

Article

The Evaluation of Agronomic Traits of Wild Soybean Accessions (*Glycine soja* Sieb. and Zucc.) in Heilongjiang Province, China

Wei Li ^{1,2}, Mu Peng ³, Zhen Wang ¹, Yingdong Bi ², Miao Liu ², Ling Wang ², Shufeng Di ², Jianxin Liu ², Chao Fan ², Guang Yang ² and Deyue Yu ^{1,*}

- ¹ National Center for Soybean Improvement, State Key Laboratory of Crop Genetics and Germplasm Enhancement, College of Agriculture, Nanjing Agricultural University, Nanjing 210095, China; nuio-3@163.com (W.L.); zhen302@163.com (Z.W.)
- ² Crop Tillage and Cultivation, Institute of Heilongjiang Academy of Agricultural Sciences (HAAS), Harbin 150086, China; ydbi308@163.com (Y.B.); liumiao8349@163.com (M.L.); lingling6958@163.com (L.W.); dishufeng2021@163.com (S.D.); wendyliujx@163.com (J.L.); beyean@163.com (C.F.); ouyanghui@163.com (G.Y.)
- ³ College of Life Science, Northeast Forestry University, Harbin 150040, China; pengmu1025@163.com
- * Correspondence: dyyu@njau.edu.cn



Citation: Li, W.; Peng, M.; Wang, Z.; Bi, Y.; Liu, M.; Wang, L.; Di, S.; Liu, J.; Fan, C.; Yang, G.; et al. The Evaluation of Agronomic Traits of Wild Soybean Accessions (*Glycine soja* Sieb. and Zucc.) in Heilongjiang Province, China. *Agronomy* **2021**, *11*, 586. <https://doi.org/10.3390/agronomy11030586>

Academic Editors: Dan Mullan, Fengjie Sun and Gustavo Caetano-Anollés

Received: 25 January 2021
Accepted: 16 March 2021
Published: 19 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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/>).

Abstract: Wild soybean germplasm is distributed widely in China, particularly in Heilongjiang Province. In this study, 242 wild soybean accessions from four agricultural divisions in Heilongjiang Province were evaluated based on six qualitative and eight quantitative traits. Results showed that a large amount of variation occurred in these evaluated traits. Abundant qualitative traits included the wild type (78.51%), purple flowers (90.50%), needle leaf (39.26%), black seed (83.88%), brown hilum (52.07%), and mud film (87.60). Results of multivariate analysis based on quantitative traits showed that 100-seed weight, seeds weight per plant, number of seeds per plant, number of effective pods, and number of invalid pods were significantly different among samples ($p < 0.05$). A total of 27 germplasms were screened. Cluster analysis identified the 242 accessions into two groups, not following the geographical distribution pattern, with rich wild soybean resources revealed in the northern site. The present study indicated that wild soybean in Heilongjiang Province should be conserved in situ. The rich genetic diversity revealed in soybeans of different sites in Heilongjiang Province suggested its significant potential utilization in genetic improvement and breakthrough for soybean breeding. This information will help to exploit and conserve wild soybean accessions in Heilongjiang Province.

Keywords: *Glycine soja* Sieb. and Zucc.; morphological traits; agronomic traits; diversity; genetic differentiation

1. Introduction

Cultivated soybean (*Glycine max* (L.) Merr.) is one of the economically important crop plants worldwide, as food for both human and livestock [1]. Additionally, cultivated soybean is the main resource for the world's protein meal and produces more than half of oil seed globally [2]. Although the soybean yield per acre has increased in the past century, the production of soybean cannot meet the demand of the growing world population. Various types of technology have been developed to breed soybean lines with high yield [3,4]. However, as one of the factors affecting high yield of soybean, the narrow genetic base (i.e., bottleneck of genetic diversity) has hindered soybean improvement [5,6], significantly limiting the development of soybean lines with high yield, high resistance traits, and environmental stress tolerance. Thus, there is an urgent need to explore rich resources of genetic diversity, to improve soybean production [7,8]. To date, it is commonly believed that wild soybean contains important genes for the adaptation to different harsh environments

stressed by salt and insects. It is naturally expected that these genes in wild soybean can be re-introduced into domesticated soybeans due to the lack of breeding barrier between wild and cultivated soybeans.

As an important resource to improve the beneficial characteristics in domesticated soybeans, wild soybean (*Glycine soja* Siebold and Zucc.) contains high genetic diversity. It has been suggested that wild soybean is native to East Asia and is distributed along a broad geographic range. In China, soybean was domesticated BC 6000–9000 years ago [9] with a wide distribution of wild soybeans. In particular, soybean is one of the main crops in Heilongjiang Province located in the north of China, harboring a large number of wild soybean resources due to its unique geographical and ecological environment. In order to advance the soybean industry, the sustainable utilization of wild soybean resources will provide the foundation for soybean innovation and production in this area. The comprehensive evaluation of agronomic traits of wild soybean will enrich the genetic basis of soybean cultivation, make breakthrough in soybean breeding, enhance the ability of soybean stress resistance, ensure the sustainable development of the soybean industry, and protect and effective use of wild soybean resources [10–12]. In this review, we investigated the agronomic traits of 242 wild soybean samples, providing a strong foundation for future application of wild soybeans, to improve the quality of cultivated soybeans.

2. Materials and Methods

2.1. Plant Materials and Field Experiments

A total of 242 wild soybean (*G. soja*) germplasms were used in this study. The samples collected from 13 cities of Heilongjiang Province were geographically distributed in four regions, namely the northern (Region I), eastern (Region II), southern (Region III), and western (Region IV) sites of Heilongjiang Province (Table 1 and Figure 1). These four regions were categorized based on their characteristics of topography, soil, and climate. Specifically, Region I showed a cool and humid climate, which was characterized by wide and rounded mountains with wide and shallow river valleys. Region II belonged to an agricultural and pastoral area in a low and flat terrain with fertile soil and plenty of water resources. Region III contained complex landforms with various vegetation types, crisscross agriculture and forestry, and abundant water resources. In Region IV, the special type of topography and soil conditions were not conducive for tree growth with the meadow steppe as the main vegetation type.

Table 1. Collecting sites of 242 wild soybean samples analyzed in this study.

| Site (Region) | Sample Size | Code | Collection Site | Area |
|-----------------------|-------------|---|--|-------------------------|
| Northern (Region I) | 88 | 42–57, 59–83, 85–89, 91–92, 103, 121, 123–125, 157–159, 161–163, 165–174, 189, 204, 206–210, 224–226, 254–256, 258–261, 264–265 | Yichun, Heihe, Daxinganling | 140,162 km ² |
| Eastern (Region II) | 51 | 109, 117–118, 130–133, 138, 156, 190–195, 197, 211–218, 227–229, 237, 242, 243, 250, 251, 263, 269, 274 | Jiamusi, Shuangyashan, Qitaihe, Jixi, Mudanjiang, Hegang | 142,493 km ² |
| Southern (Region III) | 64 | 3, 8, 11–12, 14–16, 18, 22–23, 25–35, 105–108, 116, 126–127, 129, 134–137, 139–143, 148–150, 152–155, 219–221, 231–232, 236, 240, 245–249, 252, 253, 267, 270–273 | Suihua, Harbin | 89,051 km ² |
| Western (Region IV) | 39 | 1, 9–10, 93–97, 99–101, 110–115, 146–147, 175–188, 196, 198, 238, 257, 262, 266 | Qiqihar, Daqing | 65,161 km ² |



Figure 1. Geographic distribution of 242 wild soybean samples in Heilongjiang Province.

Based on the information of four major agricultural regions in Heilongjiang Province, all wild soybean resources were collected and preserved by the field investigation (Table 1). Seeds were obtained by self-breeding. The growth habits of wild soybean were different from those of the cultivated soybeans, including stolon habits, high plant height, and thin stems. Thus, in this study, a completely randomized block design (i.e., a $0.65\text{ m} \times 1\text{ m}$ plot containing four holes) was utilized with three replications for each sample. One seedling was sown in each hole. About 2 weeks after germination, routine management was performed throughout the experiment. To avoid plant twist, the growth of each plant was supported by bamboo sticks.

All experiments were carried out at the Academy of Agricultural Sciences National Agricultural Demonstration Zone of Heilongjiang Province, in the summers of 2012 and 2013. A total of 14 agronomic traits were investigated on the soybean samples (Figure 2). The six qualitative traits were evolutionary type (“wild” as recognized by 100-seed weight less than or equal to 3 g with main stem and “semi-wild” identified as 100-seed weight $> 3.01\text{ g}$ without main stem), flower color (white and purple), leaf shape (needle, oval, ellipse, and linear), seed color (yellow, green, black, brown, and dichromatism), hilum color (black and brown), and bloom habit (mud film, mud-free film, and lustrous). The eight quantitative traits collected at maturity were 100-seed weight (g), number of seeds per plant, number of effective pods (with seeds) per plant, number of invalid pods (without seeds) per plant, number of branches (with more than two mature pods) per plant, number of nodes (from cotyledon to apex of plant) per plant, and internode length (cm; between two adjacent nodes).

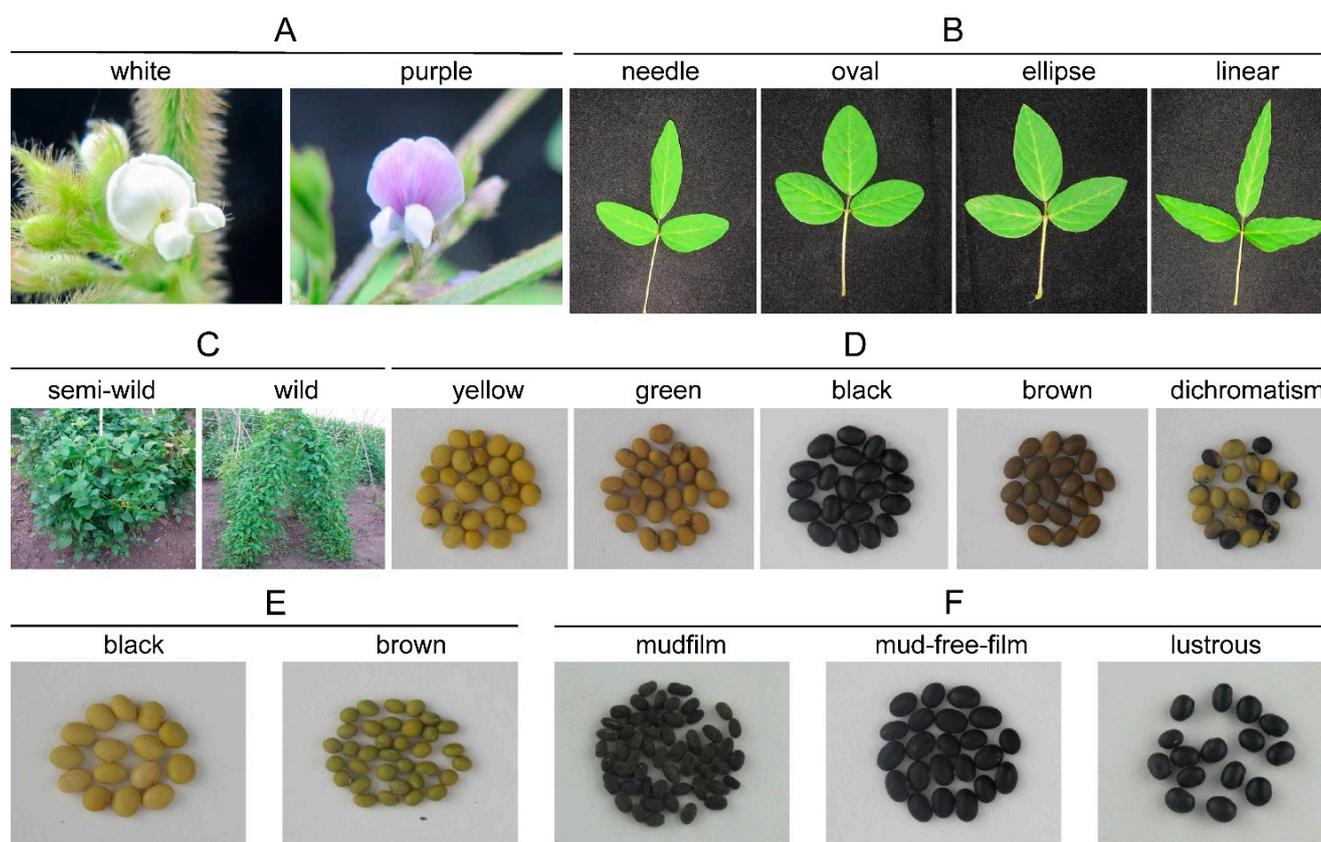


Figure 2. Qualitative traits, namely flower color (A), leaf shape (B), evolutionary type (C), seed color (D), hilum color (E), and bloom habit (F), investigated for 242 wild soybean samples in this study.

2.2. Data Processing and Statistics

The values of the variation coefficient, minimum, maximum, average, and genetic diversity index of each trait were calculated by Excel software. The genetic diversity index was calculated based on Shannon–Weave index $H' = -\sum P_i \ln P_i$ [12]. Correlation analysis, principal component analysis and cluster analysis were carried out by SPSS software. Significant differences were evaluated by using one-way ANOVA and Duncan's test at $p \leq 0.05$.

3. Results

3.1. Statistical Analysis of Qualitative Traits

The information of qualitative traits on 242 wild soybean samples are listed in Table 2 and Supplementary Materials Table S1. These results indicated that the abundant traits included wild type (78.51%), purple flower (90.50%), needle leaves (39.26%), black seed color (83.88%), brown hilum (52.07%), and mud film (87.60%). The diversity index varied from 0.31 (flower color) to 1.14 (leaf shape), with an average of 0.62.

In addition, we also calculated the percentage of different qualitative traits of 242 samples from all four sites (Table 3). The samples from the northern site contained mainly purple flower (96.59%), needle leaf (67.05%), black seed color (98.86%), brown hilum color (62.5%), and mud film (96.59%), while most accessions from the western site showed purple flower (97.44%), ellipse leaf (53.85%), black seed color (97.44%), black hilum color (61.54%), and mud film (97.44%). The germplasms from the eastern site were dominated mainly by purple flower (84.31%), ellipse leaf (49.02%), black seed color (68.63%), brown hilum color (52.94%), and mud film (76.47%). Most samples from the southern site contained purple flower (82.81%), oval leaf (53.13%), black seed color (67.19%), brown hilum color (52.69%), and mud film (78.13%).

Table 2. Summary of six qualitative traits of 242 wild soybean samples in this study.

| | Trait | Number (n) | Percent (%) | Diversity Index (H') |
|-------------------|---------------|------------|-------------|----------------------|
| Evolutionary type | wild | 190 | 78.51 | 0.52 |
| | semi-wild | 52 | 21.49 | |
| Flower color | purple | 219 | 90.50 | 0.31 |
| | white | 23 | 9.50 | |
| Leaf shape | oval | 67 | 27.68 | 1.14 |
| | needle | 95 | 39.26 | |
| | ellipse | 77 | 31.82 | |
| | linear | 3 | 1.24 | |
| Seed color | black | 203 | 83.88 | 0.61 |
| | brown | 20 | 8.26 | |
| | yellow | 3 | 1.24 | |
| | green | 2 | 0.83 | |
| | dichromatism | 14 | 5.79 | |
| Hilum color | black | 116 | 47.93 | 0.69 |
| | brown | 126 | 52.07 | |
| Bloom | mud film | 212 | 87.60 | 0.42 |
| | mud-free film | 3 | 1.24 | |
| | lustrous | 27 | 11.16 | |
| | Mean | | | 0.62 |

H', diversity index.

Table 3. Summary of 6 qualitative traits of 242 wild soybean samples based on four collecting sites.

| Site (Region) | | Northern Site (Region I) | Eastern Site (Region II) | Southern Site (Region III) | Western Site (Region IV) |
|-------------------|---------------|----------------------------|----------------------------|----------------------------|----------------------------|
| Trait | | Number (n)/ Percent (%) | Number (n)/ Percent (%) | Number (n)/ Percent (%) | Number (n)/ Percent (%) |
| Evolutionary type | wild | 82/93.18 | 30/58.82 | 43/67.19 | 35/89.74 |
| | semi-wild | 6/6.82 | 21/41.18 | 21/32.81 | 4/10.26 |
| Flower color | purple | 85/96.59 | 43/84.31 | 53/82.81 | 38/97.44 |
| | white | 3/3.41 | 8/15.69 | 11/17.19 | 1/2.56 |
| Leaf shape | oval | 8/9.09 | 19/37.25 | 34/53.13 | 6/15.38 |
| | needle | 59/67.05 | 7/13.73 | 17/26.56 | 12/30.77 |
| | ellipse | 18/20.45 | 25/49.02 | 13/20.31 | 21/53.85 |
| | linear | 3/3.41 | | | |
| Seed color | black | 87/98.86 | 35/68.63 | 43/67.19 | 38/97.44 |
| | brown | 1/1.14 | 8/15.67 | 10/15.63 | 1/2.56 |
| | yellow | | 1/1.96 | 2/3.13 | |
| | green | | 1/1.96 | 1/1.56 | |
| | dichromatism | | 6/11.76 | 8/12.5 | |
| Hilum color | black | 33/37.50 | 24/47.06 | 35/54.69 | 24/61.54 |
| | brown | 55/62.50 | 27/52.94 | 29/45.31 | 15/38.46 |
| Bloom | mud film | 85/96.59 | 39/76.47 | 50/78.13 | 38/97.44 |
| | mud-free film | 3/3.41 | | | |
| | lustrous | | 12/23.53 | 14/21.88 | 1/2.56 |

3.2. Statistical Analysis of Agronomic Traits

Several agronomic traits showed a wide range of variability in all wild soybean samples, in particular seed weight per plant (1.98–95.75 g in 2012, and 0.38–88.53 g in 2013), number of seeds per plant (126.33–5963.67 g in 2012, and 32.0–6204.00 in 2013), number of effective pods (16.67–1242.33 in 2012, and 30.00–973.33 in 2013), number of invalid pods (1.33–148.33 in 2012, and 0.67–115.33 in 2013), and number of nodes (29.33–657.33 in 2012

and 15.67–428.00 in 2013). Overall, the samples showed significantly ($p < 0.05$) higher mean values in 2012, compared to 2013, for seed weight per plant (24.51 and 20.41 g, respectively), number of effective pods (414.86 and 293.11, respectively), number of invalid pods (22.76 and 22.05, respectively), number of nodes (185.09 and 152.69, respectively) and internodes length (8.01 cm and 6.39 cm, respectively) (Table 4 and Supplementary Materials Table S2).

Table 4. Mean, range, coefficient variation, and diversity index of the agronomic traits evaluated in years 2012 and 2013.

| Trait | 2012 | | | | 2013 | | | |
|-------------------------------|-------------------------------|----------------|--------|------|-------------------------------|---------------|--------|------|
| | Mean | Range | CV (%) | H' | Mean | Range | CV (%) | H' |
| 100-seed weight (g) | 2.51 ± 1.91 ^a | 0.83–12.23 | 76 | 1.65 | 2.30 ± 1.76 ^a | 0.79–9.63 | 77 | 1.51 |
| Seeds weight per plant (g) | 24.51 ± 19.43 ^a | 1.98–95.75 | 79 | 1.77 | 20.41 ± 19.14 ^b | 0.38–88.53 | 94 | 1.7 |
| Number of seeds per plant (n) | 1102.08 ± 797.02 ^a | 126.33–5963.67 | 72 | 1.74 | 1019.54 ± 952.84 ^a | 32.00–6204.00 | 93 | 1.67 |
| Number of effective pods (n) | 414.86 ± 210.86 ^a | 16.67–1242.33 | 51 | 1.99 | 293.11 ± 180.28 ^b | 30.00–973.33 | 62 | 1.90 |
| Number of invalid pods (n) | 22.76 ± 18.19 ^a | 1.33–148.33 | 80 | 1.73 | 22.05 ± 17.74 ^b | 0.67–115.33 | 80 | 1.73 |
| Number of branches (n) | 5.18 ± 2.03 ^a | 1.00–11.67 | 39 | 2.07 | 5.44 ± 1.88 ^a | 1.67–11.67 | 34 | 1.95 |
| Number of nodes (n) | 185.09 ± 104.02 ^a | 29.33–657.33 | 56 | 1.85 | 152.69 ± 83.46 ^b | 15.67–428.00 | 55 | 2.09 |
| Internode length (cm) | 8.01 ± 1.69 ^a | 2.70–13.17 | 21 | 1.90 | 6.39 ± 1.68 ^b | 1.93–12.97 | 26 | 1.64 |
| Mean | | | 64 | 1.80 | | | 67 | 1.68 |

Means followed by the same letters in the same line are not significantly different ($p > 0.05$), while means followed by different letters in the same row are significantly different ($p < 0.05$). CV, coefficient variation; H', diversity index.

A total of five agronomic traits (i.e., 100-seed weight, seed weight per plant, number of seeds per plant, number of effective pods, and number of invalid pods) were significantly different between samples ($p < 0.05$) (Table 5). Samples in the southern site showed the highest 100-seed weight (3.26 g), seeds weight per plant (30.03 g), and number of branches (6.00 g). In contrast, the northern site was characterized by the lowest 100-seed weight per plant (1.62 g), seed weight per plant (11.07 g), number of effective pods (219.75), number of invalid pods (18.56), number of branches (4.72), number of nodes (133.91), and internodes length (6.07 cm). The samples from western site and eastern site showed intermediate values for 100-seed weight (1.67 and 2.75 g, respectively), seed weight per plant (16.57 and 27.38 g, respectively), and number of branches (5.42 and 5.97, respectively).

Table 5. Mean and range of the agronomic traits in four collecting sites, based on the means in the years of 2012 and 2013.

| Trait | Mean | | | | Range | | | |
|-------------------------------|---------------------|----------------------|----------------------|---------------------|---------------|---------------|----------------|---------------|
| | Northern Site | Western Site | Eastern Site | Southern Site | Northern Site | Western Site | Eastern Site | Southern Site |
| 100-seed weight (g) | 1.62 ^b | 1.67 ^b | 2.75 ^a | 3.26 ^a | 0.92–4.36 | 0.79–5.08 | 0.98–7.05 | 0.87–9.63 |
| Seeds weight per plant (g) | 11.07 ^c | 16.57 ^b | 27.38 ^a | 30.03 ^a | 0.38–71.27 | 0.52–44.85 | 3.94–75.29 | 1.34–88.53 |
| Number of seeds per plant (n) | 899.77 ^c | 1562.26 ^a | 1079.62 ^b | 805.63 ^b | 57.33–2652.00 | 32.33–6204.00 | 121.61–3972.33 | 32.00–2989 |
| Number of effective pods (n) | 219.75 ^b | 304.07 ^a | 354.46 ^a | 338.41 ^a | 30.00–731.67 | 32.33–800.33 | 83.67–973.33 | 63.67–905.67 |
| Number of invalid pods (n) | 18.56 ^b | 23.58 ^a | 26.65 ^a | 22.25 ^a | 0.67–115.33 | 2.00–101.00 | 7.00–107.67 | 2.67–57.67 |
| Number of branches (n) | 4.72 ^a | 5.42 ^a | 5.97 ^a | 6.00 ^a | 1.67–11.67 | 2.67–11.67 | 3.00–11.00 | 3.33–11.67 |
| Number of nodes (n) | 133.91 ^a | 158.64 ^a | 175.83 ^a | 156.45 ^a | 15.67–428 | 34.33–380.33 | 67.00–425.33 | 43.33–391.33 |
| Internodes length (cm) | 6.07 ^a | 6.42 ^a | 6.74 ^a | 6.54 ^a | 2.73–11.93 | 1.93–10.13 | 3.47–12.17 | 3.83–12.97 |

Means followed by the same letters in the same line are not significant at $p > 0.05$, while means followed by different letters in the same row are significant at $p < 0.05$.

3.3. The Principal Component, Cluster and the Correlation Analysis

The results of principal component analysis (PCA) based on all agronomic traits were shown in Figure 3 and Table 6. The first principle component (PC1) explained 35.04% of the total variation with the number of effective pods per plant, number of nodes, and number of branches as the most important variables (PC1 of 0.795, 0.785 and 0.655, respectively) in PC1, while 17.90% of the total variation was attributed to the second principle component (PC2), with the samples showing difference in internode length (0.160) and number of invalid pods (0.190). The third principle component (PC3) accounted for an additional 12% of the total variation attributed to number of seeds per plant for positive loadings (0.380) and internode length for negative loadings (-0.230). In Figure 3, all samples were clustered into two groups (Group I and Group II). There were 206 samples and 36 samples in Group I and Group II, respectively.

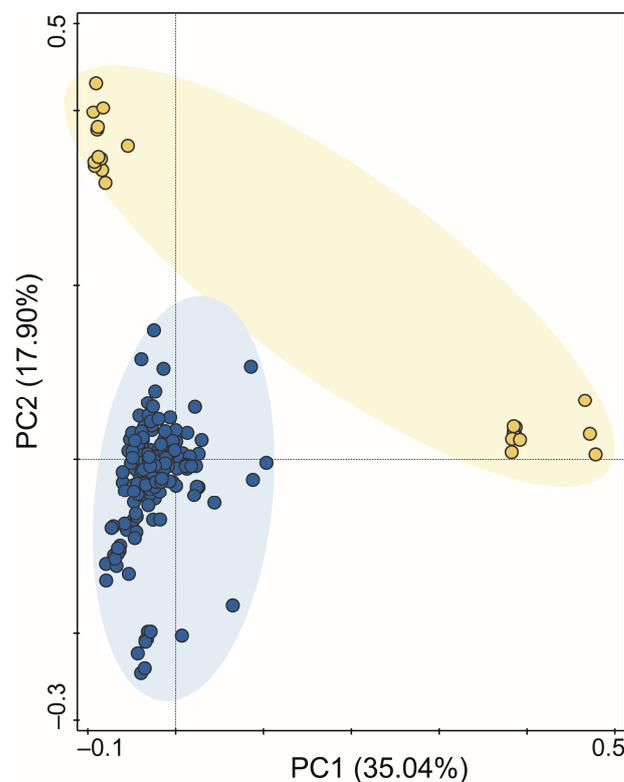


Figure 3. Graph of principal components one and two of the 242 wild soybean samples based on agronomic traits. Two groups are revealed, with 206 samples (blue dots) in Group I and 36 (yellow dots) in Group II, respectively.

The 242 wild soybean samples were revealed in two groups (Group I and Group II), based on eight agronomic traits, with 206 samples in Group I (85.12%) and 36 in Group II (14.88%), respectively (Table 7 and Figure 3). The average of variation in Groups I and II was 50.92% and 41.30%, respectively. The values of number of effective pods per plant, number of seeds per plant, and number of nodes in Group II were significantly higher than those of Group I, respectively.

The correlation coefficients between all agronomic traits are listed in Table 8. No significant correlation was observed between 100-seed weight and number of effective pods per plant ($r = 0.04$) and between 100-seed weight and number of nodes ($r = -0.55$). The 100-seed weight showed significantly negative correlation with number of seeds per plant ($r = -0.18$). Most of other agronomic traits showed significantly positive correlations with each other ($p < 0.01$), except for between number of branches and number of seeds

per plant ($r = 0.15$, $p < 0.05$) and between number of seeds per plant and internodes length ($r = 0.16$, $p < 0.16$), indicating that a close linkage occurred between these agronomic traits.

Table 6. Factor loadings (principle component one (PC1), PC2, and PC3) of the principal component analysis, based on the agronomic traits in 242 wild soybean samples, based on the mean data in the years of 2012 and 2013.

| Traits | PC1 | PC2 | PC3 |
|------------------------------------|-------|--------|--------|
| Number of branches | 0.655 | 0.115 | 0.312 |
| Number of effective pods per plant | 0.795 | −0.115 | −0.100 |
| 100-seed weight | 0.305 | 0.065 | 0.165 |
| Number of seeds per plant | 0.420 | 0.060 | 0.380 |
| Number of nodes | 0.785 | 0.075 | −0.140 |
| Internode length | 0.455 | 0.160 | −0.230 |
| Number of invalid pods | 0.465 | 0.190 | 0.065 |
| Standard deviation | 1.805 | 1.430 | 0.995 |
| Proportion of variance | 35.04 | 17.895 | 12.44 |
| Cumulative proportion | 35.04 | 52.935 | 63.37 |

Table 7. Cluster analysis in agronomic traits in 242 wild soybean samples based on the mean data of the years of 2012 and 2013. Data are presented as mean \pm standard deviation. CV, coefficient variation.

| Traits | Group I | | Group II | |
|------------------------------------|----------------------------------|--------|-----------------------------------|--------|
| | Mean \pm SD | CV (%) | Mean \pm SD | CV (%) |
| Number of branches | 5.24 \pm 1.55 ^a | 29.69 | 5.74 \pm 1.93 ^a | 33.58 |
| Number of effective pods per plant | 329.99 \pm 140.48 ^b | 42.57 | 491.29 \pm 165.11 ^a | 33.61 |
| 100-seed weight | 2.54 \pm 1.85 ^a | 72.87 | 1.65 \pm 1.34 ^b | 81.46 |
| Seed weight per plant | 21.91 \pm 17.88 ^a | 81.61 | 26.28 \pm 13.69 ^a | 52.09 |
| Number of seeds per plant | 847.74 \pm 431.63 ^b | 50.92 | 2263.54 \pm 647.56 ^a | 28.61 |
| Number of nodes | 157.77 \pm 65.97 ^b | 41.82 | 232.53 \pm 92.15 ^a | 39.63 |
| Internode length | 7.15 \pm 1.44 ^a | 20.20 | 7.50 \pm 1.12 ^a | 14.99 |
| Number of invalid pods | 21.72 \pm 14.70 ^a | 67.69 | 26.36 \pm 12.23 ^a | 46.40 |
| Average | | 50.92 | | 41.30 |

Means followed by the same letters in the same line are not significant different ($p > 0.05$), while means followed by different letters in the same row are significant different ($p < 0.05$).

Table 8. The correlation coefficients among the agronomic traits in 242 wild soybean samples based on mean data in the years of 2012 and 2013.

| Trait | Number of Branches | Number of Effective Pods per Plant | 100-Seed Weight | Seed Weight per Plant | Number of Seeds per Plant | Number of Nodes | Internodes Length |
|------------------------------------|--------------------|------------------------------------|-----------------|-----------------------|---------------------------|-----------------|-------------------|
| Number of effective pods per plant | 0.40 ** | | | | | | |
| 100-seed weight | 0.48 ** | 0.04 | | | | | |
| Seed weight per plant | 0.60 ** | 0.55 ** | 0.43 ** | | | | |
| Seeds per plant | 0.15 * | 0.43 ** | −0.18 ** | 0.19 ** | | | |
| Nodes | 0.38 ** | 0.69 ** | −0.55 | 0.40 ** | 0.51 ** | | |
| Internodes length | 0.22 ** | 0.29 ** | 0.24 ** | 0.41 ** | 0.16 * | 0.34 ** | |
| Invalid pods | 0.33 ** | 0.36 ** | 0.17 ** | 0.37 ** | 0.19 ** | 0.43 ** | 0.27 ** |

Notes: * and ** indicate significant difference set at $p < 0.05$ and $p < 0.01$, respectively.

3.4. Mining and Screening of Specific Germplasms

A total of 27 specific germplasms were screened, based on eight agronomic traits (Table 9), including eight samples showing high 100-seed weight (012, 020, 022, 025, 35, 242, 243, and 246); three samples showing more number of branches (003, 093, and 198); two with low 100-seed weight (104 and 187), less number of branches (065 and 083), high number of seeds per plant (097 and 263), high number of nodes (263 and 267), more invalid pods (190 and 238), and high internode length (061 and 197), respectively; and one with low internode length (141), less number of nodes (175), less number of seeds per plant (070), and less invalid pods (141), respectively.

Table 9. Screening of specific samples based on standard values of eight agronomic traits in 242 wild soybean samples in this study.

| Traits | Standard | No. of Samples | Standard | No. of Samples |
|---------------------------|----------|------------------------------------|----------|----------------|
| Number of branches | >9.00 | HAAS003/093/198 | <2.00 | HAAS065/083 |
| Number of effective pods | >700.00 | HAAS137/198/263 | <57.00 | HAAS174/175 |
| 100-seed weight | >6.44 | HAAS012/020/022/025/35/242/243/246 | <1.06 | HAAS104/187 |
| Seeds weight per plant | >68.76 | HAAS003 | <1.38 | HAAS174/175 |
| Number of seeds per plant | >2809.00 | HAAS097/263 | <112.00 | HAAS070 |
| Number of nodes | >399.00 | HAAS263/267 | <51.00 | HAAS175 |
| Internodes length | >9.83 | HAAS061/197 | <3.83 | HAAS141 |
| Number of invalid pods | >60.00 | HAAS190/238 | <4.00 | |

4. Discussion

The study of genetic diversity is important and necessary for understanding gene flow, evolutionary history, and time of domestication of crop plants [13]. In this study, our objective was to identify important phenotypic characteristics of wild soybeans from Heilongjiang Province for breeding of soybeans. Genetic diversity provides useful information about the diversity of plant species and greatly helps us systematically study crop breeding and understand evolutionary relationships [14]. Furthermore, understanding of the genetic diversity of wild germplasms can reveal the genetic structure, evolution, and geographical distribution of crops, and identify social factors that guide the breeding of cultivated crops. Moreover, wild germplasms with high genetic variations ultimately provide excellent genetic resources to improve the beneficial characteristics of cultivated soybeans [15]. Generally, the wild soybean is considered to be the direct progenitor of cultivated soybeans, suggesting the importance of wild soybeans in breeding programs of soybean improvement [16].

The genetic diversity of the wild soybeans has been studied based on morphological traits [17–19]. In this study, we investigated the genetic diversity of 242 wild soybean samples from four geographic sites in Heilongjiang Province, based on both qualitative and agronomic traits, in order to advance our understanding of wild soybean germplasms and their conservation strategies. All wild habitats of soybeans were considered in the unique collecting sites. Previous studies have explored the genetic diversity of wild soybeans, based on limited samples in Heilongjiang Province, based on ecological characteristics [18,20]. These results showed rich genetic diversity in wild soybeans, which could be used to improve the characteristics of cultivated soybeans. Similarly, other studies have analyzed the genetic diversity of the wild soybean from Heilongjiang Province, but based on limited populations and lacking sufficient experimental replicates [21,22]. In our study, we investigated the genetic diversity of wild soybeans based on a comprehensive dataset of a total of 242 samples collected in four agricultural sites from Heilongjiang province, using six qualitative traits and eight quantitative traits, for two consecutive years (2012 and 2013) at the Academy of Agricultural Sciences National Agricultural Demonstration Zone of Heilongjiang Province. The results of clustering based on PCA showed that all samples were divided into two groups, suggesting that some samples were clustered together, but more mixed samples were also clustered together (Figure 3). These results may be

attributed to the fact that the agronomic variations were regulated by gene, environment, and gene–environment interactions. Furthermore, the size of samples and the collecting locations can also lead to the differentiation of the wild soybean populations.

The complicated geographic environment is formed in Heilongjiang Province, stretching across 10 longitudes and 14 latitudes, locally harboring a variety of ecotypes of wild soybeans, due to its unique ecological conditions in this region [23]. Studies have shown that the high genetic diversity of wild soybeans derived from long-term selection has generated the adaptation of wild soybeans in various types of ecological habitats [24]. In the present study, results showed that higher values of several agronomic traits (i.e., 100-seed weight, seed weight per plant, number of seeds per plant, number of effective pods, number of invalid pods, number of nodes, and internode length) were observed in 2012 than those in 2013. These results may be attributed to the low temperature and wet weather in 2013, which negatively affected the growth of wild soybeans during growing seasons [25]. In this study, the morphological traits of wild soybean samples were compared among the samples collected from four geographical distribution of agriculture on wild soybean (Agricultural Division Office of Heilongjiang Province, 1985). Results showed that most of samples contained needle leaf, small seed, no evident main stem, low 100-seed weight, and the high diversity index, while these traits are generally considered as typical wild characteristics. These results are consistent with those reported previously [26]. The low genetic diversity of qualitative traits in 242 soybean samples was suggested by the diversity index H' of 0.62 (Table 2), which is lower than that reported previous [27], while significant difference was revealed in qualitative traits among four agricultural sites, suggesting that strong association between the geographic distribution and genetic clustering (Table 3). Specifically, the northern site of Heilongjiang Province is located in the Da Xing An Ling and Xiao Xing An Ling regions, where the mountains are round and vast, while the climate is cold and the natural environments are scarcely populated. Compared with other sites, the northern site is characterized as having a shorter growing season for crops, due to limited terrain, climate conditions, and human factors [28]. It is expected that the northern site may contain rich wild soybean resources. This expectation is supported by our results, showing rich wild soybean resources in the northern site. Therefore, it is recommended that the wild soybean resources in the northern site should be protected in situ. Previous studies have shown that Southern China might be the major center of diversity of the annual wild soybeans, simply due to frequent artificial partitions in wild soybeans among the southern provinces, in addition to natural selection [29]. Furthermore, the wild soybeans in other three collecting sites showed ellipse and oval leaves with big seeds, white flowers, and a main stem, which may be attributed to the rapid agricultural development and human intervention accelerating the evolution of these wild soybeans, as reported previously [26].

In the past decade, the selection and conservation of wild soybeans have attracted increasing attention worldwide, particularly because they are highly adaptive to various harsh environments compared to cultivated soybeans [27]. It is well-known that wild soybeans contain the sufficient genetic variations to adapt to many geographic, abiotic, and biotic environmental conditions [29]. Therefore, the wild soybean germplasms are generally considered as a potential genetic bank for cultivated soybeans, to improve their coping with climate change in the future. Our results showed that several agronomic traits were positively correlated with each, such as number of effective pods per plant and seed weight per plant, number of seeds per plant and number of effective pods per plant, and 100-seed weight and seed weight per plant, indicating the strong association between these agronomically beneficial traits; that is, these traits are linked to high yield of soybeans [30]. Thus, it is recommended that these traits be selected for future soybean-yield-enhancement breeding program.

The genetic variations have been established in the course of evolution, in response to adaptation to the environmental changes, generating some typical phenotypic traits, which are, in turn, used as fundamental standards for investigation of genetic diversity and screening of desired soybean resources [31]. For example, significant phenotypic

variations in wild soybeans have been identified in several southern provinces, including Jiangsu, Henan, Shanxi, and Anhui [32–34], indicating the existence of the rich pool of wild soybean resources in China. In our study, continuous variations in the phenotypic traits were identified in the same agricultural site, showing different degrees of variation with the average variation was of 59% and 94% in the years of 2012 and 2013, respectively, while the variation coefficients ranged from 21% (internode length) to 94% (seeds weight per plant). These values are much higher than those reported previously, based on a total of 210 wild soybean samples from Jiangsu Province, with the average variation of 26.68% and the variation coefficients ranging from 10.00% to 49.54% [32]. Our results showed that the average diversity index was 1.83 and 1.77 in the years of 2012 and 2013, respectively, with the individual of diversity index ranging from 1.73 (number of invalid pods) to 2.09 (number of effective pods). These values are much higher than those reported previously, based on 305 wild soybeans from Henan Province, with the individual diversity index ranging from 0.1187 to 1.0903 [33]. These results indicate that the wild soybeans in Heilongjiang Province contain high genetic diversity. Furthermore, results of agronomic traits showed that the high values were identified in 100-seed weight, seed weight per plant, number of effective pods, and number of branches among the wild soybeans in eastern, western, and southern sites, indicating that wild soybeans from multiple agricultural sites of Heilongjiang Province contain rich genetic diversity and suggesting their potential utilization in genetic improvement of soybean breeding. Based on these traits, a total of 27 wild soybean samples were screened and would be used as soybean breeding resources for gene identification and soybean improvement.

Results of clustering analysis based on PCA showed that all wild soybean samples were divided into two groups (Groups I and II), with each containing samples from all four collection sites, suggesting that some mixed samples were clustered together. These results may be attributed to the fact that the agronomic variations were regulated by various factors, including genotype, environment, and gene–environment interaction. Furthermore, the size of samples and the collecting locations can also lead to the differentiation of wild soybean populations. Group II was characterized mainly by having higher values of number of effective pods per plant, number of seeds per plant, and number of nodes, while Group I was characterized by 100-seed weight. These results indicate that there is no significant correlation between the wild soybean samples and collection sites, not consistent with and following the patterns of agricultural and ecological characteristics of Heilongjiang Province. The results also suggested that the geographical distance is not the most important factor causing the genetic differentiation of wild soybeans or frequent gene flow occurred in the same agricultural site of Heilongjiang Province. Therefore, these wild soybean resources should be locally conserved in situ.

5. Conclusions

In this study, a total of 242 wild soybean samples were evaluated based on 14 traits, showing a large amount of variation occurred in these evaluated traits. It was demonstrated that the wild soybeans from different sites of Heilongjiang Province contained rich genetic diversity, with significant potential utilization in genetic improvement and breakthrough for soybean breeding. Therefore, it is recommended that these wild soybean resources should be conserved in situ.

Supplementary Materials: The following are available online, at <https://www.mdpi.com/2073-4395/11/3/586/s1>. Table S1: The qualitative traits of 242 wild soybean samples in this study. Table S2: The agronomic traits of 242 wild soybean samples in this study.

Author Contributions: D.Y. contributed to conceiving and designing the experiments. W.L., M.P., Z.W., Y.B. and M.L. performed the experiments and data analysis. L.W., S.D., J.L., C.F. and G.Y. drew the figures and wrote sections of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by innovation project of Heilongjiang Academy of Agricultural Sciences, grant number 2019JCQN001 and National Natural Science Foundation of China (31771823).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not available.

Acknowledgments: This work is supported innovation project of Heilongjiang Academy of Agricultural Sciences (2019JCQN001) and National Natural Science Foundation of China (31771823).

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

References

- Smil, V. Magic beans. *Nature* **2000**, *407*, 567. [[CrossRef](#)]
- Eskandari, M.; Cober, E.R.; Rajcan, I. Genetic control of soybean seed oil: II. QTL and genes that increase oil concentration without decreasing protein or with increased seed yield. *Theor. Appl. Genet.* **2013**, *126*, 1677–1687. [[CrossRef](#)]
- Montoya, F.; García, C.; Pintos, F.; Otero, A. Effects of irrigation regime on the growth and yield of irrigated soybean in temperate humid climatic conditions. *Agric. Water Manag.* **2017**, *193*, 30–45. [[CrossRef](#)]
- He, J.; Jin, Y.; Turner, N.C.; Chen, Z.; Liu, H.-Y.; Wang, X.-L.; Siddique, K.H.; Li, F.-M. Phosphorus application increases root growth, improves daily water use during the reproductive stage, and increases grain yield in soybean subjected to water shortage. *Environ. Exp. Bot.* **2019**, *166*, 103816. [[CrossRef](#)]
- Lu, B.R. Conserving biodiversity of soybean gene pool in the biotechnology era. *Plant Species Biol.* **2004**, *19*, 115–125. [[CrossRef](#)]
- Liu, D.; Li, M.; Liu, Y.; Shi, L. Integration of the metabolome and transcriptome reveals the resistance mechanism to low nitrogen in wild soybean seedling roots. *Environ. Exp. Bot.* **2020**, *175*, 104043. [[CrossRef](#)]
- Herman, T.K.; Han, J.; Singh, R.J.; Domier, L.L.; Hartman, G.L. Evaluation of wild perennial Glycine species for resistance to soybean cyst nematode and soybean rust. *Plant Breed.* **2020**, *139*, 923–931. [[CrossRef](#)]
- Wang, Y.; Zhang, K.; Sun, W.; Wang, S.; Cao, H.; Dong, Q.; Zhang, H.; Li, K. Resistance identification and response of wild soybean (*Glycine soja*) to Soybean Mosaic Virus in eastern Hebei province. *Soybean Sci.* **2017**, *36*, 92–97.
- Sun, Y. From historical documents to archaeological data: Origin of cultivated soybean. *Soybean Sci.* **2014**, *33*, 124–128.
- Wang, K.-J.; Li, X.-H. Genetic diversity and gene flow dynamics revealed in the rare mixed populations of wild soybean (*Glycine soja*) and semi-wild type (*Glycine gracilis*) in China. *Genet. Resour. Crop Evol.* **2013**, *60*, 2303–2318. [[CrossRef](#)]
- Wang, K.-J.; Takahata, Y. A preliminary comparative evaluation of genetic diversity between Chinese and Japanese wild soybean (*Glycine soja*) germplasm pools using SSR markers. *Genet. Resour. Crop Evol.* **2007**, *54*, 157–165. [[CrossRef](#)]
- Hesler, L.S.; Taliencio, E. Resistance among selected wild soybean and associated soybean accessions against two virulent colonies of *Aphis glycines* (Hemiptera: Aphididae). *Phytoparasitica* **2020**. [[CrossRef](#)]
- Baloch, F.S.; Alsaleh, A.; de Miera, L.E.S.; Hatipoğlu, R.; Çiftçi, V.; Karaköy, T.; Yıldız, M.; Özkan, H. DNA based iPBS-retrotransposon markers for investigating the population structure of pea (*Pisum sativum*) germplasm from Turkey. *Biochem. Syst. Ecol.* **2015**, *61*, 244–252. [[CrossRef](#)]
- Govindaraj, M.; Vetriventhan, M.; Srinivasan, M. Importance of genetic diversity assessment in crop plants and its recent advances: An overview of its analytical perspectives. *Genet. Res. Int.* **2015**, *2015*, 431487. [[CrossRef](#)] [[PubMed](#)]
- Baloch, F.S.; Kurt, C.; Arioğlu, H.H.; Özkan, H. Assaying of diversity among soybean (*Glycine max* (L.) Merr.) and peanut (*Arachis hypogaea* L.) genotypes at DNA level. *Turk. J. Agric. For.* **2010**, *34*, 285–301.
- Lam, H.-M.; Xu, X.; Liu, X.; Chen, W.; Yang, G.; Wong, F.-L.; Li, M.-W.; He, W.; Qin, N.; Wang, B. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.* **2010**, *42*, 1053–1059. [[CrossRef](#)] [[PubMed](#)]
- Chen, G.; Zhang, Z.-L.; Wang, K.-J.; Zhang, J.-J.; Li, Z.-B.; Ceng, X.-Q.; Gu, L.; Li, X.-H. A Preliminary Analysis on Phenotypic Diversity of Regional Populations of Wild Soybean in Chongqing. *J. Plant Genet. Resour.* **2010**, *11*, 573–577.
- Li, J.; Jing, X.; Yan, X. Study on phenotypic diversity of wild soybean in Huang-Huai populations. *Soybean Sci.* **2015**, *34*, 741–751.
- Li, X.; Wang, K. Morphological Evidence of the Origin of Semi-wild Soybean and the Rare Traits in Wild Soybean. *J. Plant Genet. Resour.* **2020**, *21*, 1357–1371.
- Sun, X.; Zhao, H.; Wang, Y.; Ren, H.; Shang, Y.; Bu, H.; Du, W.; Dong, Y. Ecological character analysis of wild soybean germplasm resources in Heilongjiang Province. *Heilongjiang Agric. Sci.* **2019**, *11*, 5–8.
- Lin, H. Identification and screening of excellent wild soybean resources in Heilongjiang Province. *Zuowu Pinzhong Ziyuan* **1997**, *2*, 41–42.
- Lai, Y.; Lin, H.; Fang, W.; Yao, Z.; Qi, N.; Wang, Q.; Yang, X.; Li, H. Research that the excellent resource of wild soybeans screen appraise and utilization in Heilongjiang. *Zhongguo Nong Xue Tong Bao/Chin. Agric. Sci. Bull.* **2005**, *21*, 379–382.
- Lin, H.; Qi, N.; Li, X. New progress on wild soybean survey in Heilongjiang province. *Chin. J. Oil Crop Sci.* **2006**, *28*, 427–430.

24. Wang, L.; Lai, Y.; Li, W.; Bi, Y.; Liu, M.; Liu, M.; Di, S. Status, problems and countermeasures of wild soybean resources in cold region of Heilongjiang Province. *Heilongjiang Agric. Sci.* **2016**, *3*, 138–142.
25. Sun, S.; Pei, Y.; Wang, C.; Liu, D. Main meteorological disasters and the impact to Heilongjiang Province in 2013. *Heilongjiang Agric. Sci.* **2014**, *12*, 153–155.
26. Yan, X.F. *Genetic Diversity and Differentiation of the Wild Soybean (Glycine soja Sieb. et Zucc) in China*; Shenyang Agricultural University: Shenyang, China, 2014.
27. Nawaz, M.A.; Yang, S.H.; Rehman, H.M.; Baloch, F.S.; Lee, J.D.; Park, J.H.; Chung, G. Genetic diversity and population structure of Korean wild soybean (*Glycine soja* Sieb. and Zucc.) inferred from microsatellite markers. *Biochem. Syst. Ecol.* **2017**, *71*, 87–96. [[CrossRef](#)]
28. Guo, R. Soybean agricultural production technology in northern Heilongjiang Province. *Xiandai Nongye Keji* **2009**, *3*, 194–204.
29. Sun, L.; Zhao, H.; Zhao, F.; Dong, Y. Analysis of genetic diversity of *Glycine soja* in Northeast China. *Soybean Sci.* **2015**, *34*, 355–360.
30. Li, W.; Xiao, J.; Bi, Y.; Liu, M.; Liu, M.; Lin, H.; Zhang, B.; Lai, Y. Diversity of wild soybean resources based on agronomic and quality traits in Heilongjiang province. *Soybean Sci.* **2015**, *34*, 9–14.
31. Barrett, S.C.; Eckert, C.G. 14 Variation and Evolution of Mating Systems in Seed Plants. *Biol. Approaches Evol. Trends Plants* **2012**, *14*, 229–254.
32. Wang, Y. *The Population Trait and Genetic Diversity of Wild Soybeans in Jiangsu Province*; Nanjing Agricultural University: Nanjing, China, 2011.
33. Wang, G. *The Genetic Diversity of Annual Wild Soybeans in Henan Province*; Northwest A&F University: Yangling, China, 2006.
34. Guo, W.; Zheng, H.; Baoque, Z. Genetic diversity analysis of Shanxi's wild soybean (*Glycine soja*). *Sci. Agric. Sin.* **2008**, *41*, 2182–2190.