

Spatial genetic and epigenetic structure of *Thlaspi arvense* (field pennycress) in China

Yabin Guan^{1,2}, Peng Qu², Shugang Lu², M. James C. Crabbe^{3,4,5},
Ticao Zhang^{6*} and Yupeng Geng^{2*}

¹School of Life Sciences, Yunnan University, Kunming, Yunnan 650500, China

²Institute of Ecology and Geobotany, School of Ecology and Environmental Sciences,
Yunnan University, Kunming, Yunnan 650500, China

³Wolfson College, Oxford University, Oxford OX2 6UD, UK

⁴School of Life Sciences, Shanxi University, Taiyuan, Shanxi 030006, China

⁵Institute of Biomedical and Environmental Science and Technology,
University of Bedfordshire, Luton LU1 3JU, UK

⁶College of Chinese Material Medica, Yunnan University of Chinese
Medicine, Kunming, Yunnan 650500, China

(Received 13 May 2020, accepted 15 July 2020; J-STAGE Advance published date: 11 November 2020)

Thlaspi arvense (field pennycress) is widespread in temperate regions of the northern hemisphere. We estimated the genetic and epigenetic structure of eight *T. arvense* populations (131 individuals) in China using amplified fragment length polymorphism and methylation-sensitive amplified polymorphism molecular-marker techniques. We detected low diversity at both genetic (mean = 0.03; total = 0.07) and epigenetic (mean = 0.04; total = 0.07) levels, while significant genetic ($F_{ST} = 0.42, P < 0.001$) and epigenetic ($F_{ST} = 0.32, P < 0.001$) divergence was found across the distribution range. Using Mantel testing, we found spatial genetic and epigenetic differentiation, consistent with isolation-by-distance models. We also identified a strong correlation between genetic and epigenetic differentiation ($r = 0.7438, P < 0.001$), suggesting genetic control of the epigenetic variation. Our results indicate that mating system, natural selection and gene flow events jointly structure spatial patterns of genetic and epigenetic variation. Moreover, epigenetic variation may serve as a basis of natural selection and ecological evolution to enable species to adapt to heterogeneous habitats. Our study provides novel clues for the adaptation of *T. arvense*.

Key words: adaptive evolution, epigenetic structure, genetic structure, methylation-sensitive amplified polymorphism (MSAP), *Thlaspi arvense*

INTRODUCTION

Thlaspi arvense L. (field pennycress), an annual diploid ($2n = 2x = 14$) herbal species of the tribe Thlaspidaceae (Brassicaceae), is widely distributed in temperate regions of the northern hemisphere. In China, it occupies a wide range of elevations, from sea level to approximately 4,000 m above sea level (Warwick et al., 2002; An et al., 2015). It can adapt to highly heterogeneous environments, and is thus a promising model species for exploring the mechanisms that underpin adaptability to diverse environments. Similar to its relatives in the Brassica-

ceae, mutations in *Flowering Locus C* (*FLC*) resulted in two ecotypes: the spring-flowering habit (early flowering) is free from vernalization in the transition from vegetative to reproductive growth, whereas vernalization is necessary to initiate flowering for the winter-flowering habit (late flowering) (McIntyre and Best, 1978; Dorn et al., 2018). In recent years, the plant has attracted great attention, since *T. arvense* possesses excellent agronomic traits, and comparative genomics has revealed substantial synteny with well-studied plants. Furthermore, transgenic protocols for *Arabidopsis thaliana* have also worked well for *T. arvense* (Phippen and Phippen, 2012; Sedbrook et al., 2014; McGinn et al., 2019).

The draft genome of *T. arvense* was published in 2015 (Dorn et al., 2015); however, little has been reported on the distribution of genetic variation or the genetic structure of natural populations of *T. arvense*. Genetic varia-

Edited by Hideki Innan

* Corresponding authors. zhangticao@mail.kib.ac.cn (TZ)
ypgeng@ynu.edu.cn (YG)

DOI: <http://doi.org/10.1266/ggs.20-00025>

tion is common in nature; it serves as the substrate for natural selection, and has been an important indicator in assessing the persistence of species and the capability to adapt to changing environments (Fisher, 1930; Frankham et al., 2002). Genetic variation exists not only within populations, but also among populations. On one hand, *T. arvense* occupies wide distribution ranges, geographic populations are scattered in highly heterogeneous environments, and divergent selective pressures are likely to drive the adaptive evolution and differentiation between local populations. On the other hand, geographic distance, as well as geographic barriers, have obstructed the genetic connectivity between populations. More importantly, divergent flowering times may be another factor that contributes to population differentiation.

Genetic variation clearly underpins adaptation to ecological environments (Yang et al., 2016; Yu et al., 2016; Ma et al., 2019; Zhang et al., 2019). Recent studies have also revealed divergent phenotypes in genetically uniform plant populations, as the consequences of epigenetic variation between individuals (Wilschut et al., 2016). Epigenetics refers to any process that has the potential to regulate gene expression and thus affect phenotype, but without any changes in DNA sequence; these include DNA methylation, small RNA expression and chromatin modifications (Richards, 2011; Deichmann, 2016). The most studied and heritable epigenetic modification is DNA methylation, where the 5-position of cytosine is modified by a methyl group in three sequence contexts: CG (or CpG), CHG and CHH (where H represents A, T or C) (Jones and Takai, 2001; Fulneček et al., 2002). In general, methylation of promoter and enhancer regions represses gene expression, whereas methylation of the gene body results in increased gene expression (Klose and Bird, 2006; Jones, 2012). The origin of epigenetic variation is mainly due to genetic variation, as a consequence of environmentally induced or inherited epimutation (Richards et al., 2017). Independent of genetic variation, abundant DNA methylation variation contributes to phenotypic differentiation and ecological adaptation in animals and plants. For instance, differences in DNA methylation affect personality traits in the great tit (*Parus major*) (Verhulst et al., 2016). They also contribute to heritable variation in flowering time and adaptive divergence in the apomictic dandelion (*Taraxacum officinale*); this variation can be synchronized using a methyltransferase inhibitor (Wilschut et al., 2016). The DNA methylation pattern of alligator weed, inherited stably across generations in the field, was immediately altered following common garden transplantation (Shi et al., 2018). Variation in DNA methylation is readily influenced by the environment, and an accurate DNA methylation state is crucial to some biological and developmental processes (Zhang et al., 2018). An understanding of DNA methylation variation may provide new insights into ecological adaptation

and evolution. As such, one may predict that natural populations under different selection regimes would differ in epigenetic states as a result of directional selection and evolutionary adaptation.

Knowing how genetic and epigenetic diversity are distributed over a large landscape scale is essential for understanding adaptive evolution, and is of great importance for germplasm utilization for *T. arvense*. Here, we applied amplified fragment length polymorphism (AFLP) and methylation-sensitive amplified polymorphism (MSAP) methods in *T. arvense* to address the following questions: (i) Are there genetic and epigenetic differentiation between populations? (ii) What is the relationship between genetic and epigenetic differentiation?

MATERIALS AND METHODS

Plant sampling strategy In this study, we sampled eight populations from low (35–705 m), middle (2,129–2,140 m), and high (3,806–4,277 m) elevations (Table 1, Fig. 1). Three populations were sampled from low-elevation locations: Xi'an, Shanxi Province (Pop1), Lushan, Henan Province (Pop2) and Hefei, Anhui Province (Pop3). The middle-elevation group comprised two populations from sites in Yunnan Province: Kunming (Pop4) and Yongsheng (Pop5). Another three populations were sampled from high elevations in Tibet: Gongbo'gyamda (Pop6), Mangkang (Pop7) and Zuogong (Pop8) (Table 1, Fig. 1).

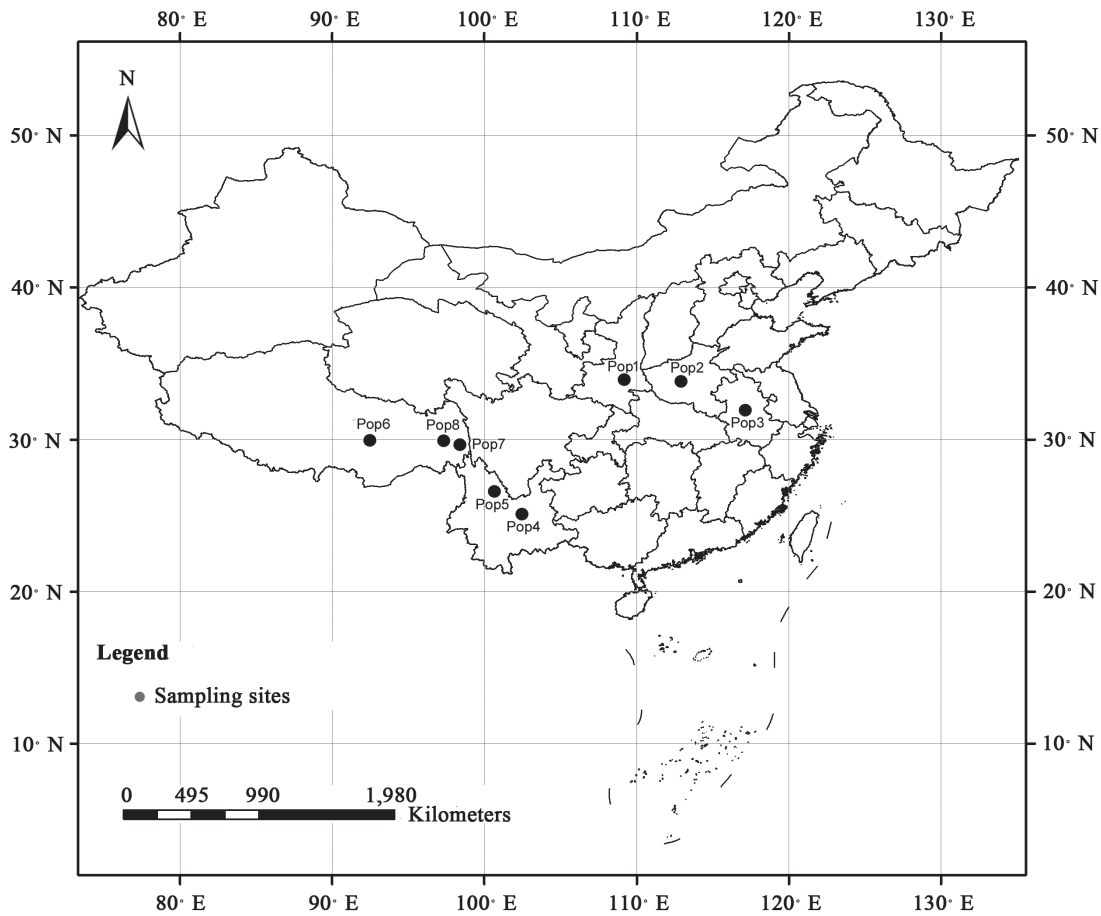
Since DNA methylation is differentiated between tissues and developmental stages, only the third and fourth true leaves were sampled when the plants started to flower, to ensure the samples were from the same developmental stage. Pairs of sampling sites were separated by a linear distance of more than 85 km. To avoid repeated sampling of individuals derived from the same mother plant or its relatives, we only sampled plants more than 30 m apart. Finally, a total of 131 samples were collected from the eight geographic populations through multiple sampling in 2017 (Table 1).

AFLP and MSAP protocol Total genomic DNA was extracted from dried leaves using a Qiagen DNeasy Plant Mini kit according to the manufacturer's protocol. The extracted DNA was tested by 1% agarose gel electrophoresis and then stored at -80°C for future use.

Population genetic diversity and structure were analyzed by AFLP according to the protocol of Vos et al. with some adjustments (Vos et al., 1995; Gao et al., 2010). We screened 10 EcoRI/MseI primer pairs to selectively amplify the following fragments in all samples: AA/CCC, AC/CAG, AC/CCT, AC/CGA, AC/CCC, AG/CGA, AG/CGC, AT/CCC, AT/CCG and AT/CGA (Supplementary Table S1). The MSAP method was applied to assess epigenetic variation within and between populations. In the AFLP-derived

Table 1. Sampling information for eight geographic populations of *T. arvense*

Elevation group	ID	Size	Location	Coordinates	Elevation (m)	Sampling date
Low	Pop1	20	Xi'an	E 109.1161°, N 34.0223°	705	15 Apr., 2017
	Pop2	20	Lushan	E 112.9109°, N 33.7603°	134	17 Apr., 2017
	Pop3	20	Hefei	E 117.1395°, N 31.9241°	35	22 Apr., 2017
Middle	Pop4	18	Kunming	E 102.5310°, N 25.0936°	2129	3 Mar., 2017
	Pop5	20	Yongsheng	E 100.7456°, N 26.6629°	2140	6 Feb., 2017
High	Pop6	9	Gongbo'gyamda	E 92.4856°, N 29.8941°	4277	12 Aug., 2017
	Pop7	14	Mangkang	E 98.3652°, N 29.6576°	3806	15 Aug., 2017
	Pop8	10	Zuogong	E 97.5123°, N 29.9556°	4006	18 Aug., 2017

Fig. 1. Map of China showing sampling locations of the eight geographic populations of *T. arvense*.

MSAP technique, first developed by Reyna-López et al. (1997), MseI used in AFLP was replaced by a pair of isoschizomers, HpaII and MspI, that have the same cleavage

site at 5'-CCGG sequences but differ in their sensitivities to the methylation state of cytosine. The following 10 selective amplification primer pairs were developed to

detect methylation variation at CCGG sites: AA/TTG, AA/TAC, AC/ATC, AC/ACT, AC/TCG, AC/TCCT, AC/ACG, AG/TGC, AG/TAG and AT/TCA. The detailed protocol can be found in Gao et al. (2010). After denaturation at 95 °C for 5 min, the final selective products were separated by 6% denaturing polyacrylamide gel electrophoresis. The gels were then stained in 0.15% AgNO₃ for 10 min and developed in 1500 ml of 2% NaOH containing 2 ml of formalin.

Genetic and epigenetic analysis The raw data were represented by a binary matrix, with band presence and absence coded as “1” and “0”, respectively. The MSAP matrix was also transformed into unmethylated marker (MISP) and methylated marker matrices in which methylation types were characterized using the “Mixed Scoring 1” classification system described by Schulz et al. (2013): for MISP markers, only condition I (fragments present in both profiles) was scored as “1”; for methylated markers, both conditions II (fragments present only in EcoRI/MspI profiles) and III (fragments present only in EcoRI/HpaII profiles) were scored as “1”.

All subsequent analyses were based on the two AFLP and methylated marker data binary matrices. The following genetic and epigenetic diversity parameters were all estimated using GenAlEx 6.503 (Peakall and Smouse, 2012): number of different alleles (Na), number of effective alleles (Ne), Shannon’s information index (I), expected heterozygosity (H_E), number of private bands (NPB), pairwise population differentiation values (PhiPT), percentage of polymorphic bands (PPB) and Nei’s genetic and epigenetic distance. Analysis of molecular variance (AMOVA) based on 1,000 permutations was conducted in Arlequin 3.1 (Excoffier et al., 2005) to calculate the partitioning of variation within and among the eight geographic populations of *T. arvense*. In addition, principal coordinate analysis (PCoA) using Nei’s genetic and epigenetic distance, which is based on the two kinds of markers, was performed to visualize relationships among the populations. The resulting distance matrices were also used to construct dendrograms based on the unweighted pair-group method with arithmetic averages in MEGA 4.1 followed by modification with the iTOL v4 online tool (<https://itol.embl.de/>).

To identify the number of subpopulations based on genetic and epigenetic backgrounds, Bayesian clustering was carried out in STRUCTURE 2.3.4 (Hubisz et al., 2009). As AFLP and MSAP markers are both dominant molecular markers, we defined “0” as the recessive allele at each locus (Falush et al., 2007). All Bayesian analyses were performed under the admixture model. To determine the optimal number of subpopulations (K), 10 runs were performed for each value of K from 2 to 8, with each run comprising a 10,000-iteration burn-in period followed by 10,000 iterations. The output results were

then uploaded to Structure Harvester to seek the optimal K -value, which corresponded to the maximum ΔK value (Earl and vonHoldt, 2011).

To test whether the observed genetic and epigenetic differentiation associated with geographic distance accorded with classical isolation-by-distance (IBD) models, and whether the one-to-one genetic and epigenetic differentiation between the eight *T. arvense* populations have the same tendency, we also performed multiple Mantel tests based on 9,999 permutations in GenAlEx 6.503 (Peakall and Smouse, 2012).

RESULTS

Genetic distance and diversity In this study, 461 and 400 repeatable loci were respectively amplified in genetic and epigenetic analyses of the 131 individuals from eight geographic populations. High genetic and epigenetic identities were found between the eight populations (Table 2), with similar patterns observed in both types of analyses. In the genetic analysis, the highest identity was found between Pop1 and Pop3 (Nei’s genetic identity = 0.993), and the lowest was between Pop4 and Pop6 (0.950). In the epigenetic analysis, the highest identity (0.991) was also between Pop1 and Pop3, while the lowest was between Pop5 and Pop6 (0.956). Cluster analyses of the genetic and epigenetic data grouped the eight populations into three clades (Fig. 2), one containing all high-elevation populations, one comprising Pop5 alone and a third consisting of all low-elevation populations plus middle-elevation Pop4.

In the genetic diversity analysis, 461 loci were clearly amplified, among which 4.12–19.96% were polymorphic among the eight geographic populations (Table 3). An average of 11.38% and 36.88% were polymorphic at population and species levels, respectively. NPB in the eight populations varied from 0 to 11 (mean = 5), and Na in each population ranged from 0.97 (Pop5) to 0.72 (Pop6), with an average of 0.85, and was 1.25 at the species

Table 2. Pairwise population matrix of Nei’s genetic and epigenetic identity between eight populations of *T. arvense*

Pop	Pop1	Pop2	Pop3	Pop4	Pop5	Pop6	Pop7	Pop8
Pop1	–	0.987	0.991	0.985	0.975	0.976	0.985	0.985
Pop2	0.988	–	0.985	0.984	0.967	0.964	0.973	0.974
Pop3	0.993	0.984	–	0.987	0.974	0.973	0.982	0.979
Pop4	0.976	0.969	0.980	–	0.967	0.964	0.974	0.970
Pop5	0.963	0.957	0.960	0.957	–	0.956	0.969	0.968
Pop6	0.976	0.967	0.971	0.950	0.955	–	0.989	0.983
Pop7	0.975	0.964	0.973	0.960	0.960	0.981	–	0.987
Pop8	0.978	0.972	0.973	0.956	0.968	0.986	0.980	–

Nei’s genetic identity is below the diagonal and Nei’s epigenetic identity is above the diagonal.

level. Similar to the Na, the highest genetic diversity index value in the eight populations was in Pop5 ($N_e = 1.12$; $I = 0.10$; $H_E = 0.07$); the lowest was in Pop6 ($N_e =$

1.03; $I = 0.02$; $H_E = 0.02$). Average values at the population level were $N_e = 1.06$, $I = 0.05$, and $H_E = 0.03$; at the species level, $N_e = 1.12$, $I = 0.12$, and $H_E = 0.07$.

In the epigenetic diversity analysis (Table 3), 400 distinct loci were amplified. PPB among the eight populations varied from 8.00% (Pop6) to 21.50% (Pop5), with an average of 13.78%, and 38.25% of bands were polymorphic at the species level. The average NPB among the eight populations was 5. Pop5 still had the highest NPB (12), while no private band was detected in Pop6. Furthermore, the maximum values of most other diversity indexes were detected in Pop5, whereas the minimum values were detected in Pop6.

According to PCoA plots (Fig. 3), low- and high-elevation populations tended to cluster. In analyses of the genetic data (Fig. 3A), middle-elevation populations Pop4 and Pop5 did not cluster together. In contrast, PCoA of the epigenetic data grouped Pop4 with low-elevation populations (Fig. 3B). In addition, Pop2 and Pop6 were slightly less associated with other populations

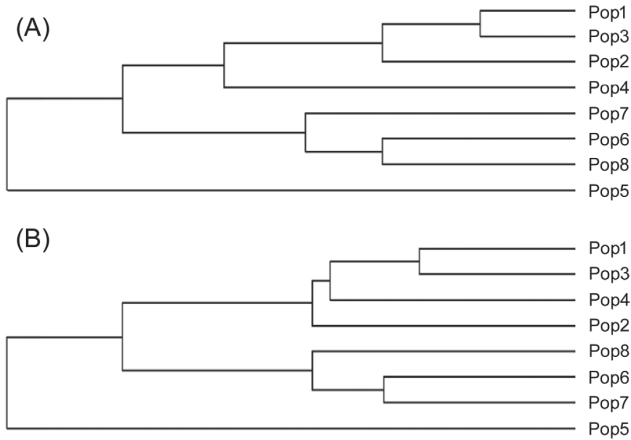


Fig. 2. Dendrograms based on Nei's distance matrices of (A) genetic and (B) epigenetic data.

Table 3. Genetic and epigenetic diversity indices among eight populations of *T. arvense*

Indices		Pop1	Pop2	Pop3	Pop4	Pop5	Pop6	Pop7	Pop8	Average	Species level
N ^a		20	20	20	18	20	9	14	10	16	131
Na ^b	Genetic	0.90	0.87	0.86	0.81	0.97	0.72	0.79	0.84	0.85	1.25
	Epigenetic	0.52	0.50	0.55	0.53	0.65	0.44	0.49	0.50	0.52	0.93
Ne ^c	Genetic	1.06	1.05	1.05	1.04	1.12	1.03	1.05	1.06	1.06	1.12
	Epigenetic	1.05	1.05	1.08	1.06	1.13	1.05	1.06	1.07	1.07	1.11
I ^d	Genetic	0.05	0.05	0.04	0.04	0.10	0.02	0.04	0.06	0.05	0.12
	Epigenetic	0.05	0.05	0.07	0.06	0.11	0.04	0.06	0.06	0.06	0.11
H _E ^e	Genetic	0.03	0.03	0.03	0.03	0.07	0.02	0.03	0.04	0.03	0.07
	Epigenetic	0.03	0.03	0.05	0.04	0.08	0.03	0.04	0.04	0.04	0.07
NPB ^f	Genetic	11	9	7	1	11	0	1	2	5	–
	Epigenetic	8	6	6	2	12	0	2	3	5	–
PPB ^g	Genetic	14.32	12.15	11.93	8.46	19.96	4.12	8.46	11.28	11.33	36.88
	Epigenetic	13.50	13.25	16.00	13.75	21.50	8.00	11.75	12.50	13.78	38.25

^aNumber of samples; ^bnumber of different alleles; ^cnumber of effective alleles; ^dShannon's information index; ^eexpected heterozygosity; ^fnumber of private bands; ^gpercentage of polymorphic bands.

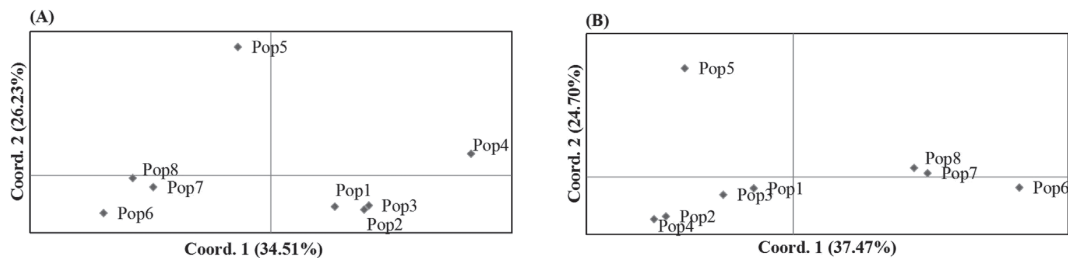


Fig. 3. Plots of the first two principal coordinate scores of eight populations of *T. arvense* based on AFLP and methylated markers. (A) Plot based on the genetic data, in which the first and second principal coordinates explained 34.51% and 26.23% of the total variation, respectively. (B) Plot based on the epigenetic data, in which the first and second principal coordinates explained 37.47% and 24.70% of the total variation, respectively.

at the same elevation compared with their genetic clustering patterns. These groupings were consistent with the patterns revealed by population pairwise genetic and epigenetic identities.

In STRUCTURE analyses of the genetic and epigenetic data, the maximum ΔK values were obtained at $K = 5$. The 131 individuals from the eight geographic populations were therefore assigned to five genetic (Fig. 4A) or epigenetic (Fig. 4B) ancestral Hardy–Weinberg equilibrium populations based on their potential genetic or epigenetic backgrounds. In the genetic clustering analysis (Supplementary Table S2), the ancestry membership participation of Pop1, Pop3 and Pop4 in cluster V was > 44% (44.7–96.5%), while Pop6, Pop7 and Pop8 in cluster I had > 63% (64.0–81.8%) of ancestry membership coefficients, and Pop2 and Pop5 had higher ancestry membership coefficients in cluster II and III, with 78.5% and 55.3%, respectively. In the epigenetic clustering (Supplementary Table S2), Pop1, Pop2 and Pop4 had > 63% (64.0–96.7%) of ancestry membership participation in cluster V, whereas Pop6, Pop7 and Pop8 had higher ancestry membership coefficients > 75% (75.7–81.9%) in cluster IV, and Pop3 and Pop5 had ancestry membership contributions of 53.6% and 55.8% to clusters II and III, respectively.

Most of the variation in the two kinds of markers uncovered by AMOVA (Table 4) existed within geographic populations (58–68%). Nevertheless, obvious genetic and epigenetic differentiation was detected: approximately 42% of the variation was assigned to population genetic differ-

entiation ($P < 0.001$), while approximately 32% could be ascribed to population epigenetic divergence ($P < 0.001$).

IBD tests and correlations between genetic and epigenetic differentiation Mantel tests uncovered significant correlations between geographic distance (km) and population differentiation (Fig. 5A, 5B). The correlation between geographic distance and population genetic differentiation was stronger ($r = 0.5362$, $P < 0.01$) than that between geographic distance and epigenetic population

Table 4. Analysis of molecular variance (AMOVA) for the 131 individuals sampled from eight populations of *T. arvense* based on AFLP and methylated markers

Source of variation	d.f.	Sum of squares	Variation components	Percentage of variation	P-value
Genetic					
Among populations	7	735.598	5.96694	41.56	0.00000
Within populations	123	1032.097	8.39103	58.44	
Total	130	1767.695	14.35797	100	
Epigenetic					
Among populations	7	533.146	4.24120	32.09	0.00000
Within populations	120	1077.190	8.97659	67.91	
Total	127	1610.336	13.21779	100	

All the AMOVAs were based on 1,000 permutations.

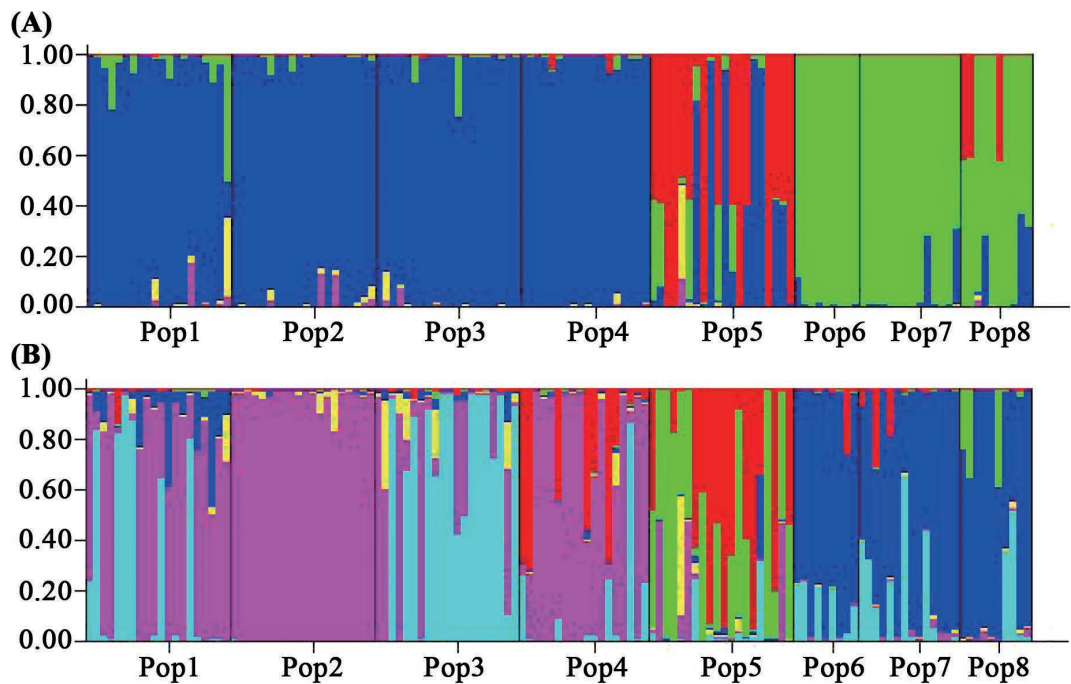


Fig. 4. Results of STRUCTURE Bayesian analysis of genetic structure ($K = 5$). (A) Genetic clusters based on AFLP markers. (B) Epigenetic clusters based on methylated markers.

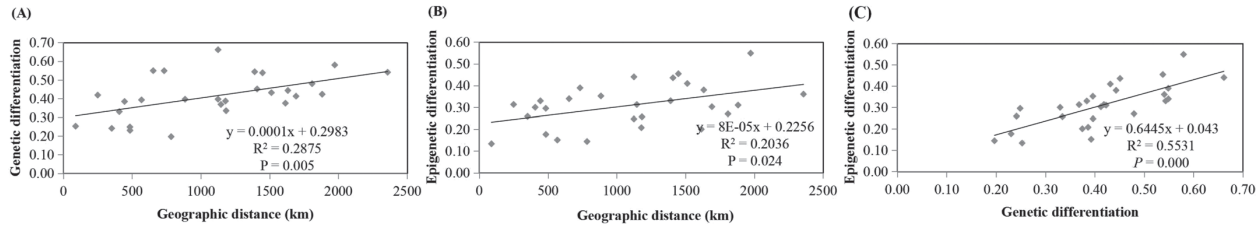


Fig. 5. Plots of Mantel test correlations between (A) geographic distance and genetic differentiation, (B) geographic distance and epigenetic differentiation, and (C) genetic and epigenetic differentiation.

differentiation ($r = 0.4512$, $P < 0.05$). This result demonstrates that population genetic and epigenetic differentiation in the analyzed *T. arvense* populations increased with geographic distance and that genetic differentiation was more evident under the IBD model. Furthermore, the most significant correlation ($r = 0.7437$, $P < 0.001$) detected by Mantel testing was between genetic differentiation and epigenetic differentiation (Fig. 5C), which indicates that local populations share similar patterns of genetic and epigenetic divergence.

DISCUSSION

Low genetic and epigenetic diversity In general, genetic diversity within allogamous species is three times larger than that within autogamous species (Ayala and Kiger, 1980), and heterozygosities within outcrossing populations are significantly higher than those within self-fertilizing ones (Nyblom, 2004). In our eight geographic populations, low expected heterozygosities were identified by genetic (mean = 0.03, total = 0.07) and epigenetic (mean = 0.04, total = 0.07) analyses. As defined by Goodwillie et al. (2005), an outcrossing species has an outcrossing rate greater than 80%, whereas outcrossing rates below 20% are characteristic of selfing species; consequently, *T. arvense*, which has a selfing rate above 80%, is a selfing species (Best and McIntyre, 1975). A previous study had revealed low genetic diversity within populations in a predominantly selfing plant, *Tacca chantrieri* (Zhang et al., 2006), and the high proportion of autogamy also contributed to a very low level of intra-population genetic diversity in *T. chantrieri* (Chen et al., 2008). These findings were similar to our observations in *T. arvense* (PPB = 36.88%, $I = 0.12$, $N_a = 1.25$). Thus, the limited genetic diversity could be largely attributed to the high proportion of autogamy in *T. arvense*. Other evolutionary processes, such as shorter colonization history, could also be related to the relatively low genetic diversity. Among the eight populations of *T. arvense*, Pop5, collected from the middle elevation, had the highest genetic diversity. Plant populations in glacial refugia and admixture zones possess the greatest genetic diversity (Jadwiszczak et al., 2015); however, a previous phylogeographic study on *T. arvense* revealed that the

richest alleles of the populations on the eastern edge of the Qinghai–Tibet Plateau were attributed to the admixture zone, rather than the glacial refugia, even though the populations were located in the universally recognized glacial refugia of Hengduan Mountain (An et al., 2015). It was likely that the location where Pop5 thrived was a transition zone of high- and low-elevation lineages, and the special climate conditions enabled them to coexist there; thus, the richest genetic diversity found within Pop5 might be explained by the hybrid zone effect.

The epigenetic diversity of the eight populations was very similar to their genetic diversity. The expected heterozygosities in epigenetics ranged from 0.03 to 0.08, with 0.04 on average and 0.07 at the species level. Only 38.25% of the 400 amplified loci were polymorphic, with the majority (61.75%) found to be monomorphic within the eight populations. These results contrast with previous reports of a higher rate of DNA methylation variation relative to nucleotide variation, and the prediction that the methylome evolves faster than the genome because of its environmental sensitivity (Jablonka and Raz, 2009; Becker et al., 2011; Foust et al., 2016). On the one hand, in our MSAP analysis, the absence of bands in both HpaII and MspI profiles (0, 0) might result from full methylation of the external cytosine, full methylation of both cytosines, hemimethylation of both cytosines or genetic variation. To avoid interference from genetic variation, “Mixed Scoring 1” (Schulz et al., 2013) was performed to split the MSAP matrix into MISP and methylated marker matrices, and only the methylated marker matrix was used to assess the epigenetic diversity and epigenetic structure. Consequently, part of the epigenetic variation may have been neglected in our analysis, resulting in the pattern that the epigenetic diversity was not higher than genetic diversity. Previous quantitative research on DNA methylation variation revealed that DNA methylation was unevenly distributed across the genome, but highly conserved between individuals that spanned divergent habitats (Trucchi et al., 2016). This is similar to our finding that up to 61.75% of total amplified loci were monomorphic; thus, DNA methylation was conserved in *T. arvense* even over a wide distribution range.

Distinguishing population divergence Accord-

ing to AMOVA, approximately 42% of genetic variation existed among populations. This result indicated considerable genetic differentiation among the eight populations ($F_{ST} > 0.25$) (Freeland, 2005). The mating system, which has effects on genetic diversity, also plays a role in genetic differentiation. The level of genetic variation among populations of outcrossing plant species is approximately 10–20%, increasing to approximately 50% in selfing species (Hamrick and Godt, 1989). The genetic differentiation observed in *T. arvense* in our study was similar to that in *Leavenworthia alabamica* ($F_{ST} = 0.45$) and *L. crassa* ($F_{ST} = 0.36$), two cedar glade endemic species that have both self-incompatible and self-compatible populations (Koelling et al., 2011), and was comparable to the average differentiation value ($F_{ST} = 0.43$) among self-fertilizing or clonal plants (Morjan and Rieseberg, 2004). Furthermore, genetic differentiation is related to gene flow, which is influenced by a wide range of factors, such as geographic distance, elevation, topography, climate and life history (Turner and Trexler, 1998; Guarnizo and Cannatella, 2013; Castillo et al., 2014; Polato et al., 2017). The frequency of gene flow across geographic regions is inversely proportional to geographic distance. In our study, the direct line distance between the eight populations ranged from 88 to 2,357 km, and, except for the middle elevation, a lower differentiation level was found between populations within the same elevation group, with 0.197–0.384 between the low-elevation populations and 0.230–0.393 between the high-elevation populations (Supplementary Table S3). Our IBD test ($r = 0.5362$, $P < 0.01$) also revealed significant genetic differentiation according to geographic distance. The eight populations of *T. arvense* spanned a large geographic scale, and the mountains, rivers, elevation and climate differences had clear effects on gene flow among geographic regions. Furthermore, we found divergent flowering time between the three elevations: the low-elevation populations flowered in the spring and the high-elevation populations flowered in summer, whereas the middle-elevation populations flowered in winter, resulting in prezygotic reproductive isolation. Apart from the gene flow, differentiated natural selection also contributed to the genetic divergence between local populations (Savolainen et al., 2007). The eight populations of *T. arvense* inhabited highly heterogeneous environments, and thus will have experienced divergent selective stresses. Several studies have described the selection on genetic variation which underpins adaptation to extreme environments, such as hypoxia, low temperature and high-dose ultraviolet radiation (Zhang et al., 2014; Zhang et al., 2016; Hallmark et al., 2019). Here, the high-elevation populations (Pop6, Pop7 and Pop8) collected from the Qinghai–Tibet Plateau had higher genetic differentiation from the other five populations collected from middle and low elevation (Supplementary Table S3). The Qinghai–Tibet

Plateau is well known for its extreme environmental conditions (low temperature, low oxygen and strong ultraviolet radiation); the three high-altitude populations of *T. arvense* should have experienced strong natural selection, and adaptive evolution may have taken place in related genes, and the limited gene flow between high elevation and middle and low elevations may therefore shape the higher differentiation level.

In our analysis of cytosine methylation variation within *T. arvense*, we showed significant epigenetic differentiation ($F_{ST} = 0.32$) among the eight natural populations from divergent habitats, and a significant correlation ($r = 0.4512$, $P < 0.05$) was also found between geographic distance and epigenetic differentiation. Variation in DNA methylation is subject to natural selection, and thus these fitness-related variations should accumulate in populations (Hirsch et al., 2012). A study of 1,001 worldwide epigenomes of *A. thaliana* showed that changes in DNA methylation were related to geographic origins and local adaptation (Kawakatsu et al., 2016). Thus, the epigenetic variation within the eight populations of *T. arvense* has been confronted with different elements for natural selection, resulting in the epigenetic differentiation between populations. Meanwhile, limited gene flow also contributed to the epigenetic differentiation and shaped the significant correlation between geographic distance and epigenetic differentiation within the natural populations.

In this study, we not only showed a significant population structure in genetics and epigenetics, but also found a strong correlation between genetic differentiation and epigenetic differentiation ($r = 0.7483$, $P < 0.001$), which is consistent with the hypothesis that epigenetic variation is mainly due to genetic variation (Richards et al., 2017). Massive methylome sequencing in *Arabidopsis* and soybean identified about 23% and 35% of differentially methylated regions that were associated with genetic polymorphisms (Schmitz et al., 2013; Shen et al., 2018). The strong correlation between genetics and epigenetics is likely attributable to genetic control of epigenetic variation.

In conclusion, we selected the widespread annual *T. arvense* as a new model system to explore the spatial genetic and epigenetic structure of natural plant populations. Our analyses revealed that this annual species has low genetic and epigenetic diversity even though it grows in highly heterogeneous habitats, but genetic and epigenetic differentiation was evident between our sampled populations. Our findings indicate that mating system, geographical distribution and habitat play important roles in shaping the genetic and epigenetic structure of natural plant populations. Moreover, we noted that the variation in the methylome is often associated with genetic polymorphism. Further research on comparative genomics and population genetics in the context of high-

throughput sequencing should accelerate the development and utilization of *T. arvense*, and provide insights into adaptation and evolution for widespread plant species.

This work was supported by the National Natural Science Foundation of China (31660057, 31770408, 32060059, 32060237). We thank Ruiwen Li, Pin Zhang, Landi Luo, Ting Lu, Jiashan Luo and Na Chang for assistance in lab experiments and field sampling.

REFERENCES

- An, M., Zeng, L., Zhang, T., and Zhong, Y. (2015) Phylogeography of *Thlaspi arvense* (Brassicaceae) in China inferred from chloroplast and nuclear DNA sequences and ecological niche modeling. *Int. J. Mol. Sci.* **16**, 13339–13355.
- Ayala, F. J., and Kiger, J. A. (1980) *Modern Genetics*. Benjamin/Cummings, Menlo Park, CA, USA.
- Becker, C., Hagmann, J., Müller, J., Koenig, D., Stegle, O., Borgwardt, K., and Weigel, D. (2011) Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature* **480**, 245–249.
- Best, K. F., and McIntyre, G. I. (1975) The biology of Canadian weeds. 9. *Thlaspi arvense* L. *Can. J. Plant Sci.* **55**, 279–292.
- Castillo, J. A., Epps, C. W., Davis, A. R., and Cushman, S. A. (2014) Landscape effects on gene flow for a climate-sensitive montane species, the American pika. *Mol. Ecol.* **23**, 843–856.
- Chen, Y.-Y., Li, X.-L., Yin, L.-Y., and Li, W. (2008) Genetic diversity of the threatened aquatic plant *Ottelia alismoides* in the Yangtze River. *Aquat. Bot.* **88**, 10–16.
- Deichmann, U. (2016) Epigenetics: the origins and evolution of a fashionable topic. *Dev. Biol.* **416**, 249–254.
- Dorn, K. M., Fankhauser, J. D., Wyse, D. L., and Marks, M. D. (2015) A draft genome of field pennycress (*Thlaspi arvense*) provides tools for the domestication of a new winter biofuel crop. *DNA Res.* **22**, 121–131.
- Dorn, K. M., Johnson, E. B., Daniels, E. C., Wyse, D. L., and Marks, M. D. (2018) Spring flowering habit in field pennycress (*Thlaspi arvense*) has arisen multiple independent times. *Plant Direct* **2**, 1–10.
- Earl, D. A., and vonHoldt, B. M. (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361.
- Excoffier, L., Laval, G., and Schneider, S. (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* **1**, 47–50.
- Falush, D., Stephens, M., and Pritchard, J. K. (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol. Ecol. Notes* **7**, 574–578.
- Fisher, R. A. (1930) *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford, UK.
- Foust, C. M., Preite, V., Schrey, A. W., Alvarez, M., Robertson, M. H., Verhoeven, K. J. F., and Richards, C. L. (2016) Genetic and epigenetic differences associated with environmental gradients in replicate populations of two salt marsh perennials. *Mol. Ecol.* **25**, 1639–1652.
- Frankham, R., Ballou, J. D., and Briscoe, D. A. (2002) *Introduction to Conservation Genetics*. Cambridge University Press, Cambridge, UK.
- Freeland, J. R. (2005) *Molecular Ecology*. John Wiley and Sons, Chichester, UK.
- Fulneček, J., Matyášek, R., and Kovařík, A. (2002) Distribution of 5-methylcytosine residues in 5S rRNA genes in *Arabidopsis thaliana* and *Secale cereale*. *Mol. Genet. Genomics* **4**, 510–517.
- Gao, L., Geng, Y., Li, B., Chen, J., and Yang, J. (2010) Genome-wide DNA methylation alterations of *Alternanthera philoxeroides* in natural and manipulated habitats: implications for epigenetic regulation of rapid responses to environmental fluctuation and phenotypic variation. *Plant Cell Environ.* **33**, 1820–1827.
- Goodwillie, C., Kalisz, S., and Eckert, C. G. (2005) The evolutionary enigma of mixed mating systems in plants: occurrence, theoretical explanations, and empirical evidence. *Annu. Rev. Ecol. Evol. Syst.* **36**, 47–79.
- Guarnizo, C. E., and Cannatella, D. C. (2013) Genetic divergence within frog species is greater in topographically more complex regions. *J. Zool. Syst. Evol. Res.* **51**, 333–340.
- Hallmark, B., Karafet, T. M., Hsieh, P., Osipova, L. P., Watkins, J. C., and Hammer, M. F. (2019) Genomic evidence of local adaptation to climate and diet in indigenous siberians. *Mol. Biol. Evol.* **36**, 315–327.
- Hamrick, J. L., and Godt, M. J. W. (1989) Allozyme diversity in plant species. In *Plant Population Genetics, Breeding and Genetic Resources*. (eds.: Brown, A. H. D., Clegg, M. T., Kahler, A. L., and Weir, B. S.), pp. 43–63. Sinauer Associates, Sunderland, MA, USA.
- Hirsch, S., Baumberger, R., and Grossniklaus, U. (2012) Epigenetic variation, inheritance, and selection in plant populations. *Cold Spring Harb. Symp. Quant. Biol.* **77**, 97–104.
- Hubisz, M. J., Falush, D., Stephens, M., and Pritchard, J. K. (2009) Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.* **9**, 1322–1332.
- Jablonka, E., and Raz, G. (2009) Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Q. Rev. Biol.* **84**, 131–176.
- Jadwiszczak, K. A., Banaszek, A., Chrzanowska, A., Kłosowski, S., and Sozinov, O. V. (2015) The admixture zone of *Betula humilis* Schrk. phylogenetic lineages follows the eastern central European suture zone. *Plant Ecol. Divers.* **8**, 323–329.
- Jones, P. A. (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13**, 484–492.
- Jones, P. A., and Takai, D. (2001) The role of DNA methylation in mammalian epigenetics. *Science* **293**, 1068–1070.
- Kawakatsu, T., Huang, S. C., Jupe, F., Sasaki, E., Schmitz, R. J., Urich, M. A., Castanon, R., Nery, J. R., Barragan, C., He, Y., et al. (2016) Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions. *Cell* **166**, 492–505.
- Klose, R. J., and Bird, A. P. (2006) Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.* **31**, 89–97.
- Koelling, V. A., Hamrick, J. L., and Mauricio, R. (2011) Genetic diversity and structure in two species of *Leavenworthia* with self-incompatible and self-compatible populations. *Heredity* **106**, 310–318.
- Ma, Y.-F., Han, X.-M., Huang, C.-P., Zhong, L., Adeola, A. C., Irwin, D. M., Xie, H.-B., and Zhang, Y.-P. (2019) Population genomics analysis revealed origin and high-altitude adaptation of Tibetan pigs. *Sci. Rep.* **9**, 11463.
- McGinn, M., Phippen, W. B., Chopra, R., Bansal, S., Jarvis, B. A., Phippen, M. E., Dorn, K. M., Esfahanian, M., Nazarene, T. J., Cahoon, E. B., et al. (2019) Molecular tools enabling pen-

- nycress (*Thlaspi arvense*) as a model plant and oilseed cash cover crop. *Plant Biotechnol. J.* **17**, 776–788.
- McIntyre, G. I., and Best, K. F. (1978) Studies on the flowering of *Thlaspi arvense* L. IV. genetic and ecological differences between early- and late-flowering strains. *Bot. Gaz.* **139**, 190–195.
- Morjan, C. L., and Rieseberg, L. H. (2004) How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Mol. Ecol.* **13**, 1341–1356.
- Nybom, H. (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Mol. Ecol.* **13**, 1143–1155.
- Peakall, R., and Smouse, P. E. (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* **28**, 2537–2539.
- Phippen, W. B., and Phippen, M. E. (2012) Soybean seed yield and quality as a response to field pennycress residue. *Crop Sci.* **52**, 2767–2773.
- Polato, N. R., Gray, M. M., Gill, B. A., Becker, C. G., Casner, K. L., Flecker, A. S., Kondratieff, B. C., Encalada, A. C., Poff, N. L., Funk, W. C., et al. (2017) Genetic diversity and gene flow decline with elevation in montane mayflies. *Heredity* **119**, 107–116.
- Reyna-López, G. E., Simpson, J., and Ruiz-Herrera, J. (1997) Differences in DNA methylation patterns are detectable during the dimorphic transition of fungi by amplification of restriction polymorphisms. *Mol. Gen. Genet.* **253**, 703–710.
- Richards, C. L., Alonso, C., Becker, C., Bossdorf, O., Bucher, E., Colomé-Tatché, M., Durka, W., Engelhardt, J., Gaspar, B., Gogol-Döring, A., et al. (2017) Ecological plant epigenetics: evidence from model and non-model species, and the way forward. *Ecol. Lett.* **20**, 1576–1590.
- Richards, E. J. (2011) Natural epigenetic variation in plant species: a view from the field. *Curr. Opin. Plant Biol.* **14**, 204–209.
- Savolainen, O., Pyhäjärvi, T., and Knürr, T. (2007) Gene flow and local adaptation in trees. *Annu. Rev. Ecol. Evol. Syst.* **38**, 595–619.
- Schmitz, R. J., Schultz, M. D., Urlich, M. A., Nery, J. R., Pelizzola, M., Libiger, O., Alix, A., McCosh, R. B., Chen, H., Schork, N. J., et al. (2013) Patterns of population epigenomic diversity. *Nature* **495**, 193–198.
- Schulz, B., Eckstein, R. L., and Durka, W. (2013) Scoring and analysis of methylation-sensitive amplification polymorphisms for epigenetic population studies. *Mol. Ecol. Resour.* **13**, 642–653.
- Sedbrook, J. C., Phippen, W. B., and Marks, M. D. (2014) New approaches to facilitate rapid domestication of a wild plant to an oilseed crop: example pennycress (*Thlaspi arvense* L.). *Plant Sci.* **227**, 122–132.
- Shen, Y., Zhang, J., Liu, Y., Liu, S., Liu, Z., Duan, Z., Wang, Z., Zhu, B., Guo, Y.-L., and Tian, Z. (2018) DNA methylation footprints during soybean domestication and improvement. *Genome Biol.* **19**, 128.
- Shi, W., Chen, X., Gao, L., Xu, C.-Y., Ou, X., Bossdorf, O., Yang, J., and Geng, Y. (2018) Transient stability of epigenetic population differentiation in a clonal invader. *Front. Plant Sci.* **9**, 1851.
- Trucchi, E., Mazzarella, A. B., Gilfillan, G. D., Lorenzo, M. T., Schönschetter, P., and Paun, O. (2016) BsRADseq: screening DNA methylation in natural populations of non-model species. *Mol. Ecol.* **25**, 1697–1713.
- Turner, T. F., and Trexler, J. C. (1998) Ecological and historical associations of gene flow in darters (Teleostei: Percidae). *Evolution* **52**, 1781–1801.
- Verhulst, E. C., Mateman, A. C., Zwier, M. V., Caro, S. P., Verhoeven, K. J. F., and van Oers, K. (2016) Evidence from pyrosequencing indicates that natural variation in animal personality is associated with DRD4 DNA methylation. *Mol. Ecol.* **25**, 1801–1811.
- Vos, P., Hogers, R., Bleeker, M., Reijmans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., et al. (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**, 4407–4414.
- Warwick, S. I., Francis, A., and Susko, D. J. (2002) The biology of Canadian weeds. 9. *Thlaspi arvense* L. (updated). *Can. J. Plant Sci.* **82**, 803–823.
- Wilschut, R. A., Oplaat, C., Snoek, L. B., Kirschner, J., and Verhoeven, K. J. (2016) Natural epigenetic variation contributes to heritable flowering divergence in a widespread asexual dandelion lineage. *Mol. Ecol.* **25**, 1759–1768.
- Yang, J., Chen, X., Bai, J., Fang, D., Qiu, Y., Jiang, W., Yuan, H., Bian, C., Lu, J., He, S., et al. (2016) The *Sinocyclocheilus* cavefish genome provides insights into cave adaptation. *BMC Biol.* **14**, 1.
- Yu, L., Wang, G.-D., Ruan, J., Chen, Y.-B., Yang, C.-P., Cao, X., Wu, H., Liu, Y.-H., Du, Z.-L., Wang, X.-P., et al. (2016) Genomic analysis of snub-nosed monkeys (*Rhinopithecus*) identifies genes and processes related to high-altitude adaptation. *Nat. Genet.* **48**, 947–952.
- Zhang, H., Lang, Z., and Zhu, J. K. (2018) Dynamics and function of DNA methylation in plants. *Nat. Rev. Mol. Cell Biol.* **19**, 489–506.
- Zhang, L., Li, Q.-J., Li, H.-T., Chen, J., and Li, D.-Z. (2006) Genetic diversity and geographic differentiation in *Tacca chantrieri* (Taccaceae): an autonomous selfing plant with showy floral display. *Ann. Bot.* **98**, 449–457.
- Zhang, Q., Gou, W., Wang, X., Zhang, Y., Ma, J., Zhang, H., Zhang, Y., and Zhang, H. (2016) Genome resequencing identifies unique adaptations of Tibetan chickens to hypoxia and high-dose ultraviolet radiation in high-altitude environments. *Genome Biol. Evol.* **8**, 765–776.
- Zhang, T., Qiao, Q., Novikova, P. Y., Wang, Q., Yue, J., Guan, Y., Ming, S., Liu, T., De, J., Liu, Y., et al. (2019) Genome of *Crucihimalaya himalaica*, a close relative of *Arabidopsis*, shows ecological adaptation to high altitude. *Proc. Natl. Acad. Sci. USA* **116**, 7137–7146.
- Zhang, W., Fan, Z., Han, E., Hou, R., Zhang, L., Galaverni, M., Huang, J., Liu, H., Silva, P., Li, P., et al. (2014) Hypoxia adaptations in the grey wolf (*Canis lupus chanco*) from Qinghai-Tibet Plateau. *PLoS Genet.* **10**, e1004466.