

Article

A Genomic Evaluation of Six Selected Inbred Lines of the Naturalized Plants of Milk Thistle (*Silybum marianum* L. Gaertn.) in Korea

Jeehyoung Shim ^{1,2,†}, Su Young Hong ^{3,†}, Jae-Hyuk Han ⁴, Yeisoo Yu ⁵, Eunae Yoo ⁶, Jungsook Sung ⁶, Joong Hyoun Chin ^{4,7,*} and O New Lee ^{8,*} 

¹ Department of Food and Nutrition, Chung Ang University, Seodong-daero 4726, Daedeok-myeon, Anseong 17546, Republic of Korea; gyoung.feb@gmail.com

² EL&I Co., Ltd., Hwaseong 18278, Republic of Korea

³ Genomics Division, National Institute of Agricultural Sciences, Rural Development Administration, Jeonju 54874, Republic of Korea; suyong@korea.kr

⁴ Food Crops Molecular Breeding Laboratory, Department of Integrative Biological Sciences and Industry, Sejong University, Seoul 05006, Republic of Korea; 0724jh@gmail.com

⁵ DNACare Co., Ltd., Seoul 06730, Republic of Korea; yeisooyu@dnacare.co.kr

⁶ National Agrobiodiversity Center, National Institute of Agricultural Sciences, Rural Development Administration, Jeonju 54874, Republic of Korea; eung77@korea.kr (E.Y.); sjs7861@korea.kr (J.S.)

⁷ Convergence Research Center for Natural Products, Sejong University, Seoul 05006, Republic of Korea

⁸ Department of Bioindustry and Bioresource Engineering, Sejong University, Seoul 05006, Republic of Korea

* Correspondence: jhchin@sejong.ac.kr (J.H.C.); onewlee@sejong.ac.kr (O.N.L.); Tel.: +82-2-3408-3897 (J.H.C.); +82-2-6935-2610 (O.N.L.)

† The authors contributed equally to this work.

Abstract: Milk thistle (*Silybum marianum*) belongs to the Asteraceae family and is a medicinal plant native to the Mediterranean Basin. Silymarin in achene is a widely used herbal product for chronic liver disease. There is growing interest in natural medicine using milk thistle in Korea, but the raw material completely relies on imports. Despite its economic importance, phenotypic evaluations of native resources of milk thistle in Korea have not been carried out. In addition, genomic research and molecular marker development are very limited in milk thistle. In this study, we evaluated 220 milk thistle resources consisting of 172 accessions collected from the domestic market, and 48 accessions isolated from 6 accessions distributed by the National Agrobiodiversity Center in Korea. Six plant characteristics (height, seed weight, number of flowers, seed weight per flower, spine length, and color at harvest) were measured, and six samples (M01–M06) were selected to represent the genetic diversity of the population for genomic research. To develop PCR-based and co-dominant insertion/deletion (InDel) markers, we performed genome-wide InDel detection by comparing the whole-genome resequencing data of the six selected accessions with the reference genome sequence (GCA_001541825). As a result, 177 InDel markers with high distinguishability and reproducibility were selected from the 30,845 InDel variants. Unknowingly imported alien plant resources could easily be genetically mixed, and jeopardized seed purity can cause continuous difficulties in the development of high value-added agricultural platforms utilizing natural products. The selected plant materials and 177 validated InDel markers developed via whole-genome resequencing analysis could be valuable resources for breeding, conservation, and ecological studies of natives to Korea, along with acceleration of *Silybum marianum* industrialization.

Keywords: *Silybum marianum*; milk thistle; InDel; resequencing; naturalized plants



Citation: Shim, J.; Hong, S.Y.; Han, J.-H.; Yu, Y.; Yoo, E.; Sung, J.; Chin, J.H.; Lee, O.N. A Genomic Evaluation of Six Selected Inbred Lines of the Naturalized Plants of Milk Thistle (*Silybum marianum* L. Gaertn.) in Korea. *Plants* **2023**, *12*, 2702. <https://doi.org/10.3390/plants12142702>

Academic Editor: Alex Troitsky

Received: 19 June 2023

Revised: 18 July 2023

Accepted: 19 July 2023

Published: 20 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Milk thistle (*Silybum marianum* L. Gaertn.) is an annual-to-biennial plant of the Asteraceae family, native to the Mediterranean area, and is now a widely cultivated officinal

plant [1–3]. It is a diploid species ($2n = 34$) and an autogamous plant with an average outcrossing rate of 2% under field conditions [4–6]. Milk thistle has a glossy, brown–black-to-greyish husk achene with cypselae. In the dry pericarp and seed coat of the achene, flavonolignans (about 70–80%), as well as polymeric and oxidized polyphenolic compounds consisting of a mixture of flavonoids, are accumulated [1,3]. The health-promoting properties of milk thistle are attributable to flavonolignans, commonly referred to as silymarin [2,7]. Silymarin content ranges from 1.5% to 3.0% of the achene dry weight, but may exceed 4.0% [7]. Silymarin is composed of the six representatives of flavonolignans: silybin, isosilybin, silychristin, isosilychristin, silydianin, and silimonin [3]. The main component of silymarin is silybin ($C_{25}H_{22}O_{30}$, molecular weight of 482.441), which is a mixture of two diastereomers: silybins A and B. Silybin has pharmacologically relevant actions for human liver diseases (e.g., liver-regenerating properties, anti-inflammatory, immunomodulating, antifibrotic, and antioxidating effects), as well as the clinical potential in patients with viral hepatitis, drug-induced liver injury, and non-alcoholic fatty liver disease [1]. Aside from silymarin, milk thistle achenes have a high oil content (20–30%) [2,4]. Byproducts generated from silymarin extraction and other fractions of biomass can be used in various ways, namely edible oil, fodder, cosmetics, and bioenergy production [4].

Milk thistle is not only rich in nutrients and offers medicinal benefits, but is also well-adapted to suboptimal growing conditions [8]. It is considered one of the most interesting alternative crops in the Mediterranean environment [9,10] and has been tested as a potential commercial seed crop in Canada and New Zealand [4,11,12]. In North America, commercial cultivation has recently become more significant due to the growing popularity of herbal supplements and increasing demand for milk thistle extract from the pharmaceutical industry [13]. In Korea, milk thistle is recognized as a naturalized plant introduced by artificial or natural methods that is able to reproduce and survive in the wild [14]. Milk thistle extract was ranked 10th, with a yield of 45.1 billion KWN (34 million USD) in the health functional food market, itself estimated to have a value of 4.03 trillion KWN (2.8 billion USD) (Ministry of Food and Drug Safety of Korea, 2021). On the other hand, domestic pharmaceutical companies rely on imported materials of milk thistle from foreign countries, such as Poland, France, the U.S., India, etc. In order to develop a sustainable milk thistle industry in Korea, it is necessary to evaluate domestic resources and develop domestic milk thistle varieties.

DNA-based molecular markers are a valuable tool in both basic and applied research, such as fingerprinting genotypes, analyzing genetic diversity, and marker-assisted breeding [15–18]. Phenotypic characters are generally influenced by the environmental factors and developmental stage of the plant, and agronomic practices. In contrast, molecular markers based on DNA sequence polymorphisms are independent of environmental factors and show high polymorphism, reproducibility, and reliable identification. NGS has allowed us to identify a massive number of single-nucleotide polymorphisms (SNPs) and insertion and deletion (InDel) polymorphisms between highly homologous genomes [19]. InDel markers are relatively easy to genotype based on their fragment-length polymorphism, without special infrastructure to perform SNP genotyping, and extensively cover the whole genome [15,20,21]. InDels are the second most abundant forms of genetic variation in plants and humans, next to SNPs [21–23]. InDel markers offer advantages in their multi-allelic nature and codominant inheritance. InDels generally have a low frequency of homoplasy, which represents the probability of two InDel mutations of exactly the same length occurring at the same genomic location, allowing InDels to be confidently related to identity-by-descent [19]. For this reason, the usefulness of InDels over SSR markers was demonstrated in analyzing the interspecific structure of cultivated citrus genetic diversity. InDels are genetic variants that can have a more significant impact on protein structure and function than single-base changes, thus allowing for their use in the development of phylogenetic markers [24]. InDel markers have been developed in various crop species, such as rice, soybean, hot pepper, and maize [25–29]. However, in milk thistle, genome-wide PCR markers have not yet been reported. There have been some recent attempts to create

shatter-resistant mutant lines using DArT (diversity array technology), and analyze genetic diversity using the SCoT (start codon-targeted) marker system [4,30]. By harnessing the first draft of the whole genome of *S. marianum*, it is now possible to detect genome-wide InDel polymorphisms among different accessions using whole-genome resequencing to guide the efficient development of PCR-based markers [18].

In this study, six accessions that showed significant differences in agricultural characteristics were selected to be used as initial materials for genomic research of milk thistle. We used resequencing data from the six accessions in comparison with the reference genome sequence (assembly ID: GCA_001541825.1) to identify 30,845 polymorphisms across the genome. Furthermore, we converted 3410 InDel polymorphisms with a separation ratio of 3:3 between the reference allele and the alternative allele among the six re-sequenced accessions to PCR-based InDel assay markers. After experimental validation, 177 InDel markers with high distinguishability and reproducibility were selected. These plant materials and InDels will be useful resources for genetic research and breeding programs of *S. marianum*.

2. Results

2.1. Phenotypic Evaluation of 220 Milk Thistle Accessions

All analyzed traits showed a continuous unimodal distribution among the 220 milk thistle accessions, except for plant color at harvest (Figure 1). The plant height ranged from 32 cm to 176 cm, with an average of 109 cm, when erect and branched in the upper part of the plant (Table 1, Figure 2). The basal leaves were large and glabrous with spiny margins. Each stem ended in a flower with a spiny bract, and the average number of flower heads was 37.5 per plant. The average seed weight per plant in 220 accessions was 112.2 g, and ranged from 0.2 g to 311.4 g. Phenotypic comparisons between M01–06 and 220 accessions showed no significant differences in SW, FHN, and PC. However, there were significant differences in PH, SW/FHN, and SL, as the value of six accessions were higher than 220 accessions (Table 1). All traits were highly correlated ($p \leq 0.01$) with each other, except for the number of flower heads and spine length (Table S1). Positive correlations were observed among the traits, while spine length showed negative correlations with all the other traits. Among the six traits, two traits (spine length and plant color at harvest) showed a low correlation coefficient and high p -value with the other traits (plant height, seed weight, flower head number, and seed weight/flower head number). Although the p -value was low and significant, it is likely that spine length and plant color had a low correlation with the other traits.

Table 1. Comparison of the phenotypic variations in the 6 accessions between M01–06 and 220 accessions (accs.). Six agronomical traits were evaluated: plant height (PH, cm), seed weight (SW, g), flower head number (FHN), SW/FHN, spine length (SL, mm), and plant color at harvest (PC). PC was classified into three categories based on color charts: green (1), brown (2), or green-brown (3). The mean \pm SD are shown in a separate column.

Variable	M01	M02	M03	M04	M05	M06	Mean		Contrast
							M01–M06	220 Accs.	M01–06:220 Accs.
PH	121.5	100.1	136.2	141.6	131.9	134.2	127.6 \pm 15.0	109.8 \pm 27.0	*
SW	116.1	79.5	118.8	97.2	142.0	109.1	110.4 \pm 21.1	112.2 \pm 85.0	NS
FHN	40.4	38.7	37.7	25.4	43.9	45.0	38.5 \pm 7.0	37.5 \pm 21.7	NS
SW/FHN	2.7	2.1	3.0	3.8	3.2	2.4	2.9 \pm 0.6	2.6 \pm 1.4	*
SL	39.7	46.4	43.9	28.7	45.8	22.9	37.9 \pm 9.8	20.7 \pm 6.0	*
PC	2.7	2.7	2.7	2.3	2.7	2.7	2.6 \pm 0.1	2.6 \pm 0.8	NS

* Significant at $p < 0.05$, NS, not significant.

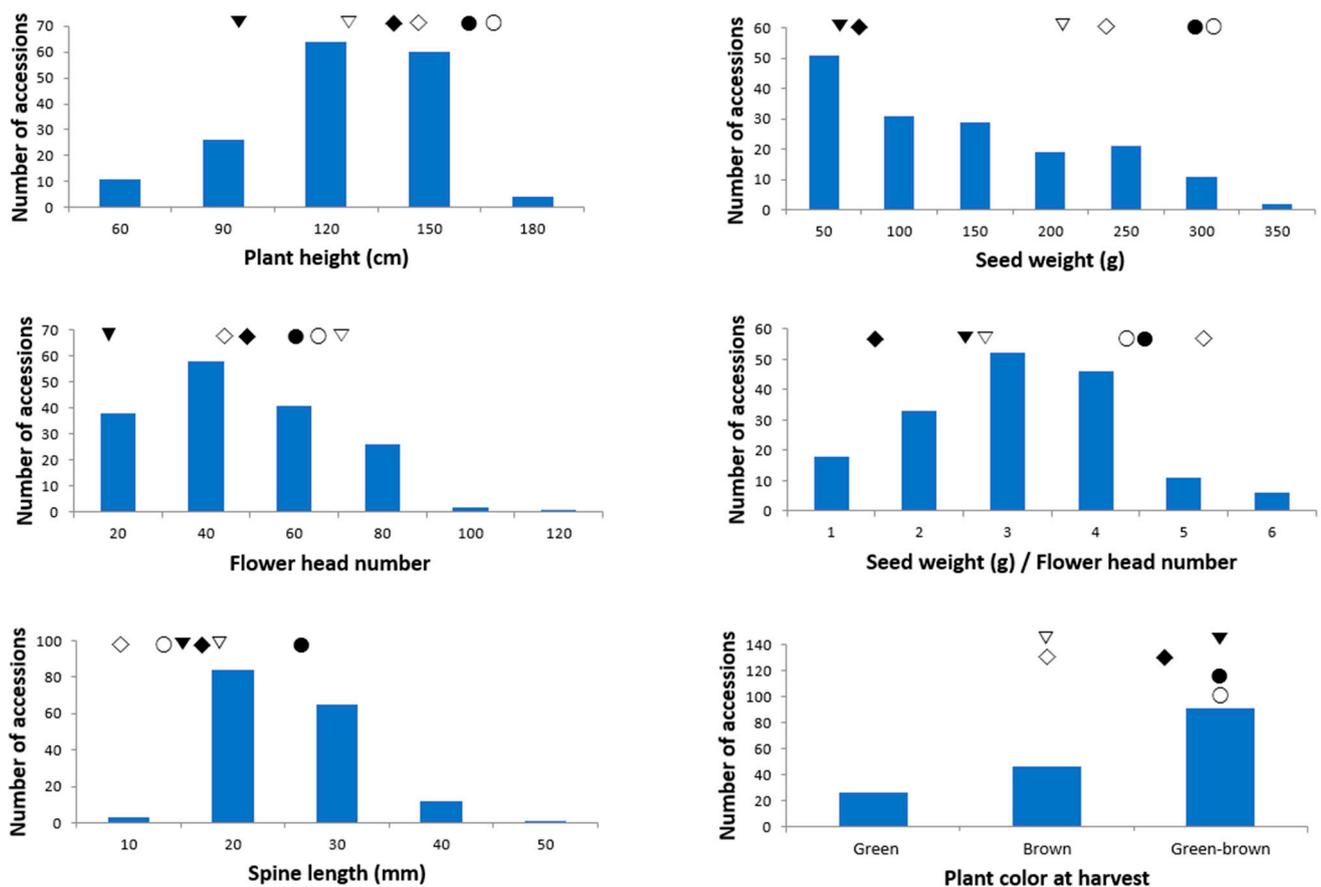


Figure 1. Frequency distribution for each trait of the 220 milk thistle accessions. Six representative milk thistle accessions are displayed as figures on the graph. M01 (▼); M02 (▽); M03 (◆); M04 (◇); M05 (●); and M06 (○). Plant color at harvest was classified into three categories based on color charts: green, brown, and green-brown.

To assess the phenotypic diversity of 220 accessions, six agronomic traits were evaluated and analyzed using principal component analysis (PCA) and cluster analysis. The first two PCs accounted for approximately 74%. PC1 represented the “crop yield” such as seed weight ($r = 0.97$), flower head number ($r = 0.91$), plant height ($r = 0.86$), and seed weight/flower head number ($r = 0.77$), which accounted for 56.2% of total variance (Table 2). PC2 represented the “appearance of plants at harvest”, such as spine length ($r = 0.84$) and plant color at harvest ($r = -0.72$), which accounted for 17.6% of total variance. Using a hierarchical cluster analysis based on the six agronomic traits, a dendrogram classified the 220 milk thistle accessions into three main groups (Figure S1). We selected six representative milk thistle accessions (M01–M06) as initial materials for genomic research, including two Korean natives (M05 and 06). M01, M03 and M04, and M02, M05, and M06 were each assigned to one of three clusters. After the selection, we re-evaluated the six representative plants for six agricultural characteristics from 2019 to 2021 (Table S2). The six accessions showed significant differences in all six traits. Plant height and involucre diameter were the highest in M04 among the six accessions, whereas the number of flower heads, 100-achene weight, and spine length were the highest in M05. Plant height was the highest in M04 (124.3 cm and 55.8 mm, respectively), but was the smallest in M03 (95.3 cm and 47.1 mm, respectively). The number of branches of M06 was the highest (23.4) among the six accessions.

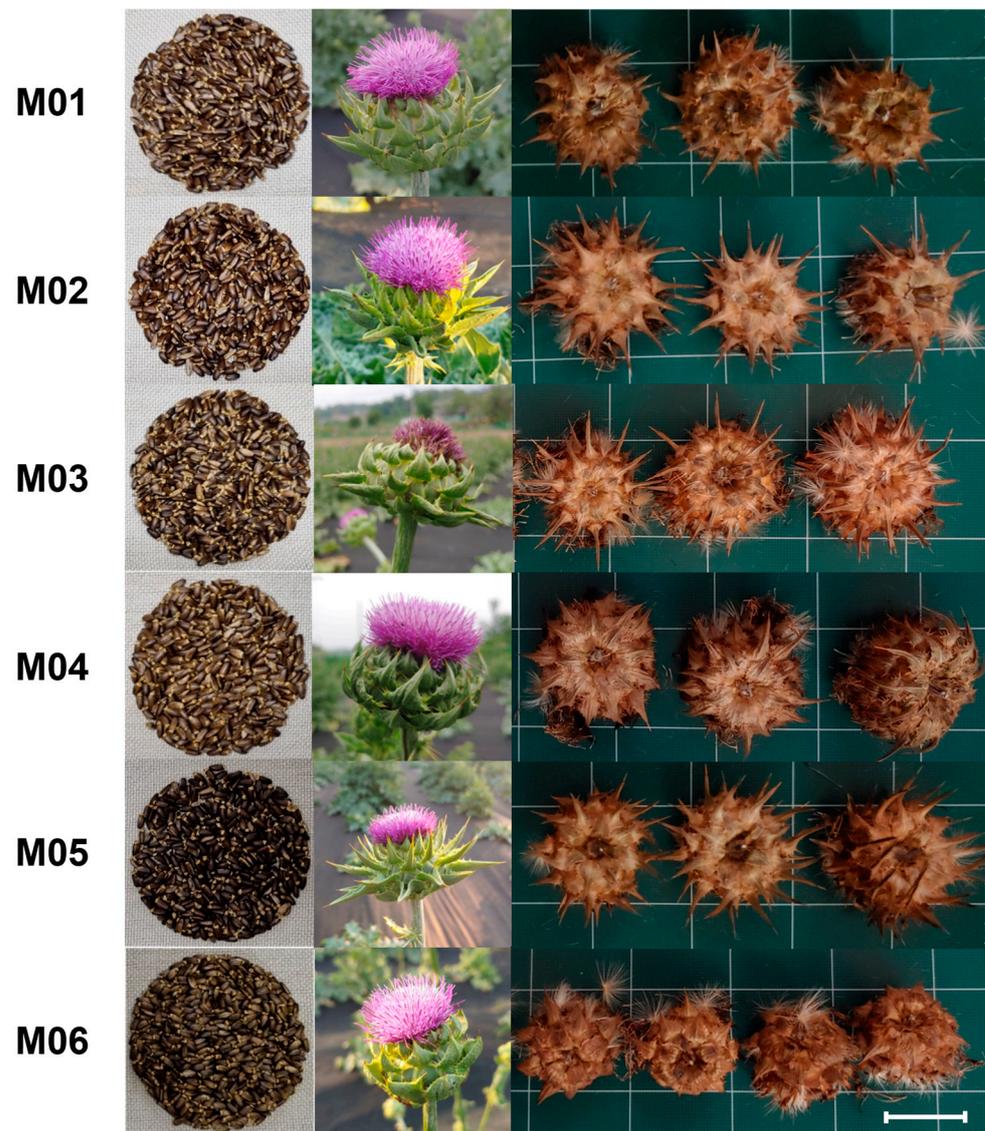


Figure 2. Morphological characteristics of the selected six milk thistle accessions: (from left) seed, flower, and dried involucre after harvest. The scale bar indicates 40 mm.

Table 2. Results of the first two axes (PC1, PC2) of the principal component analysis of morphological traits within the relationships between 220 milk thistle accessions. Six agronomical traits were evaluated: plant height (PH, cm), seed weight (SW, g), flower head number (FHN), SW/FHN, spine length (SL, mm), and plant color at harvest (PC). PC was estimated as one of three types according to color charts: green (1), brown (2), or green-brown (3).

	PC1	PC2
Eigenvalue	3.37	1.06
Variance (%)	56.2	17.6
Cumulative variance (%)	56.2	73.8
Variable		
SW	0.97	−0.30
FHN	0.91	−0.14
PH	0.86	−0.44
SW/FHN	0.77	−0.43
SL	−0.23	0.84
PC	0.33	−0.72

2.2. Sequencing and Mapping Summary

To develop genome-wide DNA markers in *S. marianum*, we produced 120.5 Gb of raw sequence across a total of six accessions, which ranged from 16.7 Gb (M01) to 22.7 Gb (M03) and about 20.1 Gb for each accession on average (Table S3). After LQ and adapter sequences were trimmed using “fastp”, with a Q-score ≥ 20 and a minimum length ≥ 36 bp, the trimmed sequence remained 117 Gb in total (97.1% of raw data). It ranged from 16.3 Gb (M01) to 21.9 Gb (M03, M05), and 19.5 Gb per accession on average.

In total, 687.7 million trimmed reads were mapped to the *S. marianum* draft assembly using BWA-mem, which showed an 88.2% average mapping rate (Table 3). On average, 114.6 million read/samples were mapped. The mapping rate per each accession ranged from 85.3% (M02) to 90.0% (M01, M03). About 92.3 million reads were unmapped (11.8%). About 83.6% of the mapped reads were properly paired, accounting for 81.9 million to 109.9 million reads for each accession.

Table 3. Percentage of reads in each accession mapping to the reference genomes of *Silybum marianum* L. Gaertn.

Accession	Total Trimmed	Mapped	Unmapped	Properly Paired	Properly Paired (%)	Mapped (%)	Unmapped (%)	Properly Paired Mapped (%)
M01	108,554,670	97,690,066	10,864,604	81,978,182	75.5	90.0	10.0	83.9
M02	136,302,598	116,215,035	20,087,563	96,982,188	71.1	85.3	14.7	83.5
M03	146,110,578	131,474,969	14,635,609	109,901,174	75.2	90.0	10.0	83.6
M04	146,000,972	130,085,152	15,915,820	108,315,504	74.2	89.1	10.9	83.3
M05	121,009,526	104,304,345	16,705,181	86,915,402	71.8	86.2	13.8	83.3
M06	122,017,832	107,898,658	14,119,174	90,503,310	74.2	88.4	11.6	83.9
Total	779,996,176	687,668,225	92,327,951	574,597,760	Mean 73.7	88.2	11.8	83.6

2.3. Identification and Validation of Genome-Wide InDels

Although both insertions and deletions (InDels) and single-nucleotide polymorphisms (SNPs) were identified in this study, we focused on InDels for further analysis and discussion. In total, 3,518,667 raw InDels were called via GATK haplotypeCaller. After variant filtering, 238,988 InDel variants remained. In addition, after the strict filtering of heterozygotes for the reference allele, 30,845 InDels were grouped by modified allele frequency (Figure S2, Table 4). Of these, 3410 InDels exhibited three homozygotes for the reference allele and three homozygotes for the alternative allele, which refers to group ‘303’ polymorphic type. To select indel markers that were highly discriminatory and easy to use for profiling the genotypes of six resources, only the genotype group ‘303’ was used for InDels marker development in this study. Of them, 234 InDels with a size difference of more than 15 bp between the reference allele and alternative allele were selected for marker design (Table 4). Length differences less than 39 bp accounted for 82% of InDels. Three InDels with length differences of more than 100 bp (107 bp, 176 bp, and 248 bp) were deleted in this study. Therefore, a total of 231 InDels were finally used, consisting of 107 (46.3%) inserted InDels and 124 (53.7%) deleted InDels. All InDels were >200 bp inward from the beginning or end of a contig, comprising sufficient sequences available for primer design. The InDel size difference generated by the tested primers ranged from 30 to 217 bp. Of them, 177 (77%) InDels yielded a single PCR fragment and showed polymorphism, 33 (14%) did not amplify a product, and 21 (9%) showed no polymorphism (Table S4). Only 10 InDels showed a ‘303’ allele segregation type among the six accessions, as expected, and 100 InDels showed a heterozygous genotype in at least one of the six accessions. A total of 171 (74%) InDels from 231 InDels showed polymorphic bands between the two native accessions of Korea, M05, and M06.

Table 4. Polymorphic types among six accessions were grouped according to modified allele frequency, which is represented by a three-digit number. The first number of three digits represents the number of accessions with homozygotes for the reference allele among six accessions, the second digit represents the number of accessions with a heterozygote allele, and the third digit represents the number of accessions with homozygotes for the alternative allele, respectively. Distribution of various sizes of InDel polymorphism (more than 15 bp, more than 20 bp, and more than 25 bp) was identified among the six selected accessions of milk thistle.

Polymorphic Types	InDel	Size of InDel Polymorphism		
		≥15 bp	≥20 bp	≥25 bp
'105'	4750	259	165	114
'204'	3334	207	127	92
'303'	3410	234	161	115
'402'	6114	480	321	221
'501'	13,237	1161	777	557
Total	30,845	2341	1551	1099

2.4. Cluster Analysis and Fingerprinting of the Six Selected Accessions

The phylogenetic relationship was constructed with dendrogram coefficients using the numerical taxonomy system of multivariate programs (NTSYS) cluster analysis (Figure 3). An unweighted pair group method with an arithmetic mean (UPGMA) dendrogram was constructed for the six milk thistle accessions based on the 177 InDel polymorphisms, in which the Jaccard's similarity coefficients ranged from 0.36 (M05 vs. M06) to 0.61 (M01 vs. M03). A pair of M01 and M03 were estimated as having the highest genetic similarity. Two accessions native to Korea (M05 and M06) were clustered separately from the four accessions (M01–M04) that originated from Canada, Germany, North Korea, and Moldova, respectively. For fingerprinting of the six selected accessions, a minimum marker set was developed using the InDel markers system. A set of six InDels (SM034, SM026, SM102, SM135, SM182, and SM176) were screened out of 177 primers based on sharp, clear, and reproducible bands (Figure 4A,B), which completely discriminated all six accessions.

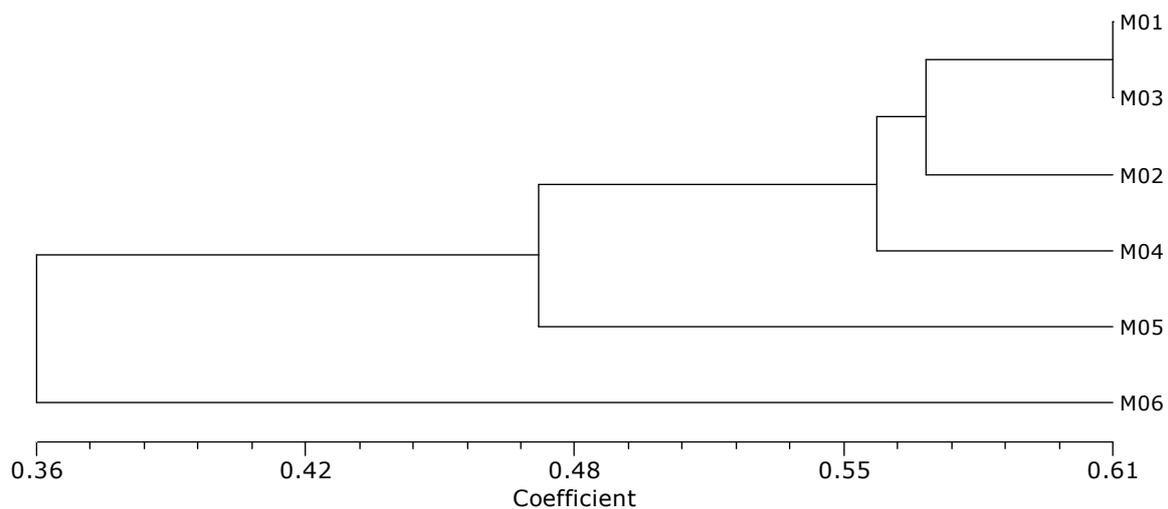


Figure 3. The phylogenetic relationship among the six selected accessions of milk thistle was constructed using NTSYS cluster analysis; a dendrogram showing the genetic similarity based on the 177 InDel polymorphisms.

A.

Primer ID	Forward primer	Reverse Primer	M01	M02	M03	M04	M05	M06
SM034	GGAATGGGTATGTCGGCAAT	GCCTGTCGATGTTATGTCTACC	AA	AA	AA	AA	AA	BB
SM026	AGTCGAAGTTGATGAGAGTTGA	AGTGATGGGTTAGTTGAGAGT	AA	AA	AA	AA	BB	AA
SM102	GCGCAATACATGGTCGGAAT	AAAGTTGGTGCAAACTGC	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	GTCGTCGTCCTCAACTAAACC	AGGTCACATGGTCACACGAT	BB	BB	AA	BB	AA	BB

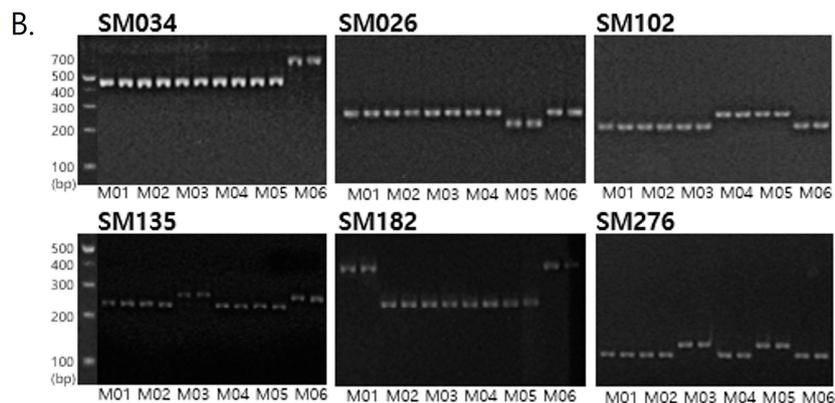


Figure 4. Minimum InDel marker set for fingerprinting the six selected accessions across six markers. (A) Primer list including primer ID, primer sequences, and allele types of six accessions. Homozygous reference alleles (AA) are shown in red, and homozygous alternative alleles (BB) in yellow. (B) Agarose gel images were analyzed via SM034, SM026, SM102, SM135, SM182, and SM176 from two samples of each accession, respectively.

3. Discussion

Milk thistle (*Silybum marianum* L. Gaertn.) is a medicinal plant that contains silymarin, a compound that is beneficial for people with chronic liver disease [1]. It is native to the Mediterranean Basin and was introduced as a crop in Europe, North and South America, Asia, and Southern Australia [10]. The leaves and achenes of milk thistle are used as raw materials for food in Korea. Milk thistle extract is also available as an over-the-counter (OTC) medication in Korea (Ministry of Food and Drug Safety, 2022). Currently, Korean pharmaceutical companies import all the milk thistle extracts they use. This reliance on imports exposes them to risks such as supply chain disruptions. By localizing and standardizing the production of milk thistle, Korean pharmaceutical companies can reduce their reliance on imports and mitigate these risks. Localizing production means that milk thistle will be grown and processed in Korea. Standardizing production means that there will be a consistent quality of milk thistle available in Korea. This will ensure that Korean pharmaceutical companies have a secure supply of milk thistle, even in the event of supply chain disruptions or other challenges.

Breeding high-quality milk thistles can be achieved through understanding the phenotypic variation and genetic diversity within the population. There are several challenges in milk thistle breeding that obstruct the cultivation and industrialization of milk thistle, such as spiny leaves, fruit dispersion, asynchronous flowering, unstable yield quality, and crop stability [4,31]. In this study, we investigated six traits associated with the breeding goals for milk thistle, over a three-year period (Table S2). In the phenotypic evaluation among the selected six accessions, the weight of 100 achenes was increased as the number of flower heads and involucre diameter increased. This is consistent with previous studies concluding that the number of seeds per plant increased, ranging from a 484 to 1359 head per plant increase [32]. Both Korean native accessions, M05 and M06, were morphologically and genetically distinctive from the other four accessions. In M05, both the weight of 100 achenes and the number of flower heads were the highest (Table 1). Shim et al. (2020)

reported that M05 contained the highest contents of silybin B in dried achenes (3.50 mg/g) among the six accessions, which is one of the major active constituents of silymarin [33,34]. Thus, M05 is a promising breeding material for milk thistle production in Korea. It has a number of desirable traits that make it well-suited to commercial production, including high yield, high silybin B content, and adaptability to a wide range of Korean climates and environments. Meanwhile, M06 showed a small spine length and small involucre diameter, along with the highest number of branches (Figure 2, Table 1). A particularly long spine length reduces work efficiency during cultivation and harvesting. The short spine length of M06 is considered a valuable trait for achieving optimal breeding target traits. M06 was genetically and morphologically distinct from other materials following phylogenetic analysis (Figure 3 and Figure S2). Overall, M06 has the potential to benefit commercial cultivation and breeding programs to reduce spine length.

Molecular breeding for milk thistle, as a non-model medicinal plant, is far behind that of model crops, as there are insufficient genome data and efficient molecular markers. A few studies on *Silybum marianum* employing RAPD, AFLP, DArT array, and ISSR primer systems were conducted [4,35–37]. Marker-assisted selection (MAS) is a breeding technique that uses molecular markers to identify and select for desired traits in plants [35]. MAS can improve the productivity and accuracy of classical plant breeding by reducing its time consumption. InDels and SNPs are the most widely used PCR-based marker systems in MAS [38]. InDels are derived from the insertion of retrotransposons or other mobile elements, unequal crossover events, or slippage in simple sequence replication [39]. It is known that mutations in DNA repair genes are also related to the occurrence of indels [40]. They have a low frequency of homoplasmy, indicating that there is an adequately low probability of two InDel mutations of exactly the same length occurring at the same genomic position [19]. InDels in genes with functional diversity between alleles are highly useful for marker-assisted selection or QTL mapping [41,42].

We conducted the first large-scale study of genome-wide InDel development in milk thistle (Table 3). Six milk thistle accessions, including two native Korean lines, were sequenced. After the strict filtering of heterozygotes for the reference allele, 30,845 InDels remained from the 238,988 InDel variants. Of the '303' type, 231 InDels were selected, expected to represent the most polymorphic allele type, and 177 InDels (77%) of comparatively high polymorphism percentages were observed among the six selected validation accessions (Table S3). These InDel primers were expected to serve as a high-potential tool for genetic discrimination among the *Silybum marianum* species. A novel marker set of 177 InDels with high amplification rates and high polymorphism can be utilized for genetic studies, such as pedigree analysis and seed purity test of a parental line and F1 hybrid (Figure 4). In allele types, there was a discrepancy in InDel sizes between the predicted and actual size. This discrepancy could be due to alignment to the primitive reference genome at the assembled contig level.

Phylogenetic analysis using 177 InDels clarified the relationship among the six genotypes (Figure 3). Genetically diverse parents have the possibility to generate higher heterosis within phenotypes [43]. In milk thistle, ecotypes of different geographical regions with various ratios of flavonolignan compounds constituted a gene pool for plant improvement [32]. There was high variability in silymarin content and composition among natural populations in Iran, Egypt, and Italy. Low variability among non-native populations in New Zealand was reported [44]. Several milk thistle varieties and lines have been developed and registered in Poland, Hungary, Germany, England, and New Zealand [32,45]. In order to breed useful milk thistle cultivars, it is necessary to investigate the agronomical traits, genetic diversity, and silymarin contents of the breeding materials.

Orphan crops are often native to a particular region and have been cultivated for centuries, but they have not received the same level of research and development as other popular crops [8]. Their germplasm collections are not complete, and lack a full understanding of the genetic diversity of the plants. Modern genetic and genomic tools can be used to improve crop breeding, which could help ensure global food and nutritional

security. In this study, we conducted a phenotypic evaluation of 220 milk thistle plants and developed the first large-scale InDel markers, which may serve as a foundation for breeding programs and genetic studies, including pedigree analysis, origin and evolutionary analysis, population structure and diversity analysis, QTL mapping, and marker-assisted selection.

4. Materials and Methods

4.1. Plant Materials

For the selection and standardization of domestic milk thistle resources, we collected 172 milk thistle plant seeds from local markets across Korea (Table S5). In addition, six accessions were distributed from the National Agrobiodiversity Center, National Institute of Agricultural Sciences, Rural Development of Administration (RDA), Korea. For phenotypic evaluation, a total of 220 accessions, composed of 172 accessions from local markets and 48 accessions isolated from 6 accessions, were grown in Hwaseong, Gyeonggi-do, Korea, in 2018 and 2019. Of them, six representative accessions were selected based on morphological characteristics. Four accessions (M01–M04) among the six selected accessions were distributed from RDA, originating from Canada, Germany, North Korea, and Moldova (Table 1). The two other accessions, M05 and M06, were collected from local markets in Korea. The six selected accessions were advanced by self-pollination and used for the resequencing analysis and phenotypic evaluation in 2019 to 2021.

4.2. Phenotypic Evaluation

To select a representative plant for consideration of its morphological characteristics and industrial value, six phenotypic traits were evaluated for 220 milk thistle accessions in 2018 and 2019, i.e., plant height (cm), seed weight (g), number of flower heads, seed weight(g)/ number of flower heads, spine length (mm), and plant color at harvest. Following the selection of six representatives, we evaluated six quantitative characteristics from 2019 to 2021, i.e., plant height (cm), number of branches, the diameter of involucre with spine tips (mm), number of flower heads, 100-achene weight (g), and spine length (mm). Statistical analyses were conducted to determine significant differences in six agronomic traits among six accessions, using analysis of variance (ANOVA) and Duncan's multiple range test in the R package. Frequency distributions were calculated for each trait using Microsoft Office Excel 2016. We used the statistical analysis package SPSS 12.0KO for Windows to calculate Pearson's correlation coefficients, principal component analysis (PCA), and hierarchical cluster analysis using Ward's method. Six agriculturally superior plants were selected that represented the phenotypic variation of traits and had the potential to be used as breeding materials from a breeder's perspective.

4.3. Genome Resequencing and Assembly

DNA from a single plant of each accession was extracted using the Cetyl Trimethyl Ammonium Bromide (CTAB) method [46]. Each DNA was quantified by NanoDrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA.), and only the high-quality DNA samples for genome sequencing were used. For the first step, high-depth resequencing was conducted on six milk thistle accessions to identify InDel markers. An Illumina paired-end (P.E.) library with a 400 bp insert size was constructed according to the manufacturer's recommendations, and the library was sequenced on Illumina Novaseq with 2×150 bp. Low-quality sequences (Phred score ≤ 20) and Illumina adapter sequences were removed in raw fastq files using Trimmomatic v.0.39 (<http://www.usadellab.org/cms/?page=trimmomatic> (accessed on 1st September 2021)). The reference genome sequence of *Silybum marianum* was downloaded from the NCBI database (Genbank assembly acc# GCA_001541825.1). The trimmed data were aligned to the reference genome using BWA-MEM (version 0.7.1.7).

4.4. Variant Calling and InDel Screening

The alignment data were transformed into a binary alignment map (BAM) format via SAMTools [47]. 'Mark duplicates' in the Picard tool (Broad Institute, Boston, MA,

USA) were applied to remove replicate reads. To reduce the inaccurate alignments, a GATK-HaplotypeCaller was used to conduct local realignment around the insertions and deletions, read base quality recalibration, and variant calling [48]. The dataset was further filtered using two approaches: (i) relaxed filtering for $\text{minDP} \geq 10$, $\text{minGQ} \geq 30$, missing data $\leq 20\%$ for each locus, and removing monomorphic variants among six milk thistle accessions; and (ii) stricter filtering for heterozygote and missing data. Then, sample genotypes were grouped according to modified allele frequencies, which are coded as a three-digit number. Genotype groups were represented as '105', '204', '303', '402', and '501' (Figure S2). The first digit represents the number of accessions with homozygotes for the reference allele among the six accessions, the second digit represents the number of accessions with the heterozygote allele, and the third digit represents the number of accessions for the alternative allele, respectively. Of the '303' group, 234 InDels were selected and PCR primers were designed from the flanking region using Primer3 [49]. Of them, 231 InDels with size differences between 15 bp and 100 bp were arbitrarily chosen for experimental resolvability under agarose gel electrophoresis.

4.5. PCR Amplification

For validating the identified InDels, PCR was performed on a SimpliAmp thermocycler (Thermo Scientific, Waltham, MA, USA) in a 20 μL reaction volume containing 50 ng of DNA template and 0.5 μL of each forward and reverse primer, making a total of 10 μL (Bioneer, Daejeon, Republic of Korea), 2 μL of $10 \times$ buffer, 0.5 μL of 2.5 mM dNTPs, and 0.1 μL of Taq polymerase (IN5001-0500; Inclone, Yongin, Republic of Korea), under the following conditions: initial denaturation at 94 °C for 4 min, followed by 35 cycles of denaturation at 94 °C for 1 min, annealing at 55 °C for 30–60 s, and extension at 72 °C for 30–60 s, with a final extension at 72 °C for 7 min. The PCR products were electrophoresed (BioFACT, Daejeon, Republic of Korea) on 1% or 2% agarose gel at 100–160 V for 20–40 min in a $0.5 \times$ TBE buffer. Gels were visualized using a gel imager (Korea Lab Tech, Seongnam, Republic of Korea). To consider the resolution of agarose gel, a single PCR product with a fragment ≥ 30 bp larger than the reference and alternative alleles was determined to comprise C and D alleles, according to the order of detection of the six selected accessions (M01–M06).

4.6. Genetic Diversity Assay

To generate molecular data matrices, the presence or absence of clear bands was scored in every accession for each primer pair, and recorded either as 1 (presence of a fragment) or 0 (absence of a fragment). A phylogenetic tree was conducted based on genetic distances and the unweighted pair group method with an arithmetic mean (UPGMA) using the Jaccard functionality of NTSYSpc Version 2.21 m (Exeter Software, Setauket, NY, USA).

5. Conclusions

Milk thistle (*Silybum marianum* L. Gaertn) is an important medicinal plant for chronic liver disease. We selected six accessions that showed significant differences in agricultural characteristics to be used as initial materials for genomic research into milk thistle. A large-scale development of genome-wide InDels using resequencing analysis was performed. As a result, 177 InDels with reliable polymorphisms were developed from the resequencing data of the six selected milk thistle accessions. These plant materials and InDels could be valuable resources for the identification, conservation, breeding programs, and industrialization of *Silybum marianum*.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/plants12142702/s1>, Figure S1. Hierarchical cluster analysis based on agronomical traits with dendrogram visualization using Ward's method via SPSS. Figure S2. An explanation of the modified allele frequency used for InDel classification. Sample genotypes were grouped according to the modified allele frequency, which is represented by a polymorphic type of a three-digit number. The first number of the three digits represents the number of accessions with

homozygotes for the reference allele among the six accessions, the second digit represents the number of accessions with a heterozygote allele, and the third digit represents the number of accessions with homozygotes for the alternative allele, respectively. Only the genotype group ‘303’ was used for InDel marker development in this study. Table S1. Correlations between six traits in the 220 milk thistle accessions. Six agronomical traits were evaluated: plant height (PH, cm), seed weight (SW, g), number of flower heads (FHN), SW/FHN, spine length (SL, mm), and plant color at harvest (PC). According to color charts, PC was estimated as being one of three types: green (1), brown (2), and green-brown (3). ** $p < 0.01$; * $p < 0.05$. Table S2. Phenotypic variation in the six traits in six milk thistle accessions measured in 2019–2021. Means separated within columns followed by different letters were significantly different based on Duncan’s test ($p \leq 0.05$). Table S3. Sequence read counts by preprocessing raw resequencing data and trimmed read data by fastq in six milk thistle accessions. Table S4. Details of the 198 InDel markers developed in this study. Table S5. Passport data of *Silybum marianum* collected in this study.

Author Contributions: Conceptualization, J.H.C., S.Y.H. and J.S. (Jungsook Sung); Data curation, O.N.L., J.S. (Jeehyoung Shim) and J.-H.H.; Resources, E.Y. and J.S. (Jungsook Sung); Investigation, J.S. (Jeehyoung Shim) and J.-H.H.; Methodology and Data analysis, Y.Y.; Writing—original draft, O.N.L., J.S. (Jeehyoung Shim), J.-H.H. and Y.Y.; Writing—review and editing, O.N.L. and J.H.C.; Funding acquisition, J.S. (Jeehyoung Shim); Supervision and Project administration, J.H.C. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by a collaboration project (Project No. PJ01418503) with the National Agrobiodiversity Center, National Institute of Agricultural Sciences, Rural Development Administration, Republic of Korea.

Data Availability Statement: All data that support the findings within this study are available at NCBI GenBank with the following accession numbers: PRJNA887484. This includes genomic sequencing data for the selected six accessions (M01–M06).

Acknowledgments: The authors are grateful to Sang Hyuk Moon and Yeong Ju Lee, Department of Integrative Biological Sciences and Industry, Sejong University, for DNA genotyping. We also express our gratefulness to Hye Gyeong An, Donggeon Kim, and Jaejun So in EL&I Co., Ltd., for their technical support on plant cultivation.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Abenavoli, L.; Izzo, A.A.; Milic, N.; Cicala, C.; Santini, A.; Capasso, R. Milk thistle (*Silybum marianum*): A concise overview on its chemistry, pharmacological, and nutraceutical uses in liver diseases. *Phytother. Res.* **2018**, *32*, 2202–2213. [\[CrossRef\]](#)
2. Fulvio, F.; Martinelli, T.; Paris, R. Selection and validation of reference genes for RT-qPCR normalization in different tissues of milk thistle (*Silybum marianum*, Gaert.). *Gene* **2021**, *768*, 145272. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Bijak, M. Silybin, a Major Bioactive Component of Milk Thistle (*Silybum marianum* L. Gaertn.)—Chemistry, Bioavailability, and Metabolism. *Molecules* **2017**, *22*, 1942. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Martinelli, T.; Potenza, E.; Moschella, A.; Zaccheria, F.; Benedettelli, S.; Andrzejewska, J. Phenotypic Evaluation of a Milk Thistle Germplasm Collection: Fruit Morphology and Chemical Composition. *Crop Sci.* **2016**, *56*, 3160–3172. [\[CrossRef\]](#)
5. Tutin, T.G.; Heywood, V.H.; Burges, N.A.; Valentine, D.H.; Walters, S.M.; Webb, D.A. *Flora Europaea: Plantaginaceae to Compositae (and Rubiaceae)*; Cambridge University Press: Cambridge, UK, 1964; Volume 4.
6. Hetz, E.; Liersch, R.; Schieder, O. Genetic investigations on *Silybum marianum* and *S. eburneum* with respect to leaf colour, outcrossing ratio, and flavonolignan composition. *Planta Med.* **1995**, *61*, 54–57. [\[CrossRef\]](#)
7. Andrzejewska, J.; Sadowska, K.; Mielcarek, S. Effect of sowing date and rate on the yield and flavonolignan content of the fruits of milk thistle (*Silybum marianum* L. Gaertn.) grown on light soil in a moderate climate. *Ind. Crops Prod.* **2011**, *33*, 462–468. [\[CrossRef\]](#)
8. Talabi, A.O.; Vikram, P.; Thushar, S.; Rahman, H.; Ahmadzai, H.; Nhamo, N.; Shahid, M.; Singh, R.K. Orphan Crops: A best fit for dietary enrichment and diversification in highly deteriorated marginal environments. *Front. Plant Sci.* **2022**, *13*, 839704. [\[CrossRef\]](#)
9. Gresta, F.; Avola, G.; Guarnaccia, P. Agronomic characterization of some spontaneous genotypes of milk thistle (*Silybum marianum* L. Gaertn.) in Mediterranean environment. *J. Herbs Spices Med. Plants* **2007**, *12*, 51–60. [\[CrossRef\]](#)
10. Marceddu, R.; Dinolfo, L.; Carrubba, A.; Sarno, M.; Di Miceli, G. Milk Thistle (*Silybum Marianum* L.) as a Novel Multipurpose Crop for Agriculture in Marginal Environments: A Review. *Agronomy* **2022**, *12*, 729. [\[CrossRef\]](#)
11. Carrier, D.J.; Crowe, T.; Sokhansanj, S.; Wahab, J.; Barl, B. Milk thistle, *Silybum marianum* (L.) Gaertn., flower head development and associated marker compound profile. *J. Herbs Spices Med. Plants* **2003**, *10*, 65–74. [\[CrossRef\]](#)

12. Martin, R.; Lauren, D.; Smith, W.; Jensen, D.; Deo, B.; Douglas, J. Factors influencing silymarin content and composition in variegated thistle (*Silybum marianum*). *N. Z. J. Crop Hortic. Sci.* **2006**, *34*, 239–245. [[CrossRef](#)]
13. Zheljajzkov, V.D.; Zhalnov, I.; Nedkov, N.K. Herbicides for weed control in blessed thistle (*Silybum marianum*). *Weed Technol.* **2006**, *20*, 1030–1034. [[CrossRef](#)]
14. Song, I. *Distribution Characteristic of Alien Plants in Seoul & Its Management Plan II*; Seoul Development Institute: Seoul, Korea, 2010; pp. 1–150, ISBN 9788980527557.
15. Schlötterer, C. The evolution of molecular markers—Just a matter of fashion? *Nat. Rev. Genet.* **2004**, *5*, 63–69. [[CrossRef](#)] [[PubMed](#)]
16. Collard, B.C.; Mackill, D.J. Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. B Biol. Sci.* **2008**, *363*, 557–572. [[CrossRef](#)] [[PubMed](#)]
17. Mohan, M.; Nair, S.; Bhagwat, A.; Krishna, T.; Yano, M.; Bhatia, C.; Sasaki, T. Genome mapping, molecular markers and marker-assisted selection in crop plants. *Mol. Breed.* **1997**, *3*, 87–103. [[CrossRef](#)]
18. Liu, B.; Wang, Y.; Zhai, W.; Deng, J.; Wang, H.; Cui, Y.; Cheng, F.; Wang, X.; Wu, J. Development of InDel markers for Brassica rapa based on whole-genome re-sequencing. *Theor. Appl. Genet.* **2013**, *126*, 231–239. [[CrossRef](#)]
19. García-Lor, A.; Luro, F.; Navarro, L.; Ollitrault, P. Comparative use of InDel and SSR markers in deciphering the interspecific structure of cultivated citrus genetic diversity: A perspective for genetic association studies. *Mol. Genet. Genom.* **2012**, *287*, 77–94. [[CrossRef](#)]
20. Mills, R.E.; Luttig, C.T.; Larkins, C.E.; Beauchamp, A.; Tsui, C.; Pittard, W.S.; Devine, S.E. An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res.* **2006**, *16*, 1182–1190. [[CrossRef](#)]
21. Powell, W.; Machray, G.C.; Provan, J. Polymorphism revealed by simple sequence repeats. *Trends Plant Sci.* **1996**, *1*, 215–222. [[CrossRef](#)]
22. Pan, Z.; Li, Z.; Zhang, J.; Bai, S.; Zhao, W.; Tong, C. Investigation of genome-wide InDel distribution and segregation in Populus with restriction-site associated DNA sequencing data. *Trop. Plant Biol.* **2022**, *15*, 171–180. [[CrossRef](#)]
23. Lü, Y.; Cui, X.; Li, R.; Huang, P.; Zong, J.; Yao, D.; Li, G.; Zhang, D.; Yuan, Z. Development of genome-wide insertion/deletion markers in rice based on graphic pipeline platform. *J. Integr. Plant Biol.* **2015**, *57*, 980–991. [[CrossRef](#)]
24. Ramakrishna, G.; Kaur, P.; Nigam, D.; Chaduvula, P.K.; Yadav, S.; Talukdar, A.; Singh, N.K.; Gaikwad, K. Genome-wide identification and characterization of InDels and SNPs in Glycine max and Glycine soja for contrasting seed permeability traits. *BMC Plant Biol.* **2018**, *18*, 141. [[CrossRef](#)] [[PubMed](#)]
25. Song, X.; Wei, H.; Cheng, W.; Yang, S.; Zhao, Y.; Li, X.; Luo, D.; Zhang, H.; Feng, X. Development of INDEL markers for genetic mapping based on whole genome resequencing in soybean. *G3 Genes Genomes Genet.* **2015**, *5*, 2793–2799. [[CrossRef](#)] [[PubMed](#)]
26. Shen, Y.-J.; Jiang, H.; Jin, J.-P.; Zhang, Z.-B.; Xi, B.; He, Y.-Y.; Wang, G.; Wang, C.; Qian, L.; Li, X. Development of genome-wide DNA polymorphism database for map-based cloning of rice genes. *Plant Physiol.* **2004**, *135*, 1198–1205. [[CrossRef](#)] [[PubMed](#)]
27. Li, W.; Cheng, J.; Wu, Z.; Qin, C.; Tan, S.; Tang, X.; Cui, J.; Zhang, L.; Hu, K. An InDel-based linkage map of hot pepper (*Capsicum annuum*). *Mol. Breed.* **2015**, *35*, 32. [[CrossRef](#)]
28. Liu, J.; Li, J.; Qu, J.; Yan, S. Development of genome-wide insertion and deletion polymorphism markers from next-generation sequencing data in rice. *Rice* **2015**, *8*, 27. [[CrossRef](#)]
29. Liu, J.; Qu, J.; Yang, C.; Tang, D.; Li, J.; Lan, H.; Rong, T. Development of genome-wide insertion and deletion markers for maize, based on next-generation sequencing data. *BMC Genom.* **2015**, *16*, 601. [[CrossRef](#)]
30. Rafizadeh, A.; Koohi-Dehkordi, M.; Sorkheh, K. Molecular insights of genetic variation in milk thistle (*Silybum marianum* [L.] Gaertn.) populations collected from southwest Iran. *Mol. Biol. Rep.* **2018**, *45*, 601–609. [[CrossRef](#)]
31. Smith, T.; Kawa, K.; Eckl, V.; Johnson, J. Sales of herbal dietary supplements in US increased 7.5% in 2015 consumers spent \$6.92 billion on herbal supplements in 2015, marking the 12th consecutive year of growth. *HerbalGram* **2016**, *111*, 67–73.
32. Karkanis, A.; Bilalis, D.; Efthimiadou, A. Cultivation of milk thistle (*Silybum marianum* L. Gaertn.), a medicinal weed. *Ind. Crops Prod.* **2011**, *34*, 825–830. [[CrossRef](#)]
33. Shim, J.; Han, J.-H.; Shin, N.-H.; Lee, J.-E.; Sung, J.-S.; Yu, Y.; Lee, S.; Ahn, K.H.; Chin, J.H. Complete Chloroplast Genome of a Milk Thistle (*Silybum marianum*) Acc.'912036'. *Plant Breed. Biotechnol.* **2020**, *8*, 439–444. [[CrossRef](#)]
34. Wang, X.-l.; Lin, F.-l.; Xu, W.; Wang, C.; Wang, Q.-q.; Jiang, R.-W. Silybin B exerts protective effect on cisplatin-induced neurotoxicity by alleviating DNA damage and apoptosis. *J. Ethnopharmacol.* **2022**, *288*, 114938. [[CrossRef](#)]
35. Khan, M.M.H.; Rafii, M.Y.; Ramlee, S.I.; Jusoh, M.; Al Mamun, M.; Halidu, J. DNA fingerprinting, fixation-index (Fst), and admixture mapping of selected Bambara groundnut (*Vigna subterranea* [L.] Verdc.) accessions using ISSR markers system. *Sci. Rep.* **2021**, *11*, 14527. [[CrossRef](#)] [[PubMed](#)]
36. Mohammadi, S.A.; Shokrpour, M.; Moghaddam, M.; Javanshir, A. AFLP-based molecular characterization and population structure analysis of *Silybum marianum* L. *Plant Genet. Resour.* **2011**, *9*, 445–453. [[CrossRef](#)]
37. Saghalli, A.; Farkhari, M.; Salavati, A.; Alamisaeid, K.; Abdali, A. Genetic diversity assessment of Milk Thistle (*Silybum marianum* L.) ecotypes using ISSR markers. *Agric. Biotechnol. J.* **2016**, *8*, 51–64.
38. Liu, S.; An, Y.; Tong, W.; Qin, X.; Samarina, L.; Guo, R.; Xia, X.; Wei, C. Characterization of genome-wide genetic variations between two varieties of tea plant (*Camellia sinensis*) and development of InDel markers for genetic research. *BMC Genom.* **2019**, *20*, 935. [[CrossRef](#)]
39. Britten, R.J.; Rowen, L.; Williams, J.; Cameron, R.A. Majority of divergence between closely related DNA samples is due to indels. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 4661–4665. [[CrossRef](#)]

40. Belfield, E.J.; Ding, Z.J.; Jamieson, F.J.; Visscher, A.M.; Zheng, S.J.; Mithani, A.; Harberd, N.P. DNA mismatch repair preferentially protects genes from mutation. *Genome Res.* **2018**, *28*, 66–74. [[CrossRef](#)]
41. Raman, H.; Raman, R.; Wood, R.; Martin, P. Repetitive indel markers within the ALMT1 gene conditioning aluminium tolerance in wheat (*Triticum aestivum* L.). *Mol. Breed.* **2006**, *18*, 171–183. [[CrossRef](#)]
42. Vasemägi, A.; Gross, R.; Palm, D.; Paaver, T.; Primmer, C.R. Discovery and application of insertion-deletion (INDEL) polymorphisms for QTL mapping of early life-history traits in Atlantic salmon. *BMC Genom.* **2010**, *11*, 156. [[CrossRef](#)]
43. Zohary, D.; Hopf, M. *Domestication of Plants in the Old World: The Origin and Spread of Cultivated Plants in West Asia, Europe and the Nile Valley*; Oxford University Press: New York, NY, USA, 2000.
44. Vaknin, Y.; Mogilevski, I. Adaptive variation in vegetative, reproductive and chemical traits of the Mediterranean *Silybum marianum* L., under desert-adjacent conditions. *Isr. J. Plant Sci.* **2019**, *66*, 112–126. [[CrossRef](#)]
45. Marmouzi, I.; Bouyahya, A.; Ezzat, S.M.; El Jemli, M.; Kharbach, M. The food plant *Silybum marianum* (L.) Gaertn.: Phytochemistry, Ethnopharmacology and clinical evidence. *J. Ethnopharmacol.* **2021**, *265*, 113303. [[CrossRef](#)] [[PubMed](#)]
46. Murray, M.; Thompson, W. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* **1980**, *8*, 4321–4326. [[CrossRef](#)] [[PubMed](#)]
47. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; Subgroup, G.P.D.P. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)] [[PubMed](#)]
48. Poplin, R.; Ruano-Rubio, V.; DePristo, M.A.; Fennell, T.J.; Carneiro, M.O.; Van der Auwera, G.A.; Kling, D.E.; Gauthier, L.D.; Levy-Moonshine, A.; Roazen, D. Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv* **2017**, 201178.
49. Rozen, S.; Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics Methods and Protocols*; Springer: Berlin/Heidelberg, Germany, 2000; pp. 365–386.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.