



Phenotypic diversity analysis of *Polygala tenuifolia* Willd germplasm resources in China

Hongling Tian · Changjuan Wu · Lijun Zhang · Shuhong Guo ·
Yaoqin Wang · Shuaishuai Pei · Xianqiang Zuo · Qiubao Wang ·
Qiang Zhang

Received: 7 March 2023 / Accepted: 4 December 2023 / Published online: 11 January 2024
© The Author(s) 2024

Abstract *Polygala tenuifolia* Willd is a perennial herb in the family Polygalaceae and is used as an expectorant and stimulant in traditional Chinese medicine. There is a large demand for *P. tenuifolia*, while the good cultivars of *P. tenuifolia* are not enough. In this study, we collected 157 germplasm resources of *P. tenuifolia* from different regions in China, and 12 phenotypic traits of these samples were analyzed using correlation analysis, principal component analysis (PCA), and cluster analysis methods. The correlation analysis showed the associations between different traits, and the first four principal components contributed 62.63% of the total variation. They contained most of the information of five above-ground traits and eight root traits. Phylogenetic clustering divided materials divided into four taxa, with characteristics of high-yielding specific material, long-rooted specific material, multi-branching

specific material, and high plant height specific material, respectively. Our study provides the material and theoretical basis for the utilization of germplasm resources, parental selection, and varietal improvement of *P. tenuifolia*.

Keywords *Polygala tenuifolia* · Phenotypic Diversity · Willd Germplasm · Parental selection · Varietal improvement

Introduction

Polygala tenuifolia Willd, also known as ‘Yuanzhi’ in Chinese, is a perennial herb in the family Polygalaceae and is widely distributed in China (Jin et al. 2014). In traditional Chinese medicine, *P. tenuifolia* is generally used as an expectorant and stimulant and is frequently applied in the treatment of forgetfulness, insomnia, *P. tenuifolia* and neurasthenia (Zhang et al. 2016). Previous studies on *P. tenuifolia* also showed that it has protective effects on the central nervous system and confirmed that the aerial parts of *P. tenuifolia* exhibit significant biological activities in ameliorating learning ability and memory impairments (Deng et al. 2020; Nguyen et al. 2020; Vinh et al. 2020; Zhao et al. 2020).

The first record of *P. tenuifolia* appeared in the Compendium of Materia Medica as a treatment for amnesia in China (Cheng 2021). It is an important medicinal herb and is one of the 42 tertiary wild

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10722-023-01826-y>.

H. Tian (✉)
Center for Agricultural Genetic Resources Research,
Shanxi Agricultural University, Jinzhong, China
e-mail: thl2003@163.com

C. Wu · L. Zhang · S. Guo · Y. Wang · S. Pei · X. Zuo ·
Q. Wang · Q. Zhang
The Industrial Crop Institute, Shanxi Agricultural
University/Shanxi Academy of Agriculture Sciences,
Jinzhong, China

species under national priority protection in China. Shanxi province is the main production area of *P. tenuifolia* in China, accounting for 70% of the national production and including 90% of the national germplasm resources. The domestic history of the *P. tenuifolia* is short, with little domestication, and breeding is also lacking. With the accelerating pace of life, the number of people with insomnia and forgetfulness is gradually increasing, the demand for *P. tenuifolia* has been growing. However, the lack of good *P. tenuifolia* cultivars has severely restricted the development of its industry.

Germplasm resources are the basis of plant breeding (Liu et al. 2020; Thudi et al. 2021). The investigation of agronomic and physiological traits of *P. tenuifolia* allows a comprehensive assessment of its various characteristics, quality and yield. At present, morphological markers are usually the simplest and most effective methods for the identification and evaluation of genetic resources (Chesnokov et al. 2020). The principal component analysis (PCA), correlation analysis and cluster analysis are the most widely used methods in morphological and genetic diversity research of genetic resources. They have been broadly used in soybean (Wang and Komatsu 2018), rice (Bhargavi et al. 2021), millet (Sharma et al. 2018), tomato (Zhou et al. 2015), among other plant cultivars, while few investigations have their focus on *P. tenuifolia*. Hence, we collected 157 germplasm resources of *P. tenuifolia* from different regions in China and analyzed their genetic diversity, estimated trait correlation, and did principal components and clustering based on 12 phenotypic traits. Our study provides the material and theoretical bases for the utilization of germplasm resources, parental selection and varietal improvement of *P. tenuifolia* in China.

Materials and methods

Experimental materials

A total of 157 germplasm resources of *P. tenuifolia* from Shanxi, Hebei, Inner Mongolia, Henan, and Shaanxi were collected in this study (Table 1). Samples were provided by the Institute of Cash Crops, Shanxi Agricultural University, and purified over three generations (Fig. 1).

Overview of experimental base

The experiment was conducted at the experimental base of the Cash Crop Research Institute, Shanxi Agricultural University in Fenyang, China. Fenyang is located in Loess Plateau, with an average altitude of 1414 m. The climate in Fenyang is temperate monsoon, with an annual average temperature of 9.7 °C, an average ground temperature of 12.6 °C, and annual average precipitation of 467.2 mm. The soil type of the experimental base belongs to a brown soil.

Experimental design

Polygala tenuifolia seeds were planted on 25th June 2016. The planting density was 3.5 kg/667m², with an area of 3 m×2 m, and three replications were set for each germplasm resource. The manual seed drill was a depth of 1.5 cm and a row spacing of 25 cm. Chicken manure 350 kg/667 m² was added before planting, and no more fertilizer was used during the reproductive period. Weeding was performed for 4–5 times a year, and treatments for all regions were consistent. Seeds were harvested from 1–7 October 2019.

Measurement indicators

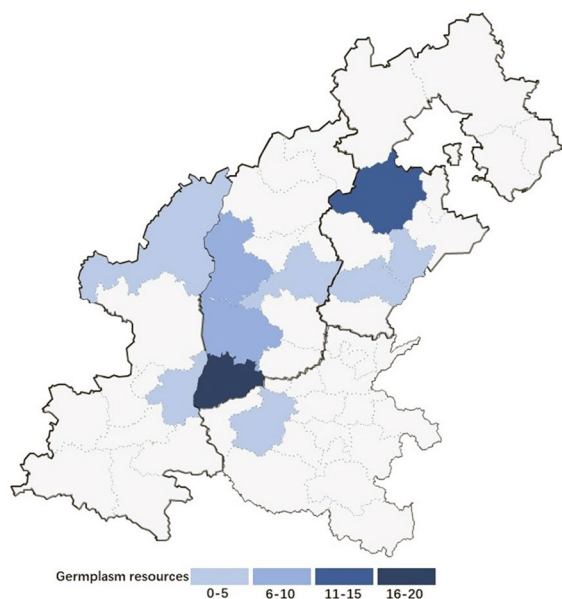
Plant traits, including the plant height, leaf shape, leaf color, and flower color, were recorded during the flowering stage in 2019. After harvesting the seeds in late July, three plants were randomly selected from each area to measure plant height. Root length, root thickness, and fresh weight were measured in October. When the moisture of roots dried to 20–24%, the bast was extracted manually, and the bast weight and wood core thickness were measured. After baking to a constant weight at 105 °C, the dry weight of the bast and wood core were measured. Qualitative traits were assigned different values (Fig. 2). For the leaf shape, thin and long was 1 (width ranging from 1.0 to 1.5 mm, length exceeding 15 mm), thin and short was 2 (width ranging from 1.0 to 1.5 mm, length ranging from 10 to 15 mm), thick and long was 3 (width exceeding 1.5 mm, length ranging from 10 to 20 mm), and thick and short was 4 (width exceeding 1.5 mm, length exceeding 20 mm). For the leaf color, green was 1, gray-green was 2, and tender green was 3. For the flower color, purple was 1, blue-purple was 2, and light purple was 3.

Table 1 The statistic of the *Polygala tenuifolia* germplasm resource collected in this study

Number	Name	Source	Number	Name	Source	Number	Name	Source
1	Z1	Shanxi	54	Jishan14-06	Shanxi	107	Yu-Lingbao-01	Henan
2	Z2	Shanxi	55	Jishan14-07	Shanxi	108	Yu-Lingbao-02	Henan
3	Z3	Shanxi	56	Jishan14-12	Shanxi	109	Yu-Lingbao-03	Henan
4	Z4	Shanxi	57	Ruicheng10-05	Shanxi	110	Yu-Lingbao-04	Henan
5	Z5	Henan	58	Ruicheng10-07	Shanxi	111	Ji-Anguo-01	Hebei
6	Z6	Henan	59	Hejin10-08	Shanxi	112	Ji-Anguo-02	Hebei
7	Z7	Henan	60	Hejin10-10	Shanxi	113	Ji-Anguo-03	Hebei
8	Z8	Shaanxi	61	Hejin10-04	Shanxi	114	Ji-Anguo-04	Hebei
9	Z9	Shaanxi	62	Hejin14-03	Shanxi	115	Ji-Anguo-05	Hebei
10	Z10	Shaanxi	63	Hejin14-04	Shanxi	116	Ji-Anguo-06	Hebei
11	Z11	Shaanxi	64	Hejin14-08	Shanxi	117	Ji-Anguo-07	Hebei
12	Z12	Shaanxi	65	Xiaxian08-01	Shanxi	118	Ji-Anguo-08	Hebei
13	Z13	Shaanxi	66	Xiaxian08-02	Shanxi	119	Ji-Anguo-09	Hebei
14	Z14	Shaanxi	67	Xiaxian08-03	Shanxi	120	Ji-Anguo-10	Hebei
15	Z15	Shaanxi	68	Xiaxian08-04	Shanxi	121	Ji-Anguo-11	Hebei
16	Yongji08-01	Shanxi	69	Xiaxian10-05	Shanxi	122	Ji-Dingzhou-01	Hebei
17	Yongji08-02	Shanxi	70	Xiaxian10-07	Shanxi	123	Ji-Dingzhou-02	Hebei
18	Xinjiang08-01	Shanxi	71	Xiaxian10-08	Shanxi	124	Ji-Dingzhou-03	Hebei
19	Xinjiang08-02	Shanxi	72	Xiaxian14-02	Shanxi	125	Ji-Dingzhou-04	Hebei
20	Xinjiang08-03	Shanxi	73	Xiaxian14-03	Shanxi	126	Ji-Dingzhou-05	Hebei
21	Xinjiang08-04	Shanxi	74	Yuanqu09-01	Shanxi	127	Ji-Dingzhou-06	Hebei
22	Xinjiang08-05	Shanxi	75	Yuanqu09-02	Shanxi	128	Ji-Dingzhou-07	Hebei
23	Xinjiang08-06	Shanxi	76	Yuanqu09-03	Shanxi	129	Ji-Dingzhou-08	Hebei
24	Jiangxian08-01	Shanxi	77	Pingshun10-01	Shanxi	130	Ji-Anping-05	Hebei
25	Jiangxian08-02	Shanxi	78	Shan-Heyang-01	Shaanxi	131	Ji-Anping-06	Hebei
26	Jiangxian08-03	Shanxi	79	Shan-Heyang-02	Shaanxi	132	Ji-Anping-07	Hebei
27	Jiangxian08-04	Shanxi	80	Shan-Heyang-03	Shaanxi	133	Ji-Anping-08	Hebei
28	Huguan08-01	Shanxi	81	Shan-Heyang-04	Shaanxi	134	Ji-Anping-09	Hebei
29	Huguan08-02	Shanxi	82	Shan-Heyang-05	Shaanxi	135	Ji-Anping-10	Hebei
30	Wanrong08-01	Shanxi	83	Shan-Heyang-06	Shaanxi	136	Ji-Raoyang-01	Hebei
31	Wanrong08-02	Shanxi	84	Shan-Chengcheng-01	Shaanxi	137	Ji-Raoyang-02	Hebei
32	Wanrong08-03	Shanxi	85	Shan-Chengcheng-02	Shaanxi	138	Ji-Raoyang-03	Hebei
33	Wanrong08-04	Shanxi	86	Shan-Chengcheng-03	Shaanxi	139	Ji-Raoyang-04	Hebei
34	Wanrong08-05	Shanxi	87	Shan-Chengcheng-04	Shaanxi	140	Ji-Zhangbei-01	Hebei
35	Wanrong08-06	Shanxi	88	Shan-Chengcheng-05	Shaanxi	141	Ji-Zhangbei-02	Hebei
36	Wanrong08-07	Shanxi	89	Shan-Chengcheng-06	Shaanxi	142	Ji-Zhangbei-03	Hebei
37	Wenxi09-01	Shanxi	90	Shan-Chengcheng-07	Shaanxi	143	Ji-Zhangbei-04	Hebei
38	Wenxi09-02	Shanxi	91	Shan-Pucheng-01	Shaanxi	144	Ji-Zhangbei-05	Hebei
39	Wenxi09-03	Shanxi	92	Shan-Pucheng-02	Shaanxi	145	Ji-Zhangbei-06	Hebei
40	Wenxi09-04	Shanxi	93	Shan-Pucheng-03	Shaanxi	146	Ji-Zhangbei-07	Hebei
41	Wenxi09-05	Shanxi	94	Shan-Pucheng-04	Shaanxi	147	Ji-Zhangbei-08	Hebei
42	Wenxi09-06	Shanxi	95	Shan-Pucheng-05	Shaanxi	148	Ji-Zhangbei-09	Hebei
43	Houma09-01	Shanxi	96	Shan-Pucheng-06	Shaanxi	149	Ji-Zhangjiakou-01	Hebei
44	Houma09-02	Shanxi	97	Shan-Dali-01	Shaanxi	150	Ji-Zhangjiakou-02	Hebei
45	Houma10-01	Shanxi	98	Shan-Dali-02	Shaanxi	151	Ji-Zhangjiakou-03	Hebei
46	Houma10-02	Shanxi	99	Shan-Dali-03	Shaanxi	152	Ji-Zhangjiakou-04	Hebei

Table 1 (continued)

Number	Name	Source	Number	Name	Source	Number	Name	Source
47	Houma10-03	Shanxi	100	Shan-Dali-04	Shaanxi	153	Ji-Zhangjiakou-05	Hebei
48	Houma14-01	Shanxi	101	Yu-Mengjin-01	Henan	154	Ji-Zhangjiakou-06	Hebei
49	Houma14-02	Shanxi	102	Yu-Mengjin-02	Henan	155	Ji-Zhangjiakou-07	Hebei
50	Houma14-03	Shanxi	103	Yu-Mengjin-03	Henan	156	NO.6	Validated varieties
51	Jishan10-01	Shanxi	104	Yu-Mengjin-04	Henan	157	NO.2	Validated varieties
52	Jishan10-02	Shanxi	105	Yu-Mengjin-05	Henan			
53	Jishan10-03	Shanxi	106	Yu-Mengjin-06	Henan			

**Fig. 1** The geographical distribution of the collected *Polygala tenuifolia* germplasm resources

Data analysis

Statistical analyses of the raw data were conducted using Microsoft Excel 2019. Key metrics calculated included the overall mean ($\bar{X}$), standard deviation (SD), and coefficient of variation (CV). Additionally, the Shannon–Wiener index (H') was employed to assess the genetic diversity of the test materials, with P_i representing the proportion of the i^{th} variant within the trait. The Shannon–Wiener index (H') was calculated using the formula: $H' = -\sum (P_i \times \ln(P_i))$, where P_i is the frequency of the i^{th} variant relative to the total number of variants for a given trait.

For correlation and cluster analyses, the data were processed using Origin 2022. Principal component analysis was performed with IBM SPSS Statistics 21 to further interpret the data and investigate underlying relationships.

Results

Variation and phenotypic diversities of different *P. tenuifolia* germplasm resources

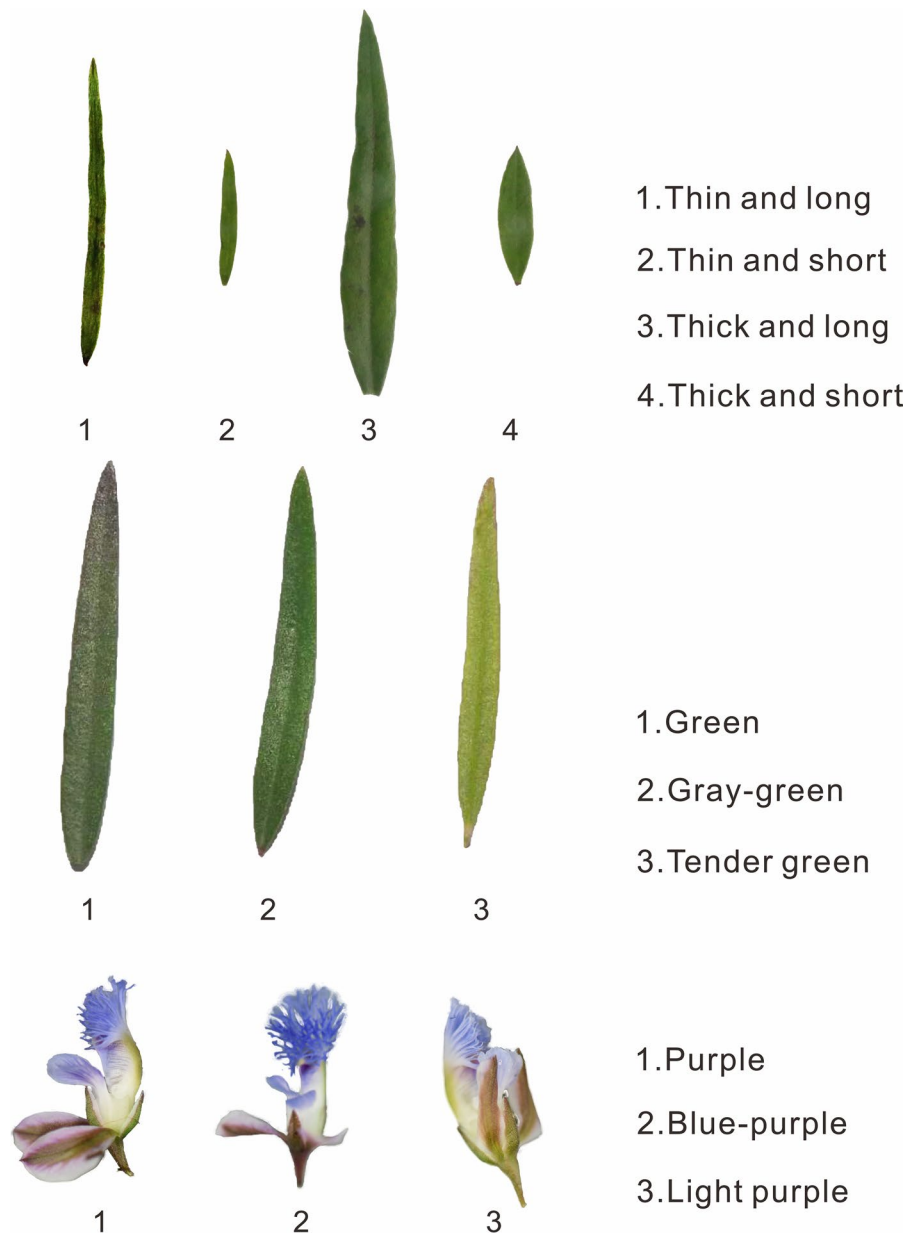
As shown in Table 2, the above-ground traits of *P. tenuifolia* varied widely, with a coefficient of variation (CV) range of 13.5–44.35%. The variation in branching was the greatest at 44.35% and the variation in plant height was the least at 13.5%. The index of genetic diversity ranged from 0.66 to 2.07, with the smallest variation was noticed for leaf color at 0.66 and the largest variation for plant height at 2.07.

According to the CV in Table 3, the root traits of *P. tenuifolia* germplasm resources varied widely, with a range of variation from 15.98 to 57.29%, of which the branching variation was the largest at 57.29% and the fresh root thickness variation was the smallest at 15.98%. The index of genetic diversity ranged from 0.95 to 2.05, with the smallest difference of 0.95 in fresh root weight and the largest difference of 2.05 in wood core thickness (Table 3).

Correlation analysis of phenotypic traits of different *P. tenuifolia* germplasm resources

The correlations were strong among root traits while were weak among above-ground traits (Table 4). The number of branches was highly significantly and positively correlated with the fresh

Fig. 2 Comparisons of different features of the collected *Polygala tenuifolia* germplasm resources



root weight, fresh root thickness, dry weight of the wood core, dry weight of root bark, and was significantly and positively correlated with above-ground branches. Fresh root weight was highly significantly correlated with root length, fresh root thickness, dry weight of the wood core, and dry weight of root bark, and was significantly correlated with plant height. Root length was highly significantly and positively correlated with the dry weight of the

wood core and the dry weight of root bark. Fresh root thickness was highly significantly and positively correlated with the dry weight of the wood core and the dry weight of root bark. The dry weight of the wood core was significantly and positively correlated with the dry weight of root bark. The dry weight of root bark was significantly correlated with plant height.

Table 2 A Variation and genetic diversity analysis of the above-ground traits of *Polygala tenuifolia*

Characteristic	Plant height	Branches	Leaf shape	Leaf color	Flower color
Min. value	20.67	1.67	1	1	1
Max. value	43.33	25.33	4	3	3
Average value	31.63	8.68	1.83	1.29	1.71
Standard deviation	4.27	3.85	0.72	0.5	0.68
Coefficient of variation (%)	13.50%	44.35%	39.30%	38.76%	39.77%
Genetic diversity index	2.07	1.96	1.04	0.66	0.99

For the leaf shape, slender was 1, thin and short was 2, thick and long was 3, and thick and short was 4. For the leaf color, green was 1, gray-green was 2, and tender green was 3. For the flower color, purple was 1, blue-purple was 2, and light purple was 3

Table 3 Variation and genetic diversity of root traits of *Polygala tenuifolia*

Characteristic	Branches	Fresh root weight	Root length	Fresh root thickness	Wood core thickness	Dry weight of wood core	Dry weight of root bark
Min. value	0.00	3.88	27.00	3.91	1.45	0.28	1.03
Max. value	7.60	25.62	62.00	9.38	3.95	3.50	8.41
Average value	2.88	11.37	39.79	6.44	2.56	0.95	3.64
Standard deviation	1.65	3.97	7.36	1.03	0.49	0.45	1.43
Coefficient of variation (%)	57.29	34.93	18.49	15.98	19.18	46.97	39.33
Genetic diversity index	2.02	0.95	1.98	2.04	2.05	1.82	1.95

PCA on phenotypic traits of different *P. tenuifolia* germplasm resources

Principal component analysis was performed on 13 phenotypic traits of 157 *P. tenuifolia* germplasm resources using SPSS24.0 software. The results showed that the cumulative contribution of the first four principal components (PC1, PC2, PC3 and PC4) amounted to 62.63% and contained most of the information of five above-ground traits and eight root traits.

The eigenvalue of PC1 was 4.5414 with a contribution of 34.9341%. Among its eigenvalue vectors, the traits with high absolute values were fresh root weight (0.4546), fresh root thickness (0.3945), and dry weight of root bark (0.4356), which are closely related to yield. Therefore, PC1 is the yield factor. The eigenvalue of PC2 was 1.4280 with a contribution of 10.9850%. In its eigenvector, the traits with high absolute values were the number of branch roots (0.3355), branches (0.5167), leaf shape (−0.3576), and leaf color (0.4800). PC2 is branch and leaf traits. The eigenvalue of PC4 was 1.1209 with a contribution rate of 8.6225, and the trait with a higher

absolute value was plant height (−0.5309), so PC3 is plant height class. The eigenvalue of PC4 was 1.0518 with a contribution rate of 8.0910%, and the trait with a higher absolute value was wood core thickness (0.7418), and PC4 is wood core class (Table 5).

Clustering analysis of phenotypic traits of different *P. tenuifolia* germplasm resources

Clustering analysis was performed on the 12 phenotypic traits of 157 *P. tenuifolia* germplasm resources by a systematic clustering method. When the genetic distance was ten, the participating materials were divided into four clusters (Supplementary Figure 1). Meanwhile, the phenotypic traits of the germplasm resources in four clusters were statistically analyzed (Table 6).

Taxon I included 54 germplasm resources, of which 28 were from Shanxi, 11 from Shaanxi, 4 from Henan, and 11 from Hebei. The main characteristics of taxon I are heavy (13.46 g) and thick (6.78 mm) fresh roots, more branch roots (3.14), large yields, short and slender leaves, and purple flowers. Taxon I belongs to the high-yielding specific material. There

Table 4 Correlations among traits of different *Polygala tenuifolia* germplasm resources

Correlation coefficient	Number of roots	Fresh root weight	Root length	Fresh root thickness	Wood core thickness	Dry weight of wood core	Dry weight of root bark	Plant height	Branches	Leaf shape	Leaf color	Flower color
Number of roots	1											
Fresh root weight	0.60**	1										
Root length	0.11	0.40**	1									
Fresh root thickness	0.56**	0.85**	0.13	1								
Wood core thickness	−0.06	−0.02	0	0.02	1							
Dry weight of wood core	0.45**	0.77**	0.30**	0.62**	−0.02	1						
Dry weight of root bark	0.41**	0.90**	0.46**	0.70**	−0.03	0.64**	1					
Plant height	0.03	0.16*	0.07	0.15	0.05	0.16	0.18*	1				
Branches	0.18*	0.08	0.06	0.12	−0.04	0.06	0.02	0.01	1			
Leaf shape	0	−0.02	0.04	−0.06	0.14	−0.01	−0.01	−0.02	−0.11	1		
Leaf color	0.14	0.03	0.09	0.01	−0.05	−0.05	0.02	0.06	0.19*	−0.12	1	
Flower color	0.01	−0.1	−0.1	−0.04	−0.09	−0.12	−0.13	−0.07	0.13	0.01	0.04	1

* $p < 0.05$, ** $p < 0.01$

Table 5 Principal components (PC) and their contributions by each trait

	PC1	PC2	PC3	PC 4
Number of roots	0.2677	0.3355	0.2285	0.1712
Fresh root weight	0.4546	0.0056	0.0636	−0.0017
Root length	0.2051	−0.1118	−0.3378	−0.1731
Fresh root thickness	0.3945	0.0871	0.1902	0.1076
Wood core thickness	−0.0075	−0.3086	0.0146	0.7418
Dry weight of wood core	0.3725	−0.0442	0.0151	−0.0738
Dry weight of root bark	0.4356	−0.0781	−0.0365	−0.0947
Plant height	0.4225	−0.1260	−0.0332	−0.0955
Branches	0.0948	−0.0355	−0.5309	0.3093
Leaf shape	0.0462	0.5167	0.0650	0.2806
Leaf color	0.0278	−0.3576	0.4461	0.3196
Flower color	0.0223	0.4800	−0.3593	0.2685
Number of roots	−0.0677	0.3486	0.4200	−0.0819
Eigenvalue	4.5414	1.4280	1.1209	1.0518
Contribution rate	34.9341	10.9850	8.6225	8.0910
Cumulative contribution rate (%)	34.9341	45.9191	54.5416	62.6326

are 33 germplasm resources of taxon II, including 11 from Shanxi, 8 from Shaanxi, 6 from Henan, and 8 from Hebei. The main characteristics of taxon II are long roots (45.56 cm), green leaves, purple-blue-purple flowers, and high yield potential. Taxon II belongs to the long-rooted specific material. Taxon III includes 46 germplasm resources, including 21 from Shanxi, 6 from Shaanxi, 2 from Henan, and 17 from Hebei. The main characteristics of taxon III are more branches in the above-ground part, high seed yield potential, slender and short leaves, and purple and blue-purple flowers. Taxon III belongs to the multi-branching specific material. There are 24 germplasm resources in Taxon IV, including 6 in Shanxi, 6 in Shaanxi, 1 in Henan, 9 in Hebei, and 2 reviewed varieties. The main characteristics of Taxon IV are high plant height (34.23 cm) and blue-purple flowers. Taxon IV belongs to the high plant height specific material.

Correlation analysis between active ingredients and agronomic traits in *P. tenuifolia*

According to the taxa of *P. tenuifolia* germplasm resources, 30% of germplasm resources of each taxon were randomly selected for active ingredient determination. Correlation analysis was performed on the active ingredients and phenotypic diversity of *P. tenuifolia* germplasm resources. The correlation between active ingredients and phenotypic traits of

P. tenuifolia was small (Table 7). The root branch of *P. tenuifolia* was significantly correlated with xanthones, and the plant height was significantly negatively correlated with 3,6'-disinapoylsucrose. There were significant positive correlations between 3,6'-disinapoylsucrose and the contents of both saponin and xanthones. All root traits were highly significant correlated with each other. The correlations among above-ground traits and among root traits were small. Leaf shape was significantly and positively correlated with fresh root weight, fresh root thickness, and wood core thickness. Branch was significantly and positively correlated with the number of branch roots.

Discussion

Germplasm resources are the cornerstone of crop improvement, and analysis of the genetic background of germplasm resources is the key to breakthroughs in plant breeding (Yang et al. 2020). In the present study, we evaluated the phenotypic characteristics of 157 *P. tenuifolia* germplasm resources collected from different regions across China, which aimed at enhancing and harnessing desirable traits in future breeding of *P. tenuifolia*. Our results revealed substantial variability in both above-ground and root traits, as indicated by a CV exceeding 13.5%. This level of variation is considerable, aligning with previous studies that interpret a CV greater than 10%

Table 6 Mean values and characteristics of phenotypic traits of the four taxa

Taxon	Number of roots	Fresh root weight	Root length	Fresh root thickness	Wood core thickness	Dry weight of wood core	Dry weight of root bark	Plant height	Branches	Leaf shape	Leaf color	Flower color
I	3.14	13.46	42.03	6.78	2.74	1.13	4.34	31.10	7.20	Mostly short and thin	Green and gray-green mostly	Mostly purple
II	2.51	10.34	45.56	6.03	2.48	0.87	3.53	28.29	8.63	Mostly slender and short	Mostly green	More purple and violet
III	2.79	9.74	38.02	6.15	2.37	0.76	3.09	31.86	14.81	Mostly short and thin	Green and gray-green mostly	More purple and violet
IV	2.86	10.70	33.98	6.50	2.52	0.91	3.29	34.23	7.45	Mostly short and thin	Mostly green	Mostly violet

as indicative of significant inter-sample variability (Toebe et al. 2018). The genetic assessments on *P. tenuifolia* also showed a high genetic diversity in its populations (Peng et al. 2018). The germplasm resources of *P. tenuifolia* involved in our study exhibit a relatively rich genetic variation, which is conducive to the screening of high-quality germplasm resources and provides abundant material for further breeding.

The comprehensive correlation analysis was conducted to investigate the relationships between active ingredients, root traits, and above-ground traits in *P. tenuifolia* germplasm. We observed robust correlations within distinct root traits and active ingredients, whereas correlations among various above-ground traits were relatively weaker. Notably, a significant positive correlation emerged between the number of branched roots and xanthonenes content, while plant height was inversely related to the levels of 3,6'-disinapoylsucrose. Additionally, 3,6'-disinapoylsucrose content displayed a positive correlation with both xanthonenes and saponin levels. Morphological traits such as leaf shape exhibited a significantly positive correlation with fresh root weight, thickness, and wood core thickness, whereas branching was significantly positive in correlation with the number of branched roots. Plant height, number of branched roots, and leaf shape were significantly correlated with 3,6'-disinapoylsucrose, xanthonenes, and overall yield, respectively.

Contrasting with the findings of previous research (Zhang et al. 2020; Zhang et al. 2017), our results suggested that the selection of *P. tenuifolia* germplasm for this study could account for these differences. Traditional selection criteria for *P. tenuifolia*, which favor attributes such as thicker tubers and fleshier, non-woody roots, were not validated by our analysis. The indicator compounds, xanthonenes, 3,6'-disinapoylsucrose, and saponin, did not exhibit a significant correlation with the traditionally valued root traits. This aligns with the outcomes reported by Wang et al. (2016), reinforcing the notion that morphological traits are crucial determinants of both yield and the medicinal quality of *P. tenuifolia*. Consequently, breeding strategies for *P. tenuifolia* should not rely solely on conventional wisdom but rather should incorporate a systematic evaluation of morphological traits in relation to key active ingredients. Our insights provide valuable guidance for the selection and cultivation of superior *P. tenuifolia* cultivars.

Table 7 Correlations between active ingredients and agronomic traits in *Polygala tenuifolia*

Correlation coefficient	Xanthones	Saponin	3,6'-Disinapoylsucrose	Number of roots	Fresh root weight	Root length	Fresh root thickness	Wood core thickness	Dry weight of wood core	Dry weight of root bark	Plant height	Branches	Leaf shape	Leaf color	Flower color
Xanthones	1														
Saponin	0.19	1													
3,6'-Disinapoylsucrose	0.28*	0.35*	1												
Number of roots	0.31*	−0.06	−0.1	1											
Fresh root weight	0.26	0.19	0.18	0.62**	1										
Root length	0.04	0.07	0.01	0.17	0.34*	1									
Fresh root thickness	0.18	0.14	0.13	0.53**	0.88**	0.1	1								
Wood core thickness	0.27	0.14	0.16	0.54**	0.84**	0.37**	0.82**	1							
Dry weight of wood core	0.25	0.13	0.02	0.47**	0.67**	0.39**	0.61**	0.69**	1						
Dry weight of root bark	0.09	0.23	0.25	0.46**	0.81**	0.33*	0.75**	0.74**	0.60**	1					
Plant height	−0.12	−0.01	−0.31*	0.27	0.09	−0.09	0.23	0.27	0.21	−0.02	1				
Branches	0.14	−0.09	−0.13	0.35*	0.15	−0.02	0.22	0.15	0.05	0.2	0.2	1			

Table 7 (continued)

Correlation coefficient	Xanthones	Saponin	3,6'-Disinapoylsucrose	Number of roots	Fresh root weight	Root length	Fresh root thickness	Wood core thickness	Dry weight of wood core	Dry weight of root bark	Plant height	Branches	Leaf shape	Leaf color	Flower color
Leaf shape	0.05	0.13	0.19	0.13	0.31*	-0.03	0.29*	0.28*	0.12	0.24	0.02	-0.06	1		
Leaf color	0	0.03	-0.18	0.21	-0.01	-0.04	-0.03	-0.03	-0.16	-0.11	0.04	0.28*	-0.1	1	
Flower color	-0.1	-0.23	-0.17	0.04	0.04	0.16	0	0.08	0.1	0.15	-0.08	0.21	-0.09	0.11	1

* $p < 0.05$, ** $p < 0.01$

Principal component analysis can simplify the classification of morphological traits by grouping morphological traits into several integrated factors to reflect the original morphological trait variables without losing the original information (Jang 2021). In this study, the PCA results indicated that the first four principal components accounted for 62.63% of the total variation, containing most of the data from five above-ground traits and eight root traits. Four principal components were correlated with each other, indicating that the PCA was instrumental in guiding the selection of both the germplasm materials and the trait indicators for further analysis.

Based on the observed values of different variables of the materials, the degree of similarity between different materials was calculated, and the cluster analysis can define different materials with similarity degrees until all materials are included clustered into a phylogenetic tree. In our analysis of 157 germplasm resources based on 12 phenotypic traits; the germplasm was divided into four distinct taxa at a genetic distance of 10. Each taxon corresponded to a specific material trait: taxon I to high-yield, taxon II to long-rooted, taxon III to multi-branching, and taxon IV to tall plant stature. Additionally, we observed that the roots of *P. tenuifolia* sourced from Shanxi (Linfen City and Yuncheng City) exhibited greater length and weight compared to those from other regions under study.

The identification of morphological traits and genetic diversity of crop germplasm resources is usually performed by using PCA, correlation analysis, and cluster analysis, but most morphological traits are affected by both genetic and environmental factors. Therefore, the identification and genetic research of germplasm resources only based on a single morphological trait have limitations on the results. In the future, cytology, biochemistry, and molecular markers should be combined to obtain more accurately genetic diversity of *P. tenuifolia* and provide a theoretical basis for efficient utilization of *P. tenuifolia* germplasm resources.

Acknowledgements We thank the researchers at our laboratories for their dedication and hard work.

Authors' contributions H.T. had the idea for and designed the study. C.W., L.Z., S.G., S.P., X.Z., Q.W., Q.Z. did the statistical analysis. All authors contributed to acquisition, analysis, or interpretation of data. H.T. wrote the draft report. All authors revised the report and approved the final version before submission.

Funding This project was supported by The Chinese Materia Medica, China Agriculture Research System (CARS-21), Shanxi Agricultural University Breeding Project (YZGC056), and the earmarked fund for Modern Agro-industry Technology Research System (2021-11).

Availability of data and material All data and materials mentioned in this article are available.

Declarations

Conflict of interest The authors declare that there is no conflict of commercial interest related to this paper.

Consent for publication The authors declare that there are Consent for publication.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bhargavi M, Maneesha K, Withanawasam DM, Aratikatla KR, Himabindu S, Prashanth M, Shanthi P, Kommana ML, Reddy DM, Reddy BR, Vemireddy LR (2021) A novel barcode system for rapid identification of rice (*Oryza sativa* L.) varieties using agro-morphological descriptors and molecular markers. *Mol Biol Rep* 48(3):2209–2221
- Cheng YG et al (2021) Chemical constituents from the aerial part of *Polygala tenuifolia*. *Nat Prod Res* 1–6
- Chesnokov YV, Kosolapov VM, Savchenko IV (2020) Morphological genetic markers in plants. *Russ J Genet* 56:1406–1415
- Deng X, Zhao S, Liu X, Han L, Wang R, Hao H, Jiao Y, Han S, Bai C (2020) *Polygala tenuifolia*: a source for anti-Alzheimer's disease drugs. *Pharm Biol* 58(1):410–416
- Jang JH (2021) Principal component analysis of hybrid functional and vector data. *Stat Med* 40(24):5152–5173
- Jin ML, Lee DY, Um Y, Lee JH, Park CG, Jetter R, Kim OT (2014) Isolation and characterization of an oxidosqualene cyclase gene encoding a β -amylin synthase involved in *Polygala tenuifolia* Willd. saponin biosynthesis. *Plant Cell Rep* 33:511–519
- Liu J, Li M, Zhang Q, Wei X, Huang X (2020) Exploring the molecular basis of heterosis for plant breeding. *J Integr Plant Biol* 62(3):287–298
- Nguyen DH, Doan HT, Vu TV, Pham QT, Khoi NM, Huu TN, Thuong PT (2020) Oligosaccharide and glucose esters from the roots of *Polygala arillata*. *Nat Prod Res* 34(20):2900–2906
- Peng YQ, Fan LL, Mao FY, Zhao YS, Xu R, Yin YJ, Chen X, Wan DG, Zhang XH (2018) Genetic diversity and population structure of a protected species: *Polygala tenuifolia* Willd. *CR Biol* 341(3):152–159
- Sharma D et al (2018) Genome wide association mapping of agro-morphological traits among a diverse collection of finger millet (*Eleusine coracana* L.) genotypes using SNP markers. *PLoS ONE* 13(8):e0199444
- Thudi M, Palakurthi R, Schnable JC et al (2021) Genomic resources in plant breeding for sustainable agriculture. *J Plant Physiol* 257:153351
- Toebe M, Machado L, Tartaglia FL, Carvalho JO, Bandeira CT, Cargnelutti Filho A (2018) Sample size for estimating mean and coefficient of variation in species of crotalaria. *An Acad Bras Cienc* 90(2):1705–1715
- Vinh LB, Heo M, Phong NV, Ali I, Koh YS, Kim YH, Yang SY (2020) Bioactive compounds from *Polygala tenuifolia* and their inhibitory effects on lipopolysaccharide-stimulated pro-inflammatory cytokine production in bone marrow-derived dendritic cells. *Plant (basel)* 9(9):1240
- Wang X, Komatsu S (2018) Proteomic approaches to uncover the flooding and drought stress response mechanisms in soybean. *J Proteom* 172:201–215
- Wang D, Bai L, Xu X, Zhang F, Xing J, Jia J, Tian H, Qin X (2016) Rationality of commodity criteria and traditional breeding of *Polygala tenuifolia* based on agronomic traits and determination of chemical components. *China J Chin Materia Med* 41(20):3733–3740
- Yang Y, Sun M, Li S, Chen Q, Teixeira da Silva JA, Wang A, Yu X, Wang L (2020) Germplasm resources and genetic breeding of *Paeonia*: a systematic review. *Hortic Res* 7:107
- Zhang TZ et al (2016) Research progress on *Polygalae Radix*. *Chin Tradit Herbal Drugs* 47:2381–2389
- Zhang F-S et al (2017) Research progress in correlation between commodity specification level and quality of *Polygala tenuifolia*. *Chin Tradit Herbal Drugs* 48(12):2538–2547
- Zhang L et al (2020) Content determination and quality analysis of five main components in *Polygala tenuifolia* from different habitats. *Chin Tradit Herb Drugs* 51(06):1609–1616
- Zhao X, Xu B, Wu P, Zhao P, Guo C, Cui Y, Zhang Y, Zhang X, Li H (2020) UHPLC-MS/MS method for pharmacokinetic and bioavailability determination of five bioactive components in raw and various processed products of *Polygala tenuifolia* in rat plasma. *Pharm Biol* 58(1):969–978
- Zhou R et al (2015) Genetic diversity of cultivated and wild tomatoes revealed by morphological traits and SSR markers. *Genet Mol Res* 14(4):13868–13879

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.