polymorphic information and discriminative power among genotypes. The radar chart and correlation heatmaps further emphasized the relevance of specific markers for identifying elite biotypes. Overall, these results highlight the potential of STS markers in marker-assisted selection (MAS) and metabolic improvement strategies for artemisinin yield enhancement. This integrated molecular and metabolite profiling approach provides a robust platform for both conventional breeding and biotechnological interventions aimed at improving *A. annua* as a sustainable source of artemisinin.

Declarations

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.

Author Contribution

Mohammad Taher Hallajian: Writing-original draft, Investigation, Data curation. Sahar Tavakoli Mohammadi: Writing-review & editing, Funding acquisition. Mohammad Ali Ebrahimi: Investigation, Visualization, Methodology. Mohammad Reza Naghavi: Methodology, Writing, review, and editing. Mojtaba Kordrostami: Conceptualization, Supervision, Writing-review & editing.

Acknowledgement

The authors extend their sincere gratitude to Negar Bazrafkan for her meticulous English proofreading of this manuscript.

Data Availability

The datasets generated and/or analyzed during the current study are available from the corresponding author, Dr. Mojtaba Kordrostami (mkordrostami@aeoi.org.ir), upon reasonable request.

References

1. Manisha, D. R. B., Begam, A. M., Chahal, K. S. & Ashok, M. A. Medicinal Plants and Traditional Uses and Modern Applications. *Journal Neonatal Surgery* **14** (2025).
2. Salehi, M., Karimzadeh, G., Naghavi, M. R. & Naghdi Badi, H. Rashidi Monfared, S. Expression of key genes affecting artemisinin content in five Artemisia species. *Sci. Rep.* **8**, 12659 (2018).
3. Morua, E., Cuyas, L. & Matías-Hernández, L. The Beneficial Use of Artemisia annua, Artemisinin, and Other Compounds in Animal Health. *Animals* **15**, 1359 (2025).
4. Fairhurst, R. M. & Dondorp, A. M. Artemisinin-resistant Plasmodium falciparum malaria. *Microbiol. Spectr.* **4** 10.1128/microbiolspec. ei10-0013-2016 (2016).
5. Liu, K. University of Macau (2021).
6. Qamar, F., Ashrafi, K., Singh, A., Dash, P. K. & Abdin, M. Artemisinin production strategies for industrial scale: Current progress and future directions. *Ind. Crops Prod.* **218**, 118937 (2024).
7. Wetzstein, H. Y., Porter, J. A., Janick, J., Ferreira, J. F. & Mutui, T. M. Selection and clonal propagation of high artemisinin genotypes of Artemisia annua. *Front. Plant Sci.* **9**, 358 (2018).
8. Wani, K. I. et al. Enhancing artemisinin content in and delivery from Artemisia annua: a review of alternative, classical, and transgenic approaches. *Planta* **254**, 1–15 (2021).
9. Huang, J. Q. & Fang, X. Amorpha-4, 11-diene synthase: a key enzyme in artemisinin biosynthesis and engineering. *Abiotech* **2**, 276–288 (2021).
10. Zhao, L. et al. From plant to yeast—Advances in biosynthesis of artemisinin. *Molecules* **27**, 6888 (2022).
11. Halder, M., Majumder, A., Ray, S. & Jha, S. Medicinal plant research at crossroads: biotechnological approaches for conservation, production and stability in tissue cultures and regenerated plants. *Medicinal Plants: Domestication Biotechnol. Reg. Importance*, 459–544 (2021).
12. Grover, A. & Sharma, P. Development and use of molecular markers: past and present. *Crit. Rev. Biotechnol.* **36**, 290–302 (2016).
13. Doveri, S., Lee, D., Maheswaran, M. & Powell, W. Molecular markers-history, features and applications. in *Principles and Practices of Plant Genomics, Vol. 1* 23–67CRC Press, (2024).
14. Shahrivari, S., Alizadeh, S., Ghassemi-Golezani, K. & Aryakia, E. A comprehensive study on essential oil compositions, antioxidant, anticholinesterase and antityrosinase activities of three Iranian Artemisia species. *Sci. Rep.* **12**, 7234 (2022).
15. Sangwan, R., Sangwan, N., Jain, D., Kumar, S. & Ranade, S. RAPD profile based genetic characterization of chemotypic variants of Artemisia annua L. *IUBMB Life*. **47**, 935–944 (1999).
16. Kayiba, N. K. et al. Spatial and molecular mapping of Pfkelch13 gene polymorphism in Africa in the era of emerging Plasmodium falciparum resistance to artemisinin: a systematic review. *Lancet. Infect. Dis*. **21**, e82–e92 (2021).
17. Biswas, P. & Kumar, N. Application of molecular markers for the assessment of genetic fidelity of in vitro raised plants: current status and future prospects. *Molecular Marker Techniques: Potential. Approach Crop Improvement*, 233–256 (2023).

18. Tarang, A. et al. Study of genetic diversity in rice (Oryza sativa L.) cultivars of Central and Western Asia using microsatellite markers tightly linked to important quality and yield related traits. *Genet. Resour. Crop Evol.* **67**, 1537–1550 (2020).
19. Kordrostami, M., Rabiei, B. & Hassani Kumleh, H. Association analysis, genetic diversity and haplotyping of rice plants under salt stress using SSR markers linked to SalTol and morpho-physiological characteristics. *Plant Syst. Evol.* **302**, 871–890 (2016).
20. Shinyuy, L. M. et al. Exploring the Phytochemical Diversity and Anti-Plasmodial Potential of Artemisia annua and Artemisia afra from Different Geographical Locations in Cameroon. *Molecules* **30**, 596 (2025).
21. Naghavi, M. et al. Characterization of morphological, phytochemical and molecular diversity of Artemisia annua accessions in Hyrcanian area of Iran. *J. Agricultural Sci. Technol.* **21**, 1265–1276 (2019).
22. Graham, I. A. et al. The genetic map of Artemisia annua L. identifies loci affecting yield of the antimalarial drug artemisinin. *science* 327, 328–331 (2010).
23. Kordrostami, M., Sanjarian, F., Shahbazi, S. & Ghasemi-Soloklui, A. A. Exploring low-dose gamma radiation effects on monoterpene biosynthesis in Thymus vulgaris: insights into plant defense mechanisms. *Environmental Sci. Pollution Research*, 1–21 (2024).
24. Ghasemi, S., Kumleh, H. H. & Kordrostami, M. Changes in the expression of some genes involved in the biosynthesis of secondary metabolites in Cuminum cyminum L. under UV stress. *Protoplasma* **256**, 279–290 (2019).
25. Poczai, P. et al. Advances in plant gene-targeted and functional markers: a review. *Plant. methods*. **9**, 1–32 (2013).
26. Wang, X. et al. Promoter variations in DBR2-like affect artemisinin production in different chemotypes of Artemisia annua. *Hortic. Res.* **10**, uhad164 (2023).
27. Kam, M. Y. Y. University of Nottingham (2021).
28. Liao, B. et al. Allele-aware chromosome-level genome assembly of Artemisia annua reveals the correlation between ADS expansion and artemisinin yield. *Mol. Plant*. **15**, 1310–1328 (2022).
29. Teoh, K. H., Polichuk, D. R., Reed, D. W., Nowak, G. & Covello, P. S. Artemisia annua L.(Asteraceae) trichome-specific cDNAs reveal CYP71AV1, a cytochrome P450 with a key role in the biosynthesis of the antimalarial sesquiterpene lactone artemisinin. *FEBS Lett.* **580**, 1411–1416 (2006).
30. Guo, Z. et al. Dihydroartemisinic acid dehydrogenase-mediated alternative route for artemisinin biosynthesis. *Nat. Commun.* **16**, 3888 (2025).
31. Lu, X. et al. A a ORA, a trichome-specific AP 2/ERF transcription factor of A rtemisia annua, is a positive regulator in the artemisinin biosynthetic pathway and in disease resistance to B otrytis cinerea. *New Phytol.* **198**, 1191–1202 (2013).
32. Hassani, D. et al. Elevation of artemisinin content by co-transformation of artemisinin biosynthetic pathway genes and trichome-specific transcription factors in Artemisia annua. *Front. Plant Sci.* **14**, 1118082 (2023).

33. Yan, T. et al. HOMEODOMAIN PROTEIN 1 is required for jasmonate-mediated glandular trichome initiation in Artemisia annua. *New Phytol.* **213**, 1145–1155 (2017).
34. Paul, S. et al. Enhancement of artemisinin content through four cycles of recurrent selection with relation to heritability, correlation and molecular marker in Artemisia annua L. *Planta Med.* **76**, 1468–1472 (2010).
35. Kraft, V., Kretzschmar, G. & Rossen, K. *Process for the production of artemisinin intermediates* (Google Patents, 2012).
36. Yuan, Y. et al. Overexpression of artemisinic aldehyde Δ11 (13) reductase gene–enhanced artemisinin and its relative metabolite biosynthesis in transgenic A rtemisia annua L. *Biotechnol. Appl. Chem.* **62**, 17–23 (2015).
37. Aquil, S., Husaini, A. M., Abdin, M. Z. & Rather, G. M. Overexpression of the HMG-CoA reductase gene leads to enhanced artemisinin biosynthesis in transgenic Artemisia annua plants. *Planta Med.* **75**, 1453–1458 (2009).
38. Banyai, W., Kirdmanee, C., Mii, M. & Supaibulwatana, K. Overexpression of farnesyl pyrophosphate synthase (FPS) gene affected artemisinin content and growth of Artemisia annua L. *Plant. Cell. Tissue Organ. Cult. (PCTOC)*. **103**, 255–265 (2010).
39. Nafis, T. et al. Enhancement of artemisinin content by constitutive expression of the HMG-CoA reductase gene in high-yielding strain of Artemisia annua L. *Plant. Biotechnol. Rep.* **5**, 53–60 (2011).
40. Hasan, N., Choudhary, S., Naaz, N., Sharma, N. & Laskar, R. A. Recent advancements in molecular marker-assisted selection and applications in plant breeding programmes. *J. Genetic Eng. Biotechnol.* **19**, 128 (2021).

Figures

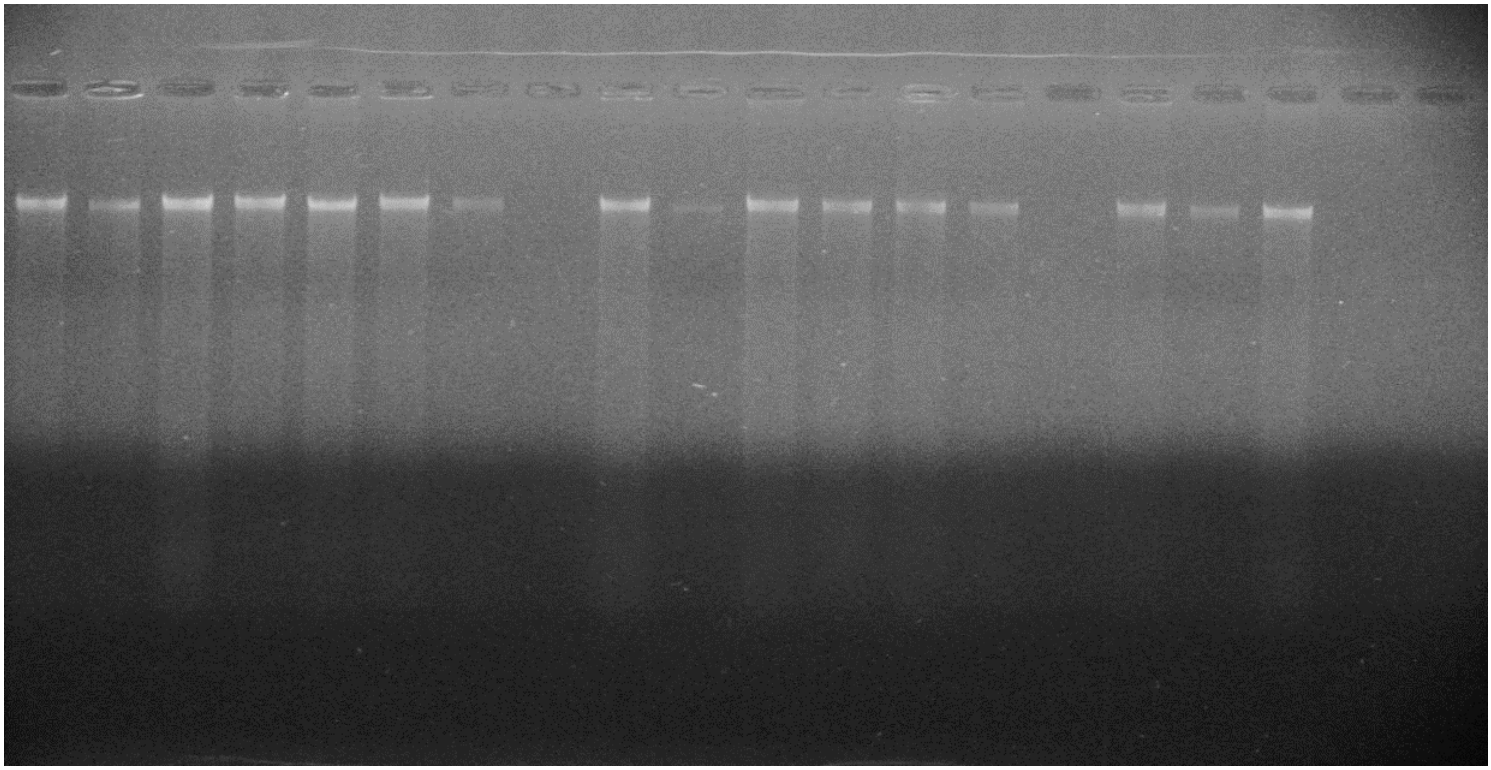


Figure 1

Agarose gel electrophoresis of genomic DNA extracted from 17 *Artemisia annua*biotypes. Each lane represents a distinct biotype. The image includes bands cropped from a single gel. White space has been inserted between groups of lanes to delineate non-adjacent lanes from the same gel. No splicing from different gels or exposures was performed. The full, uncropped gel image is provided in Supplementary Figure S1..

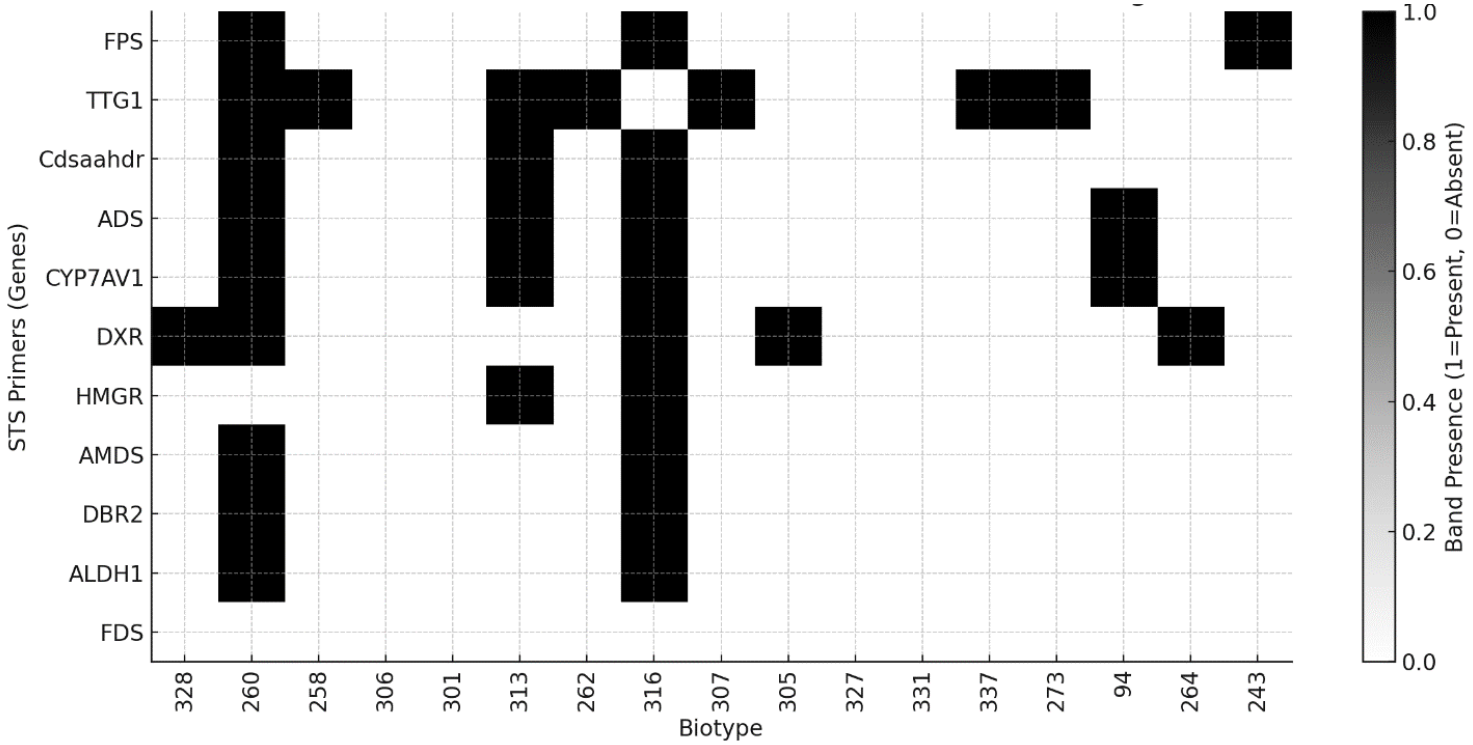


Figure 2

Allelic amplification pattern of artemisinin biosynthetic genes across *17*Artemisia annua *biotypes using gene-specific STS primers.* PCR amplification was performed using eleven primer pairs targeting key genes involved in the artemisinin biosynthetic pathway, including *ADS, CYP71AV1, DBR2, ALDH1, HMGR, FPS, DXR, AMDS, Cdsaahdr*, and *TTG1*. The presence (+) or absence (−) of bands for each marker is indicated for every biotype, revealing inter-biotype polymorphism and differential gene amplification. This pattern reflects underlying genetic diversity and potential associations with artemisinin production capacity.

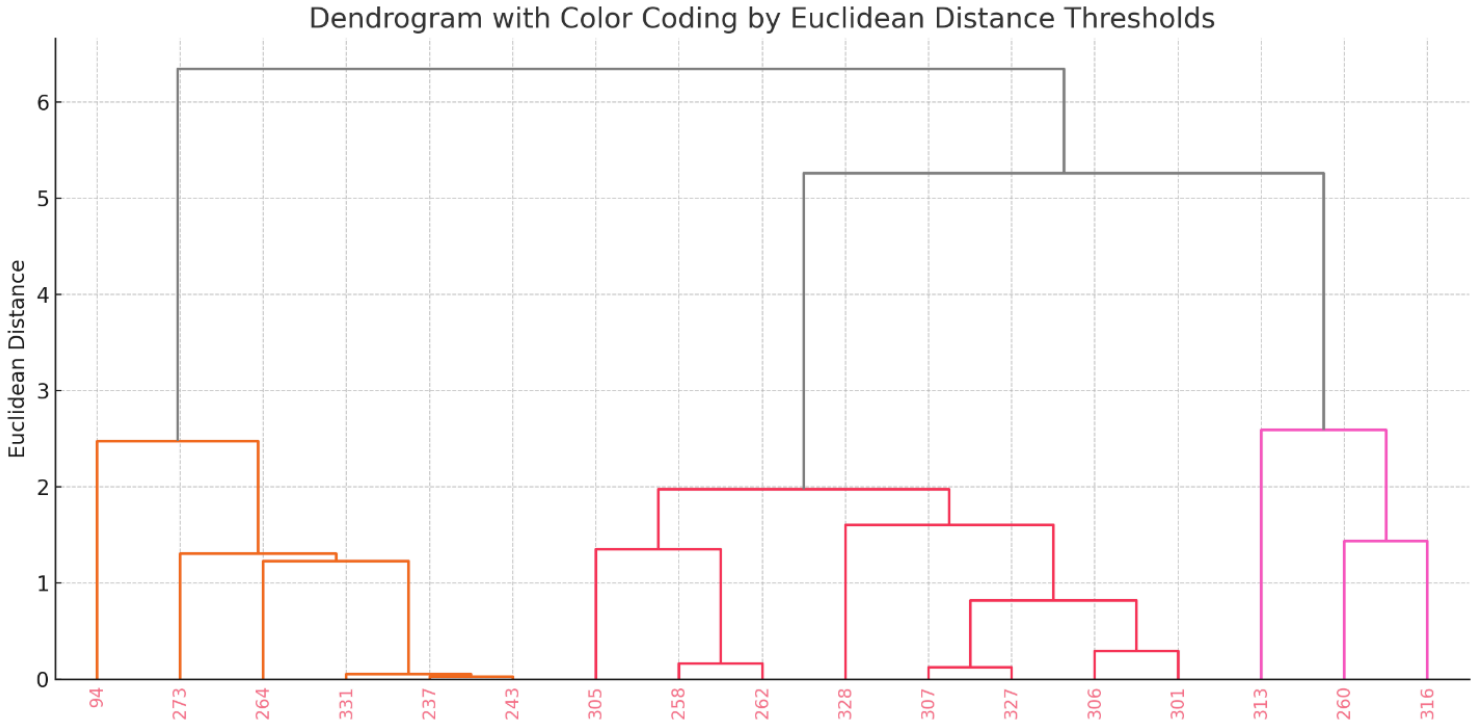


Figure 3

Hierarchical cluster analysis of *Artemisia annua* genotypes based on STS marker profiles and artemisinin content. The dendrogram was constructed using Euclidean distance as the similarity metric on a matrix combining presence/absence data from gene-specific STS markers and artemisinin content values, grouping the genotypes according to overall genetic similarity.

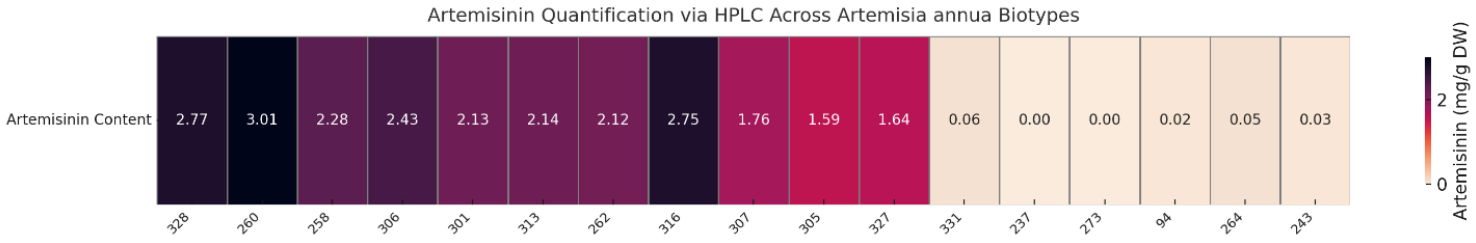


Figure 4

Heatmap representation of artemisinin quantification via HPLC in different *Artemisia annua* biotypes. The color gradient reflects the relative concentration of artemisinin (mg/g dry weight), highlighting variation in metabolite accumulation across the 17 evaluated genotypes.

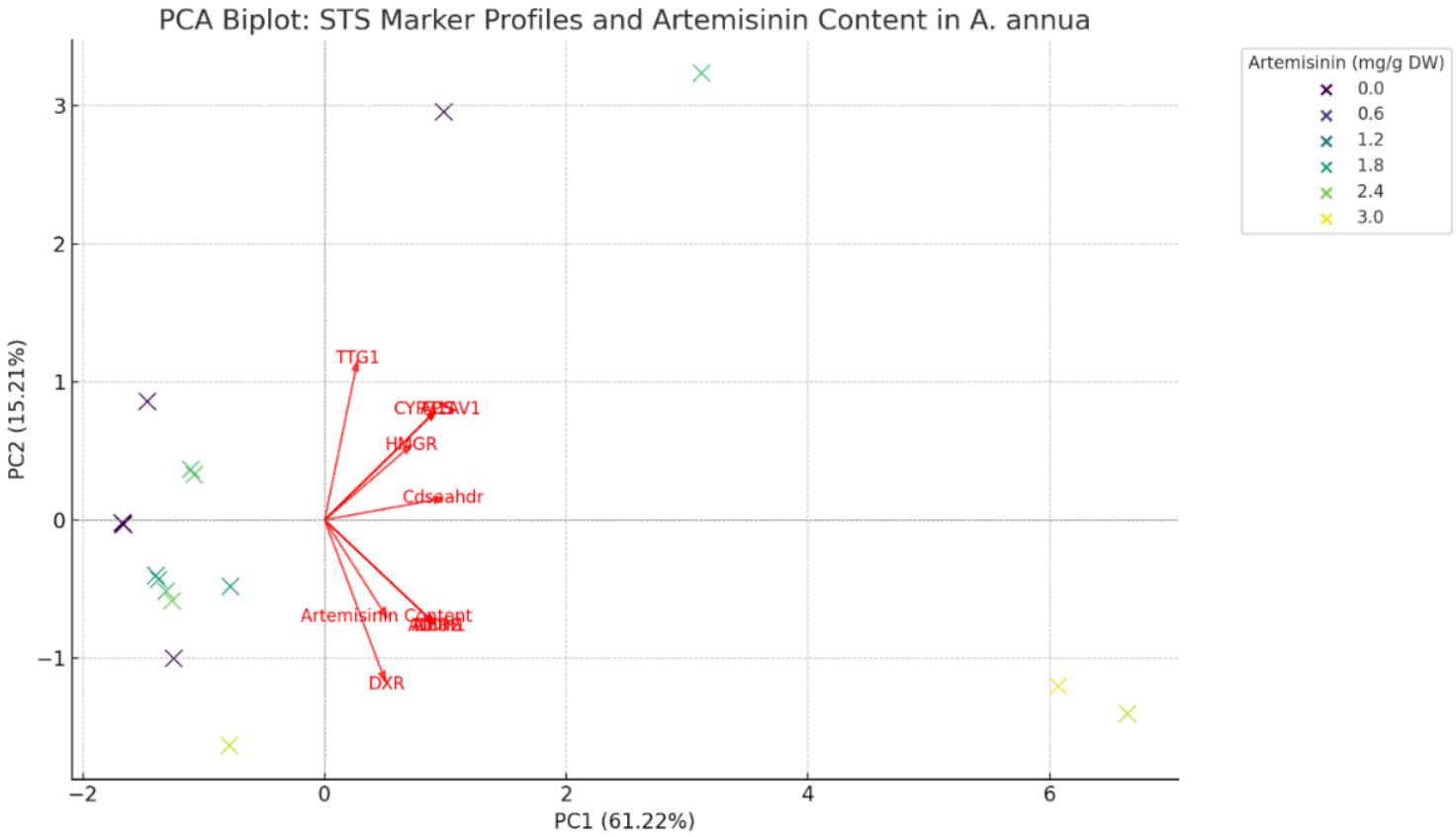


Figure 5

Principal Component Analysis (PCA) integrating gene-specific STS marker profiles and artemisinin content across *Artemisia annua* biotypes. The biplot illustrates the distribution of genotypes and the relative contribution of each marker and artemisinin content to the observed variance. Arrows represent loading vectors for each variable, allowing visualization of genotype–metabolite associations.

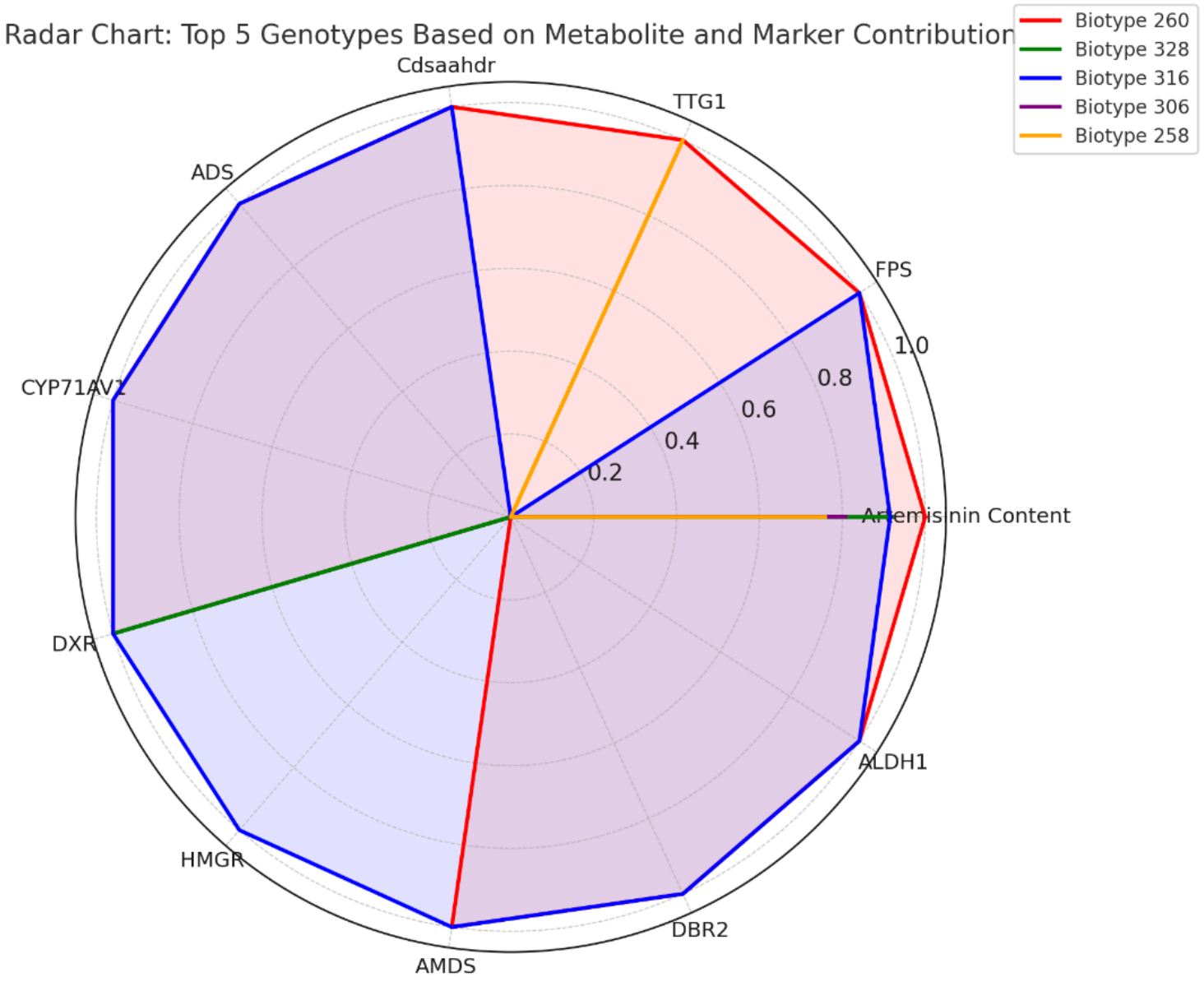


Figure 6

Radar chart depicting multitrait profiles of the top five *Artemisia annua* biotypes based on artemisinin content and STS marker amplification. Each axis represents a trait (artemisinin concentration or gene-specific marker), while the enclosed areas reflect the normalized performance of each biotype, highlighting genotypes with superior metabolite accumulation and genetic marker expression.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTableS1.pdf
- SupplementaryFigureS1.pdf