Tolangi *et al. Applied Biological Chemistry* (2024) 67:115
https://doi.org/10.1186/s13765-024-00967-7

KSABC
The Korean Society for Applied Biological Chemistry

INVITED REVIEW **Open Access**

The genetics and genomics of milk thistle: unlocking its therapeutic potential through modern breeding and biotechnological innovations

Priskila Tolangi[1,5†], Jeehyoung Shim[1,5†], Raña Mae Sumabat[1,5], Sunghan Kim[1,5], Hyun-Seung Park[1,5], Kyung Do Kim[1,5], Hyun Uk Kim[2], Sanghyun Lee[3,4*] and Joong Hyoun Chin[1,5*]

Abstract

Milk thistle (*Silybum marianum*) is a Mediterranean herb renowned for its liver-protective, antioxidant, anti-inflammatory, and detoxifying properties, primarily attributed to the bioactive compound silymarin. Recent studies have also highlighted its potential efficacy against COVID-19, contributing to the growing demand for milk thistle dietary supplements, particularly for liver health and immunity support. Milk thistle seeds, rich in silymarin and unsaturated fatty acids, hold significant industrial value as both medicinal and oilseed crops. To meet the growing demand, it is essential to develop standardized seeds, cultivation practices, and extraction methods aimed at maximizing yields of silymarin and other valuable metabolites. Recent advancements in genetic and genomic research, including the development of the first reference genome of *S. marianum*, have played a pivotal role in elucidating the biosynthesis pathways of silymarin and optimizing phytochemical production. This review highlights recent advancements in the genetics, genomics, and biochemistry of milk thistle, with particular emphasis on the importance of diverse genetic resources and AI-driven phenomics strategies, such as hyperspectral and RGB imaging, for high-yield and chemotype breeding. Further, feasibility of developing elite cultivars through molecular approaches, such as genome editing and metabolic engineering, is also discussed as the new traits obtained this way would be key to enhancing the commercial value of milk thistle in light of mass production of phytochemicals to meet rising market demands.

Keywords Chemotype breeding, Milk thistle, Molecular marker, Naturalization, Reference genome, *Silybum marianum*, Silymarin

---

[†] Priskila Tolangi and Jeehyoung Shim have contributed equally to this work.

*Correspondence:
Sanghyun Lee
slee@cau.ac.kr
Joong Hyoun Chin
jhchin@sejong.ac.kr
Full list of author information is available at the end of the article

Introduction

Milk thistle (*Silybum marianum* (L.) Gaertn.), an herbaceous plant in the *Asteraceae* family native to the Mediterranean region, has become widely distributed across the globe, including the Korean peninsula [1–3]. For over 2000 years, *S. marianum* has been utilized as a medicinal remedy for a variety of diseases and disorders, including hepatitis, jaundice, and other liver conditions, as well as gallbladder disorders. It has also been used to treat snake bites, insect stings, plant poisoning, and alcohol-related issues [1, 4–7]. In recent years, the value of milk thistle as a dietary supplement has garnered significant public interest, particularly in response to growing concerns about liver health. According to the WHO's 2024 Global Hepatitis Report, viral hepatitis is the second leading infectious cause of death worldwide, responsible for 1.3 million deaths annually, surpassed only by tuberculosis [8]. Additionally, the current trend of an aging population structure and the experience of the COVID-19 pandemic have increased awareness of the benefits of enhanced immunity and antioxidants, further driving interest in milk thistle [9].

The primary active ingredient responsible for the medicinal properties of milk thistle is a silymarin, a complex mixture of flavonolignans, flavonoids, and polyphenols [10, 11]. Among these compounds, several flavonolignans including silychristin, silydianin, silybins A and B, and isolilybins A and B, are the primary active components responsible for most of the health benefits associated with milk thistle [12]. Consistent with the known therapeutic effects of milk thistle, silymarin has been identified not only as a highly effective antioxidant with potent anti-inflammatory and antifibrotic properties but also for its role in detoxification, liver protection, and promoting liver regeneration [11, 13–16]. Furthermore, several of these compounds have been found to modulate the host's cytokine storm and inhibit viral replication in COVID-19 infections [17].

According to a recent global market trend report, the global phytochemicals market is projected to grow at a compound annual growth rate (CAGR) of 9.8% from 2021 to 2031. Specifically, the market for milk thistle dietary supplements is expected to expand at a CAGR of 8.05% over the next decade (2024–2033), reaching an estimated value of USD 213.82 million by 2033 (Fig. 1) [18]. Given its status as an oilseed crop with a diverse array of valuable phytochemicals, the rapid market growth of milk thistle underscores the necessity for developing efficient methods for the mass production of silymarins from its seeds. It is therefore crucial to ensure a consistent supply of seeds with uniform quality and improved yields, both in terms of seed production and the desired secondary metabolites. Achieving this requires the development of diverse seed varieties, from which the most favorable traits can be selected, along with the establishment of large-scale cultivation methods and the standardization of extraction processes [3, 19, 20].

This review aims to summarize the current knowledge of the genetics, genomics, and biochemistry of milk thistle (*S. marianum*). The recently constructed reference genome of *S. marianum* provides a valuable resource for understanding the biochemical pathways involved in silymarin synthesis and for developing reliable sources of natural biochemical products. Our research has explored the genetic diversity of the key chemical components of silymarin, and we will discuss future prospects for digital phytochemical breeding.

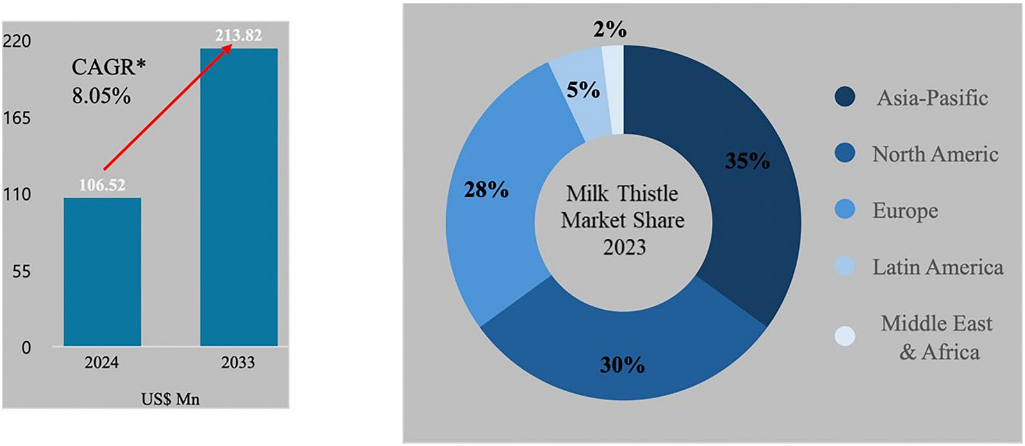


Fig. 1 Scope of the milk thistle market. Growth rate from 2024 to 2033 and regional market share in 2023. * CAGR: compound annual growth rate. *Source*: Persistence Market Research, April 2024

Key benefits of milk thistle

Phytochemicals

Silymarin is composed of 70–80% flavonolignans and 20–30% flavonoids, with the remaining portion consisting of polymeric and oxidized polyphenolics, forming a complex mixture of bioactive molecules [11]. It is primarily concentrated in the seed coat (also known as the pericarp or integument) of the seeds (achenes or fruits), accounting for approximately 1.5–4.3% of the total dry seed weight [21–23]. The main bioactive component of silymarin was first identified by Pelter and Hänsel in 1968, and in the 1970s, the WHO approved milk thistle extract as a recognized treatment for liver-related diseases.

Research into the chemistry of silymarin, including isolation and structural analysis, began in the 1960s and continues to this day [24–28]. Silymarin biosynthesis occurs via the phenylpropanoid pathway, using taxifolin and coniferyl alcohol as precursors. Taxifolin is synthesized in the floret and transported to the pericarp, where flavonolignans are formed through oxidative processes involving peroxidase enzymes [29]. Although the specific enzyme responsible for oxidative coupling remains unidentified, this process leads to the production of silybin A, silybin B, silychristin, isosilybin A, isosilybin B, and silydianin. Additionally, silybin synthesis has been successfully achieved in vitro using peroxidase (EC 1.11.1.7) from *S. marianum* cell cultures [11, 30].

Silymarin accumulation in milk thistle is classified into chemotypes A or B, based on the composition of its compounds. Chemotype A is characterized by high levels of silychristin and silybin, whereas chemotype B is distinguished by the presence of silydianin. It has been suggested that chemotype B possesses a fully functional silymarin biosynthetic pathway, encompassing all enzymatically catalyzed silymarin compounds. In contrast, chemotype A is believed to be a natural mutant lacking the ability to biosynthesize silydianin [3, 31]. Figure 2 illustrates the differences between these two chemotypes of milk thistle based on their silymarin content and geographical origins. Among the flavonolignan complexes, silybins A and B are the most abundant, constituting

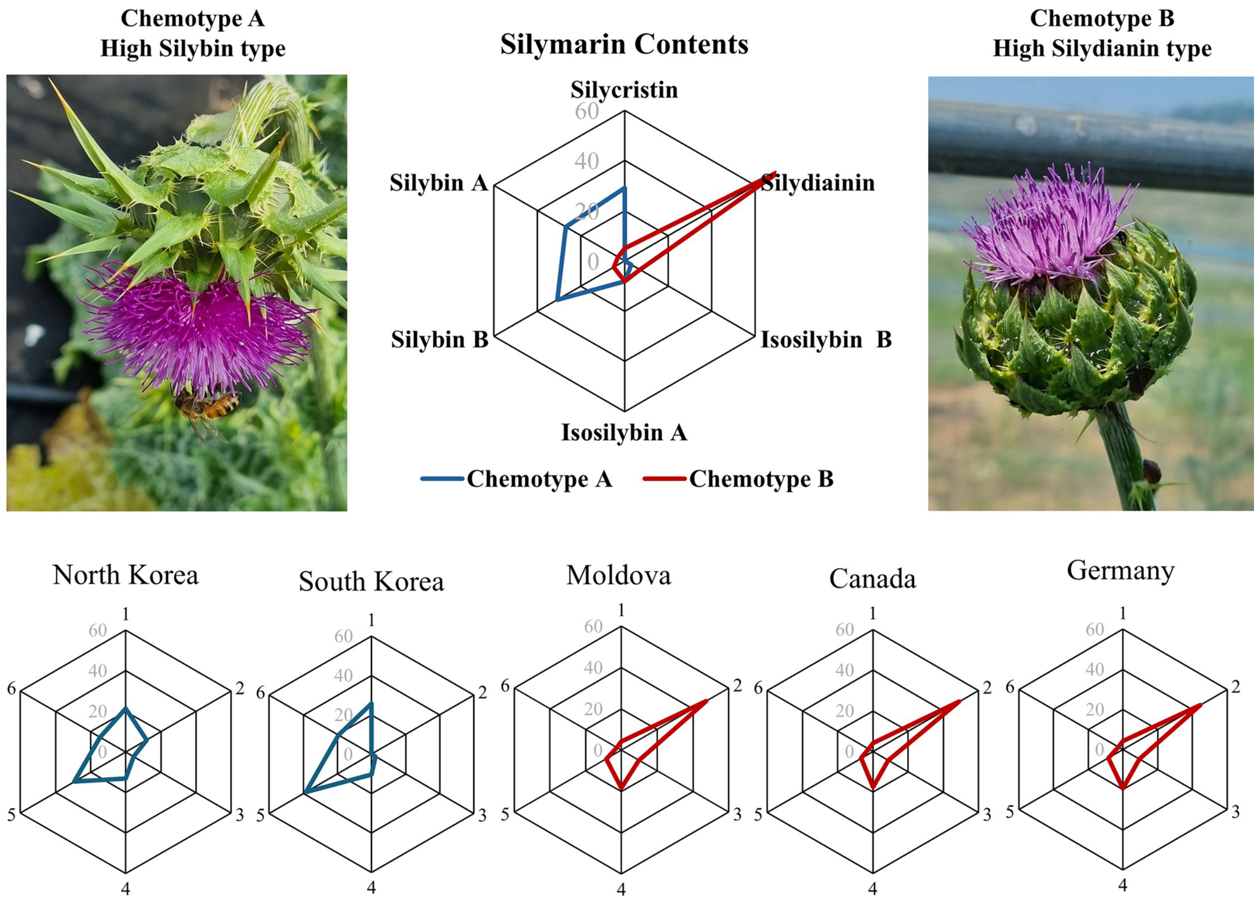


Fig. 2 Comparative analysis of two chemotypes of milk thistle (*S. marianum*) based on their silymarin content and geographical origins: (1) silychristin, (2) silydianin, (3) isosilybin B, (4) isosilybin A, (5) silybin B, and (6) silybin A

approximately 50–70% of the total silymarin extract. These are the primary bioactive components playing vital roles in various therapeutic applications [12, 32]. Furthermore, silydianin has been suggested to have anti-inflammatory potential by inhibiting oxidative products and promoting apoptosis, thereby mitigating the severity of inflammatory responses [33].

Owing to the complexity of the various flavonolignan compounds and the high lipid contents (20–30%), the purification of silymarin from milk thistle seeds in a single-step process has been a technically challenging task [1, 34]. Several extraction techniques have been explored, including supercritical CO_2 extraction [35], subcritical fluid extraction [36], microwave-assisted extraction [37, 38], enzyme-assisted extraction [39], ultrasound-assisted extraction [40], and supercritical fluid extraction method [41]. Optimizing the extraction process is one of the most critical factors in achieving the efficient purification of individual flavonolignan compounds in large quantities, while effectively removing oil constituents. Therefore, selecting the most appropriate extraction method under specific conditions is essential to obtain high-purity silymarin in a cost-effective manner. Most studies on silymarin to date have utilized analytical-scale reverse-phase high-performance liquid chromatography (HPLC) for the quantitative analysis (Fig. 3) [3, 11, 42, 43].

Oil content

In addition to its high silymarin content, a significant portion of the milk thistle seed extract consists of oils, accounting for up to 20–30% of the seed's dry weight. These oils are primarily composed of linoleic and oleic acids, constituting 39.7% and 36.7% of the total lipids, respectively, and smaller amounts of other fatty acids, including palmitic (10.2%), arachidic (3.6%), stearic (6.9%), and behenic (2.5%) acids, with trace amounts of linolenic and gadoleic acids [44]. The total lipid content of milk thistle seeds is comparable to that of well-known oilseed crops such as sunflower and soybean. Furthermore, the high proportion of unsaturated fatty acids makes it a suitable source for cooking oil [45]. However, the high linoleic acid content is considered undesirable due to its high susceptibility to lipid peroxidation, which significantly reduces the stability and shelf life of oil products derived from milk thistle. Future research efforts on milk thistle will likely focus on developing new milk thistle varieties with reduced linoleic acid content through

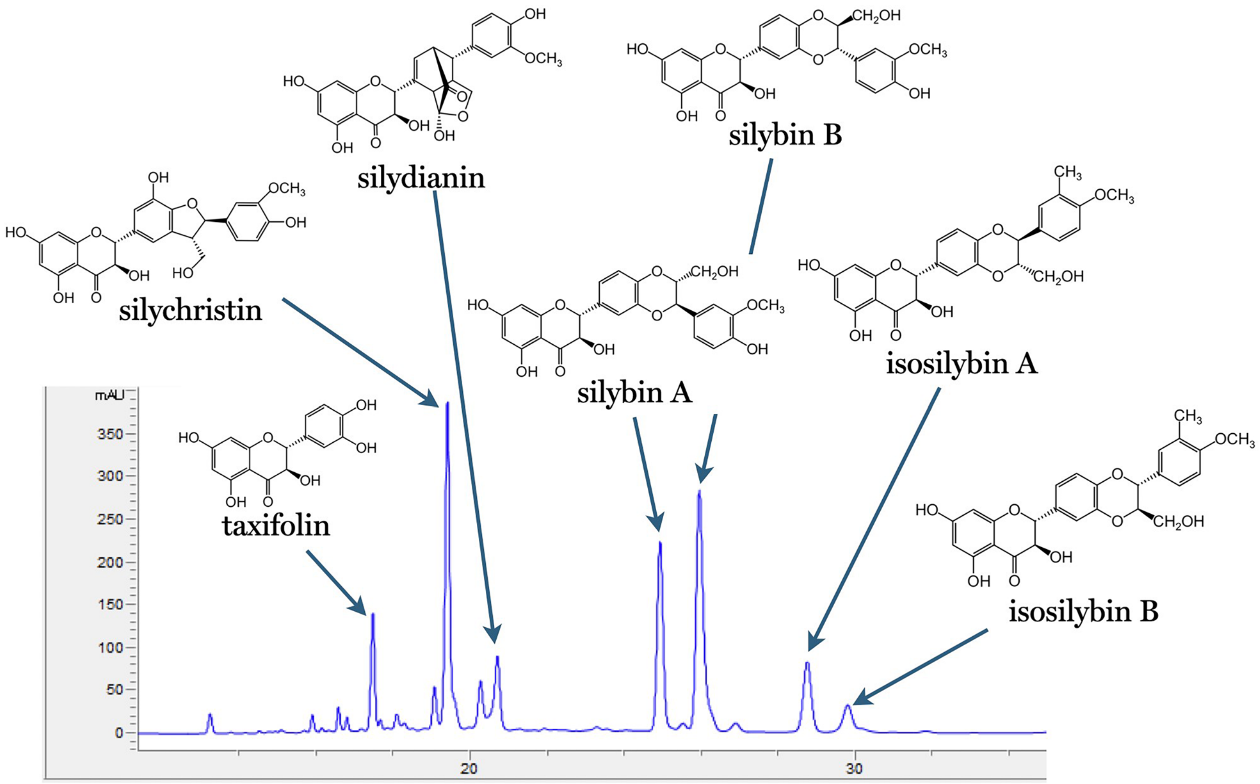


Fig. 3 HPLC chromatogram of silymarin in milk thistle with the chemical structures of silymarin components: taxifolin, silychristin, silydianin, silybin A, silybin B, isosilybin A, and isosilybin B

conventional breeding, molecular breeding, or metabolic engineering approaches [44].

Other uses

In addition to their phytochemical and oil content, milk thistle plants offer various other uses. The leaves, young stems, and sprouts are edible and can be used as fodder for livestock [46–48], as well as consumed as vegetables [46, 49]. Whole or processed seeds are rich in fiber, with calcium and potassium levels higher than many conventional vegetables, making them valuable for enriching processed foods. Furthermore, the high biomass production efficiency exhibited by milk thistle makes it a good candidate for bio-energy production. Additionally, owing to silymarin's potent antioxidant properties, milk thistle is also an ideal ingredient for cosmetic products [50].

Agronomic traits of milk thistle

Plant architecture and morphology

The genus *Silybum* comprises two species: *S. marianum*, characterized by variegated leaves, and *S. eburneum*, which has completely green leaves. Crossing experiments between these species have shown higher fruit yields compared to the parent species [51]. Milk thistle (*S. marianum*) is an annual to biennial plant, commonly found in a variety of environments [52]. Its stems are 40–200 cm tall, glabrous or slightly downy, erect, and branched at the upper part [20, 53]. The basal leaves are large, glabrous, alternate, and have spiny margins, measuring 50–60 cm in length and 20–30 cm in width. A distinctive feature of milk thistle is the milk-white veins, with smaller stem leaves [52]. Milk thistle typically produces red–purple flower heads about 5 cm in diameter [53], though white flowers have also been reported [20, 54]. These flower heads are surrounded by spiny bracts, and the florets are hermaphroditic. The plant is primarily self-pollinating, with an outcrossing rate of approximately 2% [51]. The seeds are achenes, 5–8 mm long, and vary in color from black to brown, with a white pappus. Each flower head produces approximately 190 seeds, with an average of 6350 seeds per plant. Milk thistle seeds can remain viable in the soil for up to 8 years and exhibit little to no dormancy [20, 55, 56].

A study on milk thistle leaves with white variegation, caused by air spaces between the epidermis and green chlorenchyma, examined the characteristics and differences between the white and green patches, as well as their photosynthetic efficiency at low temperatures. The findings suggest that the white veins may serve a physiological function by maintaining a higher temperature in the white areas compared to the green areas under cold conditions, providing an adaptive advantage for tolerating chilling stress [57]. The following sections summarize the results from the cultivation and study of domestically and internationally collected milk thistle resources over a 3-year period in Korea [3].

Plant architecture and morphology in Korea

Figure 4 illustrates the variations in agricultural morphology in flowers, rosettes, involucres, and seed color in milk thistle cultivated and studied in Korea. Key phenotypic traits, such as rosette width, involucre width, spine length, hundred-seed weight, plant height at flowering, and days to harvest after sowing, exhibited high heritability, averaging over 60% among the traits (Table 1) [3].

Relationship between morphology and phytochemical content

To investigate the potential correlation between the cultivar-specific morphology of milk thistle and its silymarin content, an ANOVA analysis was conducted on six genotypes over a 3-year period, accounting for environmental variations and total silymarin content. The characteristic phenotypes of each cultivar exhibited significant differences each year, suggesting highly significant genotype × year interactions [3]. However, the genetic diversity of milk thistle was insufficient to comprehensively elucidate the interactions between agronomic traits and phytochemical content, as previously reported in various studies. Future studies should focus on developing more precise and accurate non-destructive methods to measure agronomic traits, such involucre spine length, plant height, seed color, and other phenotypic characteristics, while also incorporating environmental data.

Genotype × environment interaction of silymarin components

Over the observation period, each milk thistle variety exhibited relatively consistent patterns in total silymarin content and the composition of individual components. This consistency facilitated the identification of high-silybin and high-silydianin types as key indicators for variations in silymarin composition across the two chemotypes. Notably, a significant increase in silydianin accumulation was detected in one particular variety in one of the two regions studied, indicating a genotype-by-environment interaction in silymarin accumulation (Fig. 5) [3]. Figure 5 demonstrates that each variety exhibits a genetically consistent pattern of component accumulation. Although total silymarin content and composition in each variety was fairly consistent with its respective genotypic characteristics across both study regions, one high-silydianin chemotype, M06, showed a remarkable increase in silydianin content in the southern region of Korea during a specific particular year (Haenam, 2021). This suggests that the biochemical process of silydianin

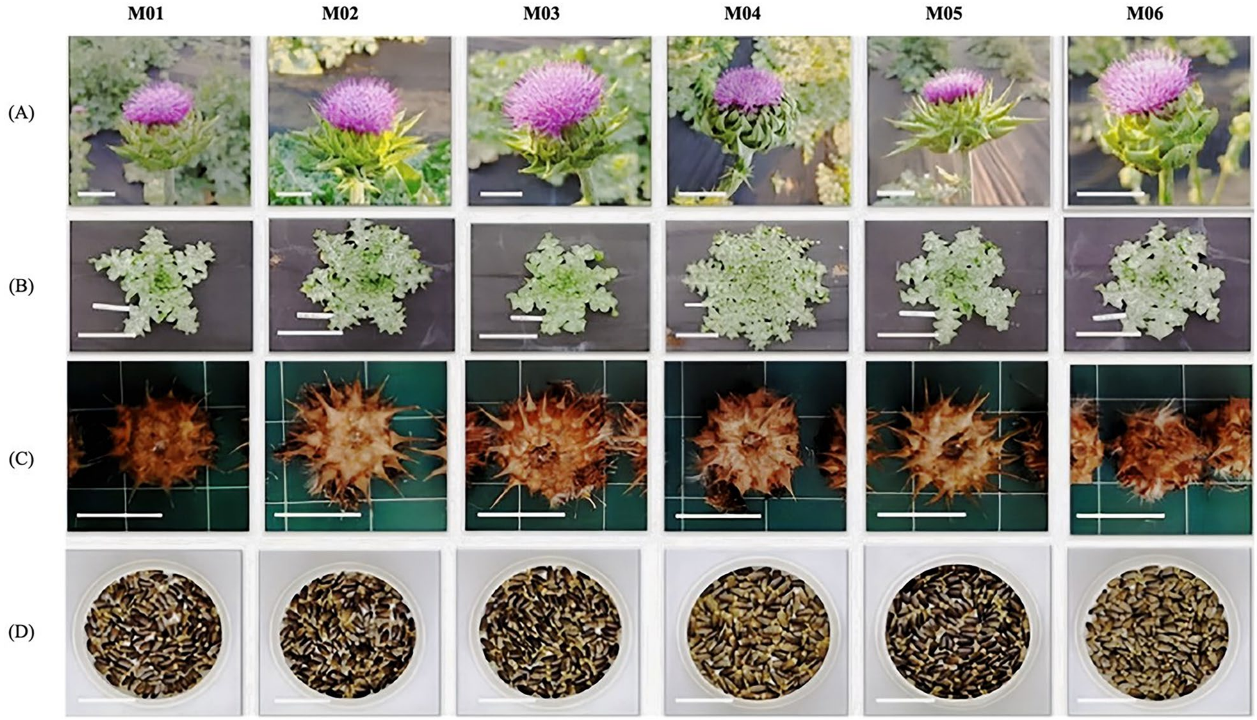


Fig. 4 Phenotypic traits of milk thistle six varieties: M01, Canada; M02, Germany; M03, North Korea; M04, Moldova; M05, South Korea; and M06, South Korea; **A** flower (scale bar = 5 cm), **B** rosette form (scale bar = 20 cm), **C** harvested involucre with spine-tipped bracts (scale bar = 5 cm), and **D** seeds (scale bar = 2 cm). *Source*: Adapted from [3] with permission

production is significantly influenced by environmental factors.

Abiotic stress tolerance

Environmental stresses, both biotic and abiotic, are major limiting factors that affect a plant's growth, development, metabolism, and availability of biologically active substances, directly impacting agricultural yields [58, 59]. Morphological characteristics such as plant height, leaf area, chlorophyll content, number of leaves or branches, and root volume clearly reflect the effects of these stress conditions [60, 61]. Plants possess a broad range of defense mechanisms that enable them to survive extreme conditions. These defense responses are often accompanied by an increase in the production of secondary metabolites, which have been implicated in mitigating both biotic and abiotic stresses [61]. Milk thistle is particularly known for its drought resistance, with studies showing an increase in total phenolic and flavonoid compounds under drought stress [62, 63]. Another study found that drought stress induces the expression of genes for chalcone synthase (*CHS1*, *CHS2*, and *CHS3*), a key and rate-limiting enzyme in the phenylalanine ammonia-lyase (PAL) pathway, which is involved in the synthesis of flavonoids and silymarins [59]. The white variegation of milk thistle leaves, caused by sub-epidermal air spaces, may be a unique adaptation that helps protect the plant from cold stress. Additionally, certain salinity-tolerant varieties, such as 'Royston' and the Iranian wild type, have been reported to withstand high salinity levels (15 dS m^{-1}) while maintaining high silymarin content in the seeds, suggesting a potential role of silymarin in salinity tolerance [19, 64].

Genomics of milk thistle

The first reference genome of milk thistle

Recently, the genome of *S. marianum* was successfully assembled and annotated at the chromosome level, identifying 53,552 predicted protein-coding genes (Fig. 6a, b) [65]. Of these, 94% (50,329 genes) were annotated with known functions available in public databases. This advancement enables a more detailed understanding of the molecular mechanisms underlying of silymarin synthesis and its varying compositions through comprehensive functional genomic analyses. The *S. marianum* genome also revealed that transposable elements constitute 58% of the genome. Compared to other well-studied genomes, such as those of *Arabidopsis* (up to 20%), rice (up to 40%), and humans (approximately 45%), the relatively large size of transposons in the genome

Table 1 Six phenotypic traits of six varieties over 3 years in Hwaseong

Variety	Year	RW (cm)	FPH (cm)	IW (mm)	SPL (mm)	HSW (g)	SHAS (days)
M01	2019*	54.0	100.0	45.0	15.1	2.37	101.0
	2020	117.7 ± 6.7	108.7 ± 5.7	40.8 ± 2.8	11.4 ± 5.3	2.14 ± 0.3	110.8 ± 4.5
	2021	146.8 ± 6.3	145.7 ± 5.5	42.8 ± 5.6	19.0 ± 3.5	2.06 ± 0.1	130 ± 0.0
M02	2019	94.0	124.0	41.6	19.9	2.4	99.0
	2020	116.3 ± 2.1	99.7 ± 4.2	38.4 ± 1.6	20.3 ± 4.2	2.61 ± 0.1	106.8 ± 1.8
	2021	122.2 ± 2.8	102.7 ± 5.9	41.0 ± 0.6	20.9 ± 2.4	2.09 ± 0.1	130 ± 0.0
M03	2019	105.0	140.0	52.8	16.5	1.73	111.0
	2020	131.3 ± 8.7	129.0 ± 11.4	39.3 ± 1.8	11.3 ± 4.0	2.14 ± 0.1	105.4 ± 1.3
	2021	130.8 ± 6.0	147.0 ± 7.2	38.9 ± 0.3	22.1 ± 2.6	2.01 ± 0.1	130.8 ± 1.8
M04	2019	108.0	144.0	58.2	8.5	2.53	109.0
	2020	125.3 ± 9.1	122.3 ± 13.9	47.9 ± 4.8	9.5 ± 4.1	2.38 ± 0.2	114.0 ± 0.0
	2021	133.5 ± 3.9	152.7 ± 4.2	48.1 ± 1.7	17.8 ± 0.7	2.27 ± 0.0	126.0 ± 0.0
M05	2019	120.0	160.0	41.7	27.3	2.70	110.0
	2020	136.3 ± 0.6	138.3 ± 13.5	40.8 ± 1.8	22.6 ± 5.3	2.80 ± 0.0	106.0 ± 0.0
	2021	134.5 ± 3.5	163.7 ± 8.1	36.2 ± 1.1	20.5 ± 0.9	2.43 ± 0.1	126.8 ± 1.8
M06	2019	90.0	145.0	43.2	14.3	2.33	110.0
	2020	106.0 ± 11.5	118.7 ± 12.2	37.9 ± 3.9	5.2 ± 2.9	2.40 ± 0.1	105.3 ± 1.5
	2021	111.3 ± 5.0	157.7 ± 4.6	35.8 ± 0.7	7.9 ± 2.3	2.27 ± 0.0	126.0 ± 0.0
Heritability (%)**	2020	61.3	54.7	61.0	74.9	75.8	87.0
	2021	88.1	96.2	76.2	80.4	86.7	67.7
	Ave	74.7	75.4	68.6	77.6	81.3	77.4

RW width at rosette, *FPH* plant height at flowering, *IW* involucre width, *SPL* spine length, *HSW* hundred seed weight, *SHAS* starting harvest days after sowing

Source: Adapted from [3] with permission

* Individual plants for each variety as an initial selection in 2019 (no population)

** Heritability (%) by two-way ANOVA

of milk thistle suggests the presence of more complex gene regulatory networks, with several other potential implications.

Chloroplast and mitochondrial genomes of milk thistle

In a recent study, a complete chloroplast genome of a local *S. marinum* variety with high silybin B content was sequenced and annotated to compared against other milk thistle accessions of different genotypes [66]. The chloroplast genome was found to be approximately 150 kb in length, comprising 87 protein-coding genes (out of 132 transcripts). These genes primarily function in photosynthesis and the translation of chloroplast transcripts. While the overall chloroplast sequences confirmed the unique identity of *S. marianum* as distinct from its closest relatives, *Cirstium* spp., no sequence polymorphisms were identified among the six different genotypes analyzed in the study. Therefore, the chloroplast genome sequence obtained from the study can be utilized to develop *S. marianum*-specific DNA markers. These markers would be instrumental for differentiating various *S. marianum* varieties and distinguishing them from other members of the *Asteraceae* family, including *Cirsium* species. Given that many *Cirsium* species are morphologically similar to milk thistle and are also used as medicinal plants, the development of such DNA markers would significantly facilitate the screening and classification of milk thistle varieties, enabling the creation of novel varieties with improved traits for industrial applications.

It was found that *S. marianum* has a relatively large mitochondrial genome, approximately 400 kb in length, containing only 27 protein-coding genes (74 transcripts), which is typical of plant mitogenomes compared to their animal counterparts [67]. Similar to the chloroplast genome, these genes are primarily involved in energy metabolism and the translation of the mitochondrial genome. However, due to the abundance of repetitive sequences, the *S. marianum* mitogenome is less suitable than the chloroplast genome for developing DNA markers to distinguish milk thistle varieties from closely related species. Instead, the mitochondrial genome holds greater potential for differentiating individual *S. marianum* genotypes among various landraces and indigenous varieties. In the study referenced in this review [67], a comparative mitogenomic analysis was conducted

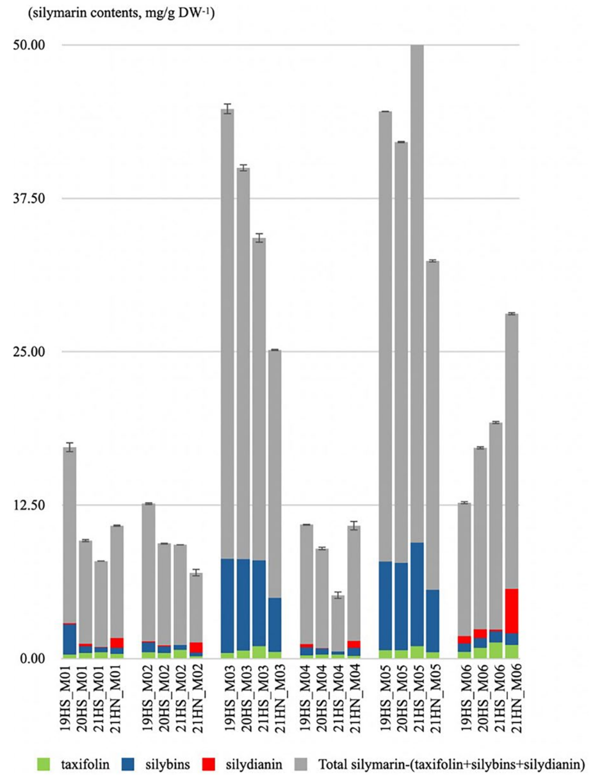


Fig. 5 Cumulative silymarin component contents of taxifolin, silybins, and silydianin. Quantitative analysis conducted over 3 years (2019, 2020, and 2021), comparing two different cultivation regions in 2021: Hwaseong (HS) and Haenam (HN); the analysis included six selected varieties: M01 from Canada, M02 from Germany, M03 from North Korea, M04 from Moldova, M05 from South Korea, and M06 from South Korea; 19HS_2019, Hwaseong; 20HS_2020, Hwaseong; 21HS_2021, Hwaseong; 21HN_2021, Haenam. *Source*: Modified from [10] with permission

with close relatives of *S. marianum*, including *Saussurea costus*, *Arctium lappa*, and *Arctium tomentosum*. The analysis confirmed the feasibility of using sequence information to decipher phylogenetic relationships among closely related species.

Candidate genes for phytochemical traits

Extensive research has been conducted to identify the key genes responsible for the biosynthesis of milk thistle-specific phytochemicals, particularly silymarins. It is well established that the pathway for silymarin biosynthesis originates from the PAL pathway of flavonoid biosynthesis, branching into the synthesis of taxifolin, from which various silymarin compounds are derived [22]. Therefore, chalcone synthase (CSH), a rate-limiting key regulator in the PAL pathway, is considered the top candidate gene for identifying alleles that enhance flavonoid and silymarin production. Additionally, the gene encoding ascorbate peroxidase (APX1), which regulates the initial step of silymarin synthesis from taxifolin [29], has garnered significant attention as a strong candidate target. This enzyme is thought to play a pivotal role in regulating the synthesis of all silymarin components. However, the identities of the enzymes involved in the synthesis of individual silymarin compounds remain largely unknown. It is speculated that many of these enzymes may belong to the oxygenase or peroxidase families, specifically of the cytochrome P450 type [31].

To gain deeper insight into the genetic and molecular basis of various phytochemical traits in milk thistle, thorough analyses of the expression patterns of candidate genes associated with these traits are essential. In RT-qPCR analyses, it is crucial to include a reliable and standardized set of reference genes to normalize and validate gene expression data across different genotypes and environmental conditions. These reference genes must accurately reflect the physiology and environmental interactions unique to the species under investigation. In a recent study, Fulvio et al. proposed a set of ten candidate reference transcripts that can be reliably used in *S. marianum*, based on an analysis of expression patterns across various tissues [68]. Many of these genes have been annotated as homologs of reference genes commonly used in other plant systems, including those for actin (*SmACT*), RNase (*SmRNASE3*), tubulin (*SmTUBα*), ribosomal RNA (*Sm18S*), and PP2A (*SmPP2A*).

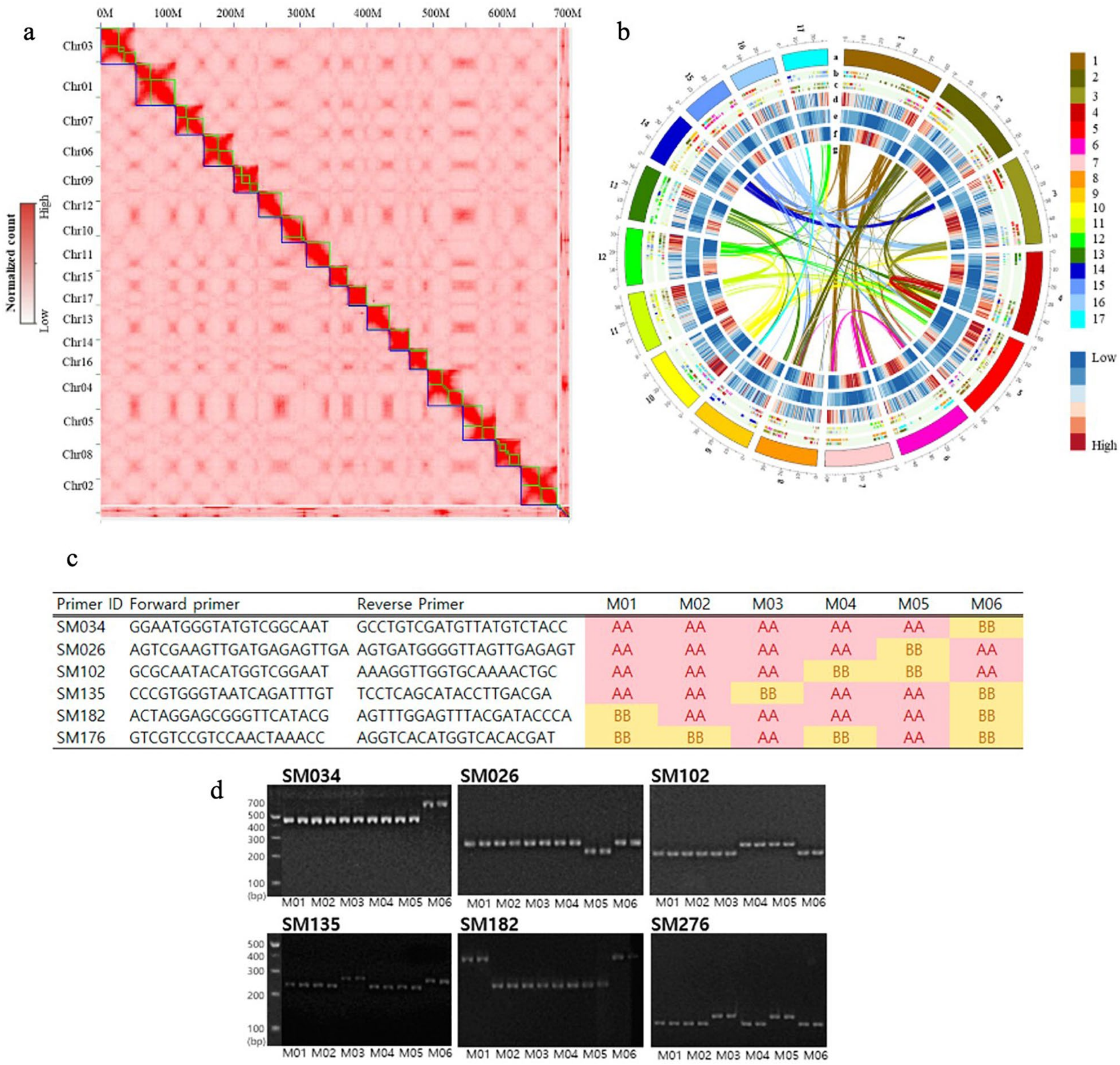
Naturalization of milk thistle in Korea

Evidence of naturalization in Korea

Initially considered a harmful weed, milk thistle has increasingly been cultivated on various farms across Korea as its benefits for liver health have become more widely recognized. The plant, locally known as "white-veined thistle," is now included in the national standard cultivation plant list [69]. Recent studies on milk thistle collected from both domestic and international sources revealed that domestically sourced plants exhibit more uniform growth patterns, suggesting a higher degree of domestication [3, 11].

Comparison of major milk thistle types in Korea

A 3-year study involving six selected milk thistle varieties with distinct agricultural traits was conducted for genomic research. The study evaluated 220 plants derived from both domesticated and foreign seed sources cultivated in Korea. Agricultural traits and silymarin content were assessed, revealing two genetically distinct types: one with high silybin content and another with high



Primer ID	Forward primer	Reverse Primer	M01	M02	M03	M04	M05	M06
SM034	GGAATGGGTATGTCGGCAAT	GCCTGTCGATGTTATGTCTACC	AA	AA	AA	AA	AA	BB
SM026	AGTCGAAGTTGATGAGAGTTGA	AGTGATGGGGTTAGTTGAGAGT	AA	AA	AA	AA	BB	AA
SM102	GCGCAATACATGGTCGGAAT	AAAGGTTGGTGCAAAACTGC	AA	AA	AA	BB	BB	AA
SM135	CCCGTGGGTAATCAGATTTGT	TCCTCAGCATACCTTGACGA	AA	AA	BB	AA	AA	BB
SM182	ACTAGGAGCGGGTTCATACG	AGTTTGGAGTTTACGATACCCA	BB	AA	AA	AA	AA	BB
SM176	GTCGTCCGTCCAACTAAACC	AGGTCACATGGTCACACGAT	BB	BB	AA	BB	AA	BB

Fig. 6 Overview of the genomic landscape of *Silybum marianum*. **a** Pore-C interaction heatmap for the *S. marianum* genome assembly illustrating interactions across the chromosomes, with the intensity of interactions represented by varying shades of red, corresponding to the number of Pore-C reads. **b** Genome features of *S. marianum* mapped across its 17 chromosomes, with each track displayed in 500 kb windows. The tracks include: a. chromosomes of *S. marianum*; b. synteny regions between *Cynara cardunculus* and *S. marianum*; c. synteny regions between *Helianthus annuus* and *S. marianum*; d. gene count of *S. marianum* in 500 kb intervals; e. DNA TE count of *S. marianum* in 500 kb intervals; f. LTR TE count of *S. marianum* in 500 kb intervals. g. Curved lines at the center indicate segmental duplication regions in *S. marianum*. Each color labeled in the tracks a, b, c, and g corresponds to a specific chromosome. *Source*: Adapted from [67] with permission. **c** DNA fingerprinting of six selected accessions using a minimum set of InDel markers. Primer list with corresponding sequences and allele types. Homozygous reference alleles (AA) and homozygous alternative alleles (BB). **d** Agarose gel images using selected InDel markers from two samples per accession. *Source*: Adapted from [19] with permission

silydianin content. Resources with high silydianin content, in particular, showed significant regional differences in silydianin accumulation, indicating that environmental factors play a role in influencing silydianin levels (Fig. 5) [3, 23].

Molecular marker development for genetic identification in milk thistles

In both basic and applied research, DNA-based molecular markers are essential for genotype fingerprinting, marker-assisted selection and breeding, as well as analyzing genetic diversity analysis [23, 70–73]. Recently, six milk thistle accessions exhibiting notable differences in agricultural traits were selected for sequencing against the reference genome. This data was subsequently used to develop genome-wide DNA markers [23]. Of the 177 markers generated, six InDel markers (SM182, SM176, SM135, SM102, SM034, SM026) were successfully employed to differentiate the six accessions (Fig. 6c, d) [23].

In another study, 31 *S. marianum* accessions, collected from geographically diverse regions with heterogeneous genetic backgrounds, along with one *S. eburneum* accession, were subjected to diversity arrays technology (DarT) sequencing. This identified 5178 allelic polymorphisms [21]. Among these, seven SNP markers (Smar03g006250, Smar10g014580, Smar05g013800, Smar05g040540, Smar05g014220, Smar10g000170, Smar11g021760) showed significant trait-specific associations with accessions of varying oil and flavonolignan content, demonstrating their potential as screening tools for genes related to these traits. The study further revealed that *S. eburneum* is split into two distinct genetic groups, indicating that some accessions previously classified as *S. eburneum* based solely on the absence of leaf variegation had been misclassified.

Phytochemical content of naturalized milk thistle

A recent study compared the phytochemical content of *S. marianum* seeds from various origins, including domestically collected resources from South Korea and internationally sourced seeds from North Korea, Germany, Canada, and Moldova. These resources were cultivated and harvested under identical conditions in Korea over a 3-year period (2020–2022). HPLC/DAD analysis revealed that the samples from South Korea consistently exhibited the highest silymarin content throughout the study period [11].

This indicates that the naturalized resources collected in South Korea are already well domesticated and highly adapted to the local environment, exhibiting relatively consistent phenotypes and yield stability each year compared to resources from other countries. Additionally, the study observed a high degree of consistency in chemotype-specific total silymarin content and its compositional ratio. These findings indicate that a significant cultivar selection for establishing a domestic breeding program to enhance silymarin productivity is feasible. Furthermore, utilizing domestic sources can be more cost-effective in meeting the market demands of pharmaceutical companies [3].

Further studies on milk thistle

Trait identification through quantitative trait loci analysis and multi-omics approaches

Shim et al. [3] conducted a comprehensive study on six diverse accessions of milk thistle, focusing on silymarin content, key agronomic traits, and quantitative trait loci (QTL) analysis related to silymarin accumulation in seeds. The application of the data obtained from this study can be significantly enhanced by the recent completion of the chromosome-level genome assembly of *S. marianum* [65], with transcriptome analysis across various tissues expected to follow.

Integrating genome annotation information into the identified QTL regions will facilitate the efficient verification and validation of candidate QTLs associated with the agronomic traits of interest. Additionally, transcriptome data related to flowering and seed maturation can be used to identify candidate genes for developing specific molecular markers associated with the traits of interest, such as flowering time, seed maturation, shedding, yield, and silymarin content. This, in turn, could lead to the efficient development of cultivars with improved traits in these areas.

To complement tissue-specific transcriptome data, which often lacks temporal continuity, single-cell RNA-seq analyses can be employed to analyze different cell types within a tissue. This approach would help provide deeper insights into complex biochemical interactions between various cell types within a tissue, which may be masked by the averaging effect if investigated solely through conventional transcriptome analysis [74]. Comparative analyses of single-cell and spatial transcriptomics can yield a more precise map of gene expression patterns at specific developmental stages, which can expedite the identification of genes involved in flower development, seed maturation, and the biosynthesis of individual silymarin compounds.

Possibility of phenomics application

Phenomics holds significant promise for improving milk thistle seed productivity and bioactive compound yields, particularly silymarin. By collecting multi-dimensional phenotypic data on floral traits (e.g., total involucre length, thorn angle) and seed development stages, researchers can identify critical growth phases and optimize seed yield. Advanced imaging technologies, such as hyperspectral and RGB imaging, can capture detailed phenotypic variations, providing insights into how different developmental stages influence seed quality. For example, a study successfully detected elaiosomes in milk thistle seeds using image-based learning. By employing the Detectron2 deep learning algorithm, elaiosomes were identified with an accuracy of 99.9%. This method can be extended to high-throughput analysis of various seed morphological traits, significantly accelerating breeding programs aimed at cultivar improvement [75]. Detailed phenotyping provides valuable insights into optimizing growing conditions and stages, ultimately enhancing seed production. This directly impacts silymarin accumulation, making it a crucial tool for improving agricultural traits.

Moreover, the integration of phenomics with artificial intelligence (AI) has the potential to revolutionize silymarin production and metabolic optimization. Phenomics data, combined with AI-driven models, can predict seed counts based on floral structure, silymarin content per seed, and the proportions of major silymarin components like silybin and silydianin. This multi-dimensional approach allows comprehensive assessments of growth conditions, photosynthetic efficiency, and overall plant health, contributing to the development of intelligent breeding strategies. In particular, AI can leverage phenotypic markers for high yields of silymarin and silydianin, optimizing metabolic pathways to increase bioactive compound concentrations. This would significantly enhance the productivity and economic value of milk thistle cultivation.

Breeding targets

Several factors critically impact the key agricultural traits related to the overall productivity of milk thistle and its silymarin yield. These include the rate of seedling establishment, water management under both irrigated and water-deficit conditions [76], seed shedding characteristics (such as involucre shape, crown formation, and involucre opening and bending) [19], and spiny leaves [77], all of which directly influence silymarin production in seeds. Additionally, silymarin production is directly proportional to achene yield [78, 79]. Therefore, breeding efforts have focused on improving these traits by targeting cross-chemotype breeding for higher silymarin content in seeds, developing cultivars that exhibit traits such as synchronized flowering, reduced thorns, and minimized achene shedding [19].

Genome editing for phytochemical metabolic engineering

Recent advances in genome editing techniques have transformed the field of genetic engineering. Through precise modifications to specific regions of the genome, it is now theoretically possible to create new traits for crops that do naturally exist. Several gene editing techniques have been successfully applied to enhance the production of secondary metabolites with medicinal or nutritional value [80, 81].

A similar approach can be applied to improve the quality of oils produced from milk thistle. Milk thistle seeds predominantly contain oleic acid (18:1) and linoleic acid (18:2) as major fatty acids, while linolenic acid, an omega-3 fatty acid, is nearly absent. As previously mentioned in the part of 'Key benefits of milk thistle' in this review, the high proportion of linoleic acid (an omega-6 fatty acid) lowers the commercial value of oils derived from milk thistle, omega-6 fatty acids are prone to lipid peroxidation. By employing genome editing techniques, the metabolic pathway for fatty acid synthesis in milk thistle could be modified to alter the properties of the enzymes involved, thereby increasing the omega-3 fatty acid content and reducing the production of omega-6 fatty acids.

Furthermore, genetic engineering technologies, such as gene editing, metabolic engineering, and synthetic biology, can be utilized to identify, modify, and enhance the genes involved in the movement, transport, and final synthesis of silymarin precursors. These approaches could elucidate the underlying mechanisms and improve high-yielding silymarin genotypes for phytochemical production.

Author contributions

Tolangi P, Shim J, and Sumabat R wrote the first draft. Kim S, Tolangi P, and Shim J were involved in manuscript writing, including figures. Park HS, Kim KD, Kim HU, and Lee S contributed to the text revision. Lee S and Chin JH supervised the study. Chin JH organized the funding and was responsible for formatting the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant, funded by the Korean government (MSIT) (No. RS-2024-00410854).

Availability of data and materials

Not applicable.

Declarations

Competing interests

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

Author details

[1] Department of Integrative Biological Sciences and Industry, Sejong University, Seoul 05006, Republic of Korea. [2] Department of Bioindustry and Bioresource Engineering, Sejong University, Seoul 05006, Republic of Korea. [3] Department of Plant Science and Technology, Chung-Ang University, Anseong 17546, Republic of Korea. [4] Natural Product Institute of Science and Technology, Anseong 17546, Republic of Korea. [5] Convergence Research Center for Natural Products, Sejong University, Seoul 05006, Republic of Korea.

Received: 30 September 2024 Accepted: 8 December 2024

References

1. Marceddu R, Dinolfo L, Carrubba A, Sarno M, Di Miceli G (2022) Milk thistle (*Sylibum marianum* L.) as a novel multipurpose crop for agriculture

in marginal environments: a review. Agronomy 12:729. https://doi.org/10.3390/agronomy12030729

2. Groves RH, Kaye PE (1989) Germination and phenology of seven introduced thistle species in southern Australia. Aust J Bot 37:351–359. https://doi.org/10.1071/BT9890351
3. Shim J, Cho H, Sung JS, Yoo E, Chin JH, Lee S (2024) Agricultural phenotype and silymarin content variations of cultivated milk thistle in Korea. Hortic Environ Biotechnol 65:891–901
4. Hamilton WR, Stohs SJ (1998) Hepatic effects of herbal remedies, Chapter 3. In: Miller LG, Murray WJ (eds) Herbal medicinals. A clinician's guide. Pharmaceutical Products Press, New York, pp 37–63
5. Radko L, Cybulski W (2007) Application of silymarin in human and animal medicine. J Pre-Clin Clin Res 6:166–173
6. Kren V, Walterova D (2005) Silybin and silymarin—new effects and applications. Biomed Pap Fac Univ Palacky Olomouc Czech Repub 149:29–41. https://doi.org/10.5507/bp.2005.002
7. Morazzoni P, Bombardelli E (1995) *Silybum marianum* (*Carduus marianus*). Fitoterapia 66:3–42
8. World Health Organization (2024) WHO sounds alarm on viral hepatitis infections claiming 3500 lives each day. https://www.who.int/news/item/09-04-2024-who-sounds-alarm-on-viral-hepatitis-infections-claiming-3500-lives-each-day. Accessed 07 Sep 2024
9. Fact.MR (2022) Milk thistle supplements market to garner US$ 204 Mn by 2032: exclusive fact. MR study. https://www.factmr.com/report/662/milk-thistle-supplements-market#:~:text=Milk%20Thistle%20Supplements%20Demand%20Outlook%20(2022%2D2032)&text=In%20future%2C%20prospects%20appear%20more,aforementioned%20period%20is%20around%207.06%25.&text=Market%20Share%20of%20the%20U.S. Accessed 07 Sept 2024
10. Javed S, Kohli K, Ali M (2011) Reassessing bioavailability of silymarin. Altern Med Rev 16:239–249
11. Tran GH, Lee HD, Shim J, Lee S, Nguyen DD, Lee S (2024) Quantitative analysis of silymarin in *Silybum marianum* seeds collected from different countries during 3 year cultivation. Sci Hort 335:113358. https://doi.org/10.1016/j.scienta.2024.113358
12. Bijak M (2017) Silybin, a major bioactive component of milk thistle (*Silybum marianum* L. Gaernt.)—chemistry, bioavailability, and metabolism. Molecules 22:1942. https://doi.org/10.3390/molecules22111942
13. Serce A, Toptanci BC, Tanrikut SE, Altas S, Kizil G, Kizil S, Kizil M (2016) Assessment of the antioxidant activity of *Silybum marianum* seed extract and its protective effect against DNA oxidation, protein damage and lipid peroxidation. Food Technol Biotechnol 54:455–461. https://doi.org/10.17113/ftb.54.04.16.4323
14. Karimi G, Hassanzadeh-Josan S, Memar B, Esmaeili SA, Riahi-Zanjani B (2018) Immunomodulatory effects of silymarin after subacute exposure to mice: a tiered approach immunotoxicity screening. J Pharmacopunct 21:90–97. https://doi.org/10.3831/KPI.2018.21.011
15. Rambaldi A, Jacobs BP, Iaquinto G, Gluud C (2005) Milk thistle for alcoholic and/or hepatitis B or C liver diseases—a systematic cochrane hepato-biliary group review with meta-analyses of randomized clinical trials. Am J Gastroenterol 100:2583–2591. https://doi.org/10.1111/j.1572-0241.2005.00262.x
16. Martinelli T, Fulvio F, Pietrella M, Bassolino L, Paris R (2023) *Silybum marianum* chemotype differentiation is genetically determined by factors involved in silydianin biosynthesis. J Appl Res Med Aromat Plants 32:100442
17. Bosch-Barrera J, Martin-Castillo B, Buxo M, Brunet J, Encinar JA, Menendez JA (2020) Silibinin and SARS-CoV-2: dual targeting of host cytokine storm and virus replication machinery for clinical management of COVID-19 patients. J Clin Med 9:1770. https://doi.org/10.3390/jcm9061770
18. Precedence Research (2024) Milk thistle market size to hit USD 213.82 million by 2033. https://www.precedenceresearch.com/milk-thistle-market. Accessed 07 Sept 2024
19. Alemardan A, Karkanis A, Salehi A (2013) Breeding objectives and selection criteria for milk thistle [*Silybum marianum* (L.) Gaertn.] improvement. Not Bot Horti Agrobot Cluj-Napoca 41:340–347. https://doi.org/10.15835/nbha4129298
20. Karkanis A, Bilalis D, Efthimiadou A (2011) Cultivation of milk thistle (*Silybum marianum* L. Gaertn.), a medicinal weed. Ind Crop Prod 34:825–830. https://doi.org/10.1016/j.indcrop.2011.03.027
21. Puglisi D, Pasquariello M, Martinelli T, Paris R, De Vita P, Pecchioni N, Esposito S, Bassolino L (2024) Genetic diversity of a *Silybum marianum* (L.) Gaertn germplasm collection revealed by DNA Diversity Array Technology (DArTseq). PLoS ONE 19:e0308368. https://doi.org/10.1371/journal.pone.0308368
22. Drouet S, Tungmunnithum D, Laine E, Hano C (2020) Gene expression analysis and metabolite profiling of silymarin biosynthesis during milk thistle (*Silybum marianum* (L.) Gaertn.) fruit ripening. Int J Mol Sci 21:4730–4747. https://doi.org/10.3390/ijms21134730
23. Shim J, Hong SY, Han JH, Yu Y, Yoo E, Sung J, Chin JH, Lee ON (2023) A genomic evaluation of six selected inbred lines of the naturalized plants of milk thistle (*Silybum marianum* L. Gaertn.) in Korea. Plants 12:2702. https://doi.org/10.3390/plants12142702
24. Althagafy HS, Meza-Avina ME, Oberlies NH, Croatt MP (2013) Mechanistic study of the biomimetic synthesis of flavonolignan diastereoisomers in milk thistle. J Org Chem 78:7594–7600. https://doi.org/10.1021/jo4011377
25. Janiak B, Hänsel R (1960) Phytochemisch-pharmakognostische untersuchungen über fructus cardui mariae. Planta Med 8:71–84. https://doi.org/10.1055/s-0028-1101539
26. Pelter A, Hänsel R (1968) The structure of silybin (silybum substance E6), the first flavonolignan. Tetrahedron Lett 9:2911–2916. https://doi.org/10.1016/S0040-4039(00)89610-0
27. Hänsel R, Schulz J, Pelter A, Rimpler H, Rizk AF (1969) Zur struktur des silybins: synthese von unsymmetrisch substituierten 2.3-trans-benzdioxanen. Tetrahedron Lett 10:4417–4420. https://doi.org/10.1016/S0040-4039(01)88710-4
28. Hänsel R, Kaloga M, Pelter A (1976) Zur konstitution des silychristins. Tetrahedron Lett 17:2241–2244. https://doi.org/10.1016/0040-4039(76)80038-X
29. Lv Y, Gao S, Xu S, Du G, Zhou J, Chen J (2017) Spatial organization of silybin biosynthesis in milk thistle [*Silybum marianum* (L.) Gaertn.]. Plant J 92:995–1004. https://doi.org/10.1111/tpj.13736
30. Sanchez-Sampedro MA, Fernandez-Tarrago J, Corchete P (2007) Silymarin synthesis and degradation by peroxidases of cell suspension cultures of *Silybum marianum*. J Plant Physiol 164:669–674. https://doi.org/10.1016/j.jplph.2006.06.015
31. Martinelli T, Whittaker A, Benedettelli S, Carboni A, Andrzejewska J (2017) The study of flavonolignan association patterns in fruits of diverging *Silybum marianum* (L.) Gaertn. chemotypes provides new insights into the silymarin biosynthetic pathway. Phytochemistry 144:9–18. https://doi.org/10.1016/j.phytochem.2017.08.013
32. Lee JI, Narayan M, Barrett JS (2007) Analysis and comparison of active constituents in commercial standardized silymarin extracts by liquid chromatography-electrospray ionization mass spectrometry. J Chromatogr B 845:95–103. https://doi.org/10.1016/j.jchromb.2006.07.063
33. Zielinska-Przyjemska M, Wiktorowicz K (2006) An in vitro study of the protective effect of the flavonoid silydianin against reactive oxygen species. Phytother Res 20:115–119. https://doi.org/10.1002/ptr.1812
34. Wianowska D, Wisniewski M (2015) Simplified procedure of silymarin extraction from *Silybum marianum* L. Gaertner. J Chromatogr Sci 53:366–372. https://doi.org/10.1093/chromsci/bmu049
35. Honjo M, Mishima K, Matsuyama K, Sekiguchi H, Ando S, Irie K, Mishima K, Fujiwara M (2009) Extraction of silymarins from milk thistle seeds using supercritical carbon dioxide with methanol. Solvent Extr Res Dev Jpn 16:111–120
36. Platonov IA, Nikitchenko NV, Onuchak LA, Arutyunov YI, Kurkin VA, Smirnov PV (2010) Subcritical water extraction of biologically active substances from milk thistle seed (*Silybum murianum* L.). Russ J Phys Chem B 4:1211–1216. https://doi.org/10.1134/S1990793110080063
37. Saleh IA, Vinatoru M, Mason TJ, Abdel-Azim NS, Shams KA, Aboutabl E, Hammouda FM (2017) Extraction of silymarin from milk thistle (*Silybum marianum*) seeds—a comparison of conventional and microwave-assisted extraction methods. J Microw Power Electromagn Energy 51:124–133. https://doi.org/10.1080/08327823.2017.1320265
38. Zheng X, Wang X, Lan Y, Shi J, Xue SJ, Liu C (2009) Application of response surface methodology to optimize microwave-assisted extraction of silymarin from milk thistle seeds. Sep Purif Technol 70:34–40. https://doi.org/10.1016/j.seppur.2009.08.008
39. Liu H, Du X, Yuan Q, Zhu L (2009) Optimisation of enzyme assisted extraction of silybin from the seeds of *Silybum marianum* by Box-Behnken

experimental design. Phytochem Anal 20:475–483. https://doi.org/10.1002/pca.1149

40. Saleh IA, Vinatoru M, Mason TJ, Abdel-Azim NS, Aboutabl EA, Hammouda FM (2015) Ultrasonic-assisted extraction and conventional extraction of silymarin from *Silybum marianum* seeds; a comparison. Res J Pharm Biol Chem Sci 6:709–717
41. Palaric C, Atwi-Ghaddar S, Gros Q, Hano C, Lesellier E (2023) Sequential selective supercritical fluid extraction (S3FE) of triglycerides and flavonolignans from milk thistle (*Silybum marianum* L., Gaertn.). J CO2 Util 77:102609. https://doi.org/10.1016/j.jcou.2023.102609
42. Uy NP, Shim J, Lee HD, Sung JS, Yoo E, Chin JH, Lee S (2023) Analysis of silybin A and silybin B in different accessions of *Silybum marianum* seeds. J Appl Biol Chem 67:159–164. https://doi.org/10.3839/jabc.2024.022
43. Kim J, Paje LA, Choi JW, Lee HD, Shim J, Geraldino PJL, Lee S (2020) Determination of silymarin and silybin diastereomers in Korean milk thistle using HPLC/UV analysis. Korean J Pharmacogn 51:297–301. https://doi.org/10.22889/KJP.2020.51.4.297
44. Malekzadeh M, Mirmazloum SI, Mortazavi SN, Panahi M, Angorani HR (2011) The physicochemical properties and oil constituents of milk thistle (*Silybum marianum* Gaertn. cv. Budakalászi) under drought stress. J Med Plants Res 5:1485–1488
45. Menasra A, Fahloul D (2019) Quality characteristics of biscuit prepared from wheat and milk thistle seeds (*Silybum marianum* (L.) Gaertn.) flour. Carpathian J Food Sci Technol 11:5–19. https://doi.org/10.34302/crpjfst/2019.11.4.1
46. Abu Jadayil S, Tukan SK, Takruri HR (1999) Bioavailability of iron from four different local food plants in Jordan. Plant Foods for Human Nutr 54:285–294. https://doi.org/10.1023/A:1008195019618
47. Beskow WB, Harrington KC, Hodgson J (2006) Dispersal of weed seeds in goat dung. In: Preston C, JH Watts, ND Crossman (eds) Proceedings of the 15th Australian weeds conference, Council of Australian Weed Societies, p 894
48. Mojadam A, Chaji M, Mohammadabadi T, Tabatabaei Vakili S (2015) Feeding value of *Silybum marianum* for sheep and its effect on fiber and protein digestion. Iran J Anim Sci Res 7:267–277. https://doi.org/10.22067/ijasr.v7i3.34753
49. Ballero M, Mura L, Maxia A (2007) Usi alimentari e terapeutici nella tradizione popolare del Goceano (*Sardegna Centrale*). Atti Soc Toscana Sci Nat Mem Ser B 114:45–56
50. Singh RP, Agarwal R (2009) Cosmeceuticals and silibinin. Clin Dermatol 27:479–484. https://doi.org/10.1016/j.clindermatol.2009.05.012
51. Hetz E, Liersch R, Schieder O (1995) Genetic investigations on *Silybum marianum* and *S. eburneum* with respect to leaf colour, outcrossing ratio, and flavonolignan composition. Planta Med 61:54–57. https://doi.org/10.1055/s-2006-957999
52. Gresta F, Avola G, Guarnaccia P (2007) Agronomic characterization of some spontaneous genotypes of milk thistle (*Silybum marianum* L. Gaertn.) in Mediterranean environment. J Herbs Spices Med Plants 12:51–60. https://doi.org/10.1300/J044v12n04_05
53. Montemurro P, Fracchiolla M, Lonigro A (2007) Effects of some environmental factors on seed germination and spreading potentials of *Silybum marianum* Gaertner. Ital J Agron 2:315–320. https://doi.org/10.4081/ija.2007.315
54. Vaknin Y, Hadas R, Schafferman D, Murkhovsky L, Bashan N (2008) The potential of milk thistle (*Silybum marianum* (L.)), an Israeli native, as a source of edible sprouts rich in antioxidants. Int J Food Sci Nutr 59:339–346. https://doi.org/10.1080/09637480701554095
55. Dodd J (1989) Phenology and seed production of variegated thistle, *Silybum marianum* (L.) Gaertn., in Australia in relation to mechanical and biological-control. Weed Res 29:255–263. https://doi.org/10.1111/j.1365-3180.1989.tb00910.x
56. Sindel BM (1991) A review of the ecology and control of thistles in Australia. Weed Res 31:189–201. https://doi.org/10.1111/j.1365-3180.1991.tb01758.x
57. Shelef O, Summerfield L, Lev-Yadun S, Villamarin-Cortez S, Sadeh R, Herrmann I, Rachmilevitch S (2019) Thermal benefits from white variegation of *Silybum marianum* leaves. Front Plant Sci 10:688. https://doi.org/10.3389/fpls.2019.00688
58. Bohnert HJ, Nelson DE, Jensen RG (1995) Adaptations to environmental stresses. Plant Cell 7:1099–1111. https://doi.org/10.1105/tpc.7.7.1099
59. ElSayed AI, El-Hamahmy MAM, Rafudeen MS, Mohamed AH, Omar AA (2019) The impact of drought stress on antioxidant responses and accumulation of flavonolignans in milk thistle (*Silybum marianum* (L.) Gaertn.). Plants 8:611. https://doi.org/10.3390/plants8120611
60. Pradhan J, Sahoo SK, Lalotra S, Sarma RS (2017) Positive impact of abiotic stress on medicinal and aromatic plants. Int J Plant Sci 12:309–13
61. Jan R, Asaf S, Numan M, Lubna KKM (2021) Plant secondary metabolite biosynthesis and transcriptional regulation in response to biotic and abiotic stress conditions. Agronomy 11:968. https://doi.org/10.3390/agronomy11050968
62. Selmar D, Kleinwächter M (2013) Influencing the product quality by deliberately applying drought stress during the cultivation of medicinal plants. Ind Crops Prod 42:558–566. https://doi.org/10.1016/j.indcrop.2012.06.020
63. Zahir A, Abbasi BH, Adil M, Anjum S, Zia M, Ihsan-Ul-Haq (2024) Synergistic effects of drought stress and photoperiods on phenology and secondary metabolism of *Silybum marianum*. Appl Biochem Biotechnol 174:693-707. https://doi.org/10.1007/s12010-014-1098-5
64. Ghavami N, Ramin AA (2008) Grain yield and active substances of milk thistle as affected by soil salinity. Commun Soil Sci Plant Anal 39:2608–2618. https://doi.org/10.1080/00103620802358672
65. Kim KD, Shim J, Hwang JH, Kim D, El Baidouri M, Park S, Song J, Yu Y, Lee K, Ahn BO, Hong SY, Chin JH (2024) Chromosome-level genome assembly of milk thistle (*Silybum marianum* (L.) Gaertn.). Sci Data 11:342. https://doi.org/10.1038/s41597-024-03178-3
66. Shim J, Han JH, Shin NH, Lee JE, Sung JS, Yu Y, Lee S, Ahn KH, Chin JH (2020) Complete chloroplast genome of a milk thistle (*Silybum marianum*). Plant Breed Biotechnol 8:439–444. https://doi.org/10.9787/PBB.2020.8.4.439
67. Lee J, Zoclanclounon YAB, Jung H, Lee T, Kim J, Park G, Lee K, An K, Shim J, Chin JH (2022) Assembly and comparative analysis of complete mitogenome of *Silybum marianum* (L.) Gaertner. Korean J Breed Sci 54:294–304. https://doi.org/10.9787/KJBS.2022.54.4.294
68. Fulvio F, Martinelli T, Paris R (2021) Selection and validation of reference genes for RT–qPCR normalization in different tissues of milk thistle (*Silybum marianum* Gaert.). Gene 768:145272. https://doi.org/10.1016/j.gene.2020.145272
69. Korea National Arboretum, National Standard Plant List. http://www.nature.go.kr/kpni/index.do. Accessed 07 Sept 2024
70. Schlotterer C (2004) The evolution of molecular markers—just a matter of fashion? Nat Rev Genet 5:63–69. https://doi.org/10.1038/nrg1249
71. Mohan M, Nair S, Bhagwat A, Krishna TG, Yano M, Bhatia CR, Sasaki T (1997) Genome mapping, molecular markers and marker-assisted selection in crop plants. Mol Breed 3:87–103. https://doi.org/10.1023/A:1009651919792
72. Liu B, Wang Y, Zhai W, Deng J, Wang H, Cui Y, Cheng F, Wang X, Wu J (2013) Development of InDel markers for *Brassica rapa* based on whole-genome re-sequencing. Theor Appl Genet 126:231–239. https://doi.org/10.1007/s00122-012-1976-6
73. Collard BC, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc Lond B Biol Sci 363:557–572. https://doi.org/10.1098/rstb.2007.2170
74. Bawa G, Liu Z, Yu X, Tran LP, Sun X (2024) Introducing single cell stereo-sequencing technology to transform the plant transcriptome landscape. Trends Plant Sci 29:249–265. https://doi.org/10.1016/j.tplants.2023.10.002
75. Kim Y, Abebe AM, Kim J, Hong S, An K, Shim J, Baek J (2024) Deep learning-based elaiosome detection in milk thistle seed for efficient high-throughput phenotyping. Front Plant Sci 15:1395558. https://doi.org/10.3389/fpls.2024.1395558
76. Afshar RK, Chaichi MR, Assareh MH, Hashemi M, Liaghat A (2014) Interactive effect of deficit irrigation and soil organic amendments on seed yield and flavonolignan production of milk thistle (*Silybum marianum* L. Gaertn.). Ind Crops Prod 58:166–172. https://doi.org/10.1016/j.indcrop.2014.03.043
77. Khan SA, Hamid S, Sabir AW (1988) Development of spineless *Silybum marianum*. Pak J Sci Ind Res 31:585–586
78. Ram G, Bhan MK, Gupta KK, Thaker B, Jamwal U, Pal S (2005) Variability pattern and correlation studies in *Silybum marianum* L. Gaertn. Fitoterapia 76:143–147. https://doi.org/10.1016/j.fitote.2004.10.006

79. Andrzejewska J, Sadowska K (2008) Effect of cultivation conditions on the variability and interrelation of yield and raw material quality in milk thistle (*Silybum marianum* (L.) Gaertn.). Acta Sci Pol Agric 7:3–11
80. Dey A (2021) CRISPR/Cas genome editing to optimize pharmacologically active plant natural products. Pharmacol Res 164:105359. https://doi.org/10.1016/j.phrs.2020.105359
81. Mitra S, Anand U, Ghorai M, Kant N, Kumar M, Radha JNK, Swamy MK, Prockow J, Pérez de la Lastra JM, Dey A (2023) Genome editing technologies, mechanisms and improved production of therapeutic phytochemicals: opportunities and prospects. Biotechnol Bioeng 120:82–94. https://doi.org/10.1002/bit.28260

Publisher's Note