

## Resource Status of Wild Hawthorn in Xinjiang and Research Progress of Genetic Diversity in Hawthorn

Chao Ma<sup>1,3,a,#</sup>, Lingli Meng<sup>1,b,#</sup>, Lingxiao Liu<sup>2,c</sup>, Yaoyao Li<sup>1,d,\*</sup>, Yunguo Liu<sup>1,e,\*</sup>

<sup>1</sup>College of Life Sciences, Linyi University, Linyi, 276005, China;

<sup>2</sup> Linyi Academy of Agricultural Sciences, Linyi, 276012, China;

<sup>3</sup>Center for International Education, Philippine Christian University, Manila, 1004, Philippines.

<sup>a</sup>machaofood@126.com, <sup>b</sup>lingli444@163.com, <sup>c</sup>liulingxiao2002@163.com, <sup>d</sup>liyaoyao678@163.com, <sup>e</sup>yguoliu@163.com

**Abstract.** Wild hawthorn is a rare wild plant resource endemic to the western Tianshan Mountain in Xinjiang, containing three species: *Crataegus songorica*, *C. alorocarpa* and *C. sanguirnta*. However, the area of wild hawthorn forest has shrunken under the disturbance of human activities, genetic diversity has also suffered serious damage. Therefore, it is urgent to carry out the present status investigation of genetic diversity and assess the genetic differentiation of geographic populations of wild hawthorns. This paper summarizes the resource status of wild hawthorn in Xinjiang and research progress of genetic diversity in hawthorn, and developed an approach to screen molecular markers of single nucleotide polymorphism (SNP) and simple sequence repeat (SSR) using Restriction-site Associated DNA sequencing (RAD-seq). SNP and SSR markers will be used to assess genetic diversity, genetic differentiations and genetic relationships of different populations of wild hawthorns.

**Keywords:** Wild hawthorns; Resource status; Genetic diversity; Molecular marker.

### 1. Introduction

Xinjiang has a vast land area, rich in light and heat resources, with huge mountains and vast plains and valleys. Diverse and specialised habitat types have produced a rich and distinctive range of wild fruit trees, the difference in species diversity of wild fruit trees between mountains and plains is very obvious in Xinjiang. For example, there are only 19 species in the Tarim Basin and the Junggar Basin, but there are as many as 80 species of wild fruit trees in the Tianshan Mountains alone[1]. This is because the high-altitude mountains are humid areas in arid areas, which is conducive to the growth and distribution of most wild fruit trees[2]. Among the major mountains in Xinjiang, the Tianshan Mountains have the most abundant species of wild fruit trees, up to 82 species, accounting for 78.4% of the total species of wild fruit trees in Xinjiang. Due to the large amount of precipitation in the western Tianshan Mountains, a large area of natural wild fruit forests is distributed. This area is one of the birthplaces of cultivated fruit trees and a treasure trove of fruit tree resources left by the unique geographical environment since ancient times[3,4]. In particular, Xinjiang Ili and West Tianshan Nature Reserve (Ili, Xinyuan, Huocheng, Gongliu) is one of the special areas of biodiversity in China, and many precious wild fruit trees are distributed in nature reserves. Xinjiang wild fruit trees occupy a special position in China's biodiversity with its unique and important.

### 2. Present Situation of Wild Hawthorn Resources in Western Tianshan Mountains

Hawthorn (*Crataegus* spp.) is a kind of important plants in Maloideae, Rosaceae. There are more than 1000 species in the world, which are widely distributed in Asia, Europe, North America and the north of South America. There are 18 species and 6 varieties of *Crataegus* in China [5]. Hawthorn is a deciduous tree or shrub, with pure white or red flower color, and red, orange, yellow-green, yellow or black fruit, which has high nutritional value and medicinal value [6]. Hawthorn is not only an important fruit tree, but also a good garden ornamental plant and greening

tree species. *C. songorica*, *C. alorocarpa* and *C. sanguirnta* are the main wild hawthorns distributed in the western Tianshan Mountains of Xinjiang[7,8]. *C. songorica* grows at an altitude of 1000 to 1600 meters, is mainly distributed in valleys and dry gravel slopes in Ili, Tacheng, Altai and other places. In addition, *C. songorica* is one of the main tree species that make up the wild fruit forest in Huocheng County and Yining County. It is widely distributed in Daxigou, Xiaoxigou, Dadonggou, Xiaodonggou, Saerbulakegou in Huocheng County, Qulihaigou and Jiliglanggou in Yining County[7]. *C. alorocarpa* also known as yellow-fruited hawthorn, growing at an altitude of 450 to 1900 meters. It is mainly distributed in the mountain valleys and slopes of Altai, Ili, Bole, Tacheng, Manas, Urumqi, and scattered distribution in wild fruit forests in Huocheng County and Yining County[7]. *C. sanguirnta* growing at an altitude of 1200 to 1900 meters, is mainly distributed in the jungle edge and river bank of Altai, Ili and Tacheng. In Xinyuan County, Gongliu County, Tekes County, Zhaosu County, wild fruit forest and mountainous areas are sporadically distributed[7]. Among them, the *C. songorica* has been listed in China Priority Species List, National priority protected species of international significance for biodiversity, and China's third-level protected plants[9].

For a long time, the western region of Tianshan Mountains has been seriously damaged by over-exploitation of land, overgrazing of grassland and deforestation under the interference of human activities, resulting in the destruction of a large area of wild fruit forests, the decline and evolution of ecosystems, and the imbalance of ecological balance. Many companion tree species have been destroyed in an attempt to reap the benefits of wild fruit forests. Over the years, such competition for land between agriculture, forestry and animal husbandry has led to a drastic reduction in the area of wild fruit forests on a wide scale. Wild hawthorn is no exception, and in recent years the area of wild hawthorn forests has been shrinking, and biodiversity has suffered serious damage. Wild hawthorn forest as a natural hawthorn gene pool, due to the reduction of natural resources, excellent genetic resources have also been destroyed. Genetic diversity is a precious natural resource and the basis for human survival. The evolutionary potential and the ability to resist adversity of a species depend on the magnitude of genetic variation. The richer its genetic diversity, the stronger its ability to adapt to environmental changes, the wider its natural distribution range, the stronger its ability to adapt to the environment, and the more advantageous in economic traits such as growth, fecundity and stress resistance. In addition, genetic diversity can also give a species the ability to adapt to environmental changes and resist pathogens. Maximizing the level of intraspecific genetic diversity is a prerequisite for maintaining its genetic potential and sustainable utilization of germplasm resources. Therefore, it is very urgent to carry out the protection of wild hawthorn resources in the western Tianshan Mountains of Xinjiang and carry out related basic research, especially for the investigation of genetic diversity and the study of genetic differentiation of different geographical populations. Strengthening the preservation of germplasm resources, protecting existing species, especially wild rare species and various ecosystems, and protecting the genetic resources of wild relatives of plants is an important task for natural conservation and sustainable development of agriculture and forestry.

### 3. Research status of genetic diversity of hawthorn

Currently, DNA molecular markers and chloroplast DNA-RFLP sequencing technology are widely used in genetic variation, genetic diversity and population genetic structure analysis, resource evaluation, in vitro regeneration, genetic relationship and conservation genetics of hawthorn[10]. In the study of genetic variation, genetic diversity and population genetic structure, Fineschi et al.[11] used chloroplast DNA (cp-DNA) PCR-RFLP and microsatellite markers to analyze the genetic diversity of 21 European hawthorn populations. Albarouki[12] and Ferrazzini et al.[13] used RAPD markers to analyze the genetic diversity of six hawthorn populations within 20,000 square kilometers in northern Italy to evaluate the correlation between geographical distance and locus variation. Beigmohamadi and Rahmani[14] used RAPD markers to analyze the genetic

variation of five Iranian hawthorn populations. Li et al.[15] compared the total DNA extraction methods of hawthorn. Dai et al. [16] used the improved CTAB method and QIAGEN kit method to extract total DNA, established and optimized the ISSR analysis system of hawthorn, and screened 10 ISSR primers suitable for genetic analysis of hawthorn. Subsequently, Dai et al.[17,18] used RAPD and ISSR markers to analyze the DNA polymorphism of 35 hawthorn (*C. pinnatifida* Bge.) resources preserved in the National Fruit Tree Germplasm Shenyang Hawthorn Nursery, including 5 wild hawthorn resources, but did not include Xinjiang wild hawthorn resources. A total of 110 clear bands were amplified by 12 RAPD primers, of which 89 showed polymorphism, with an average of 7.4 polymorphic bands per primer. A total of 110 clear bands were amplified by 13 ISSR primers, of which 94 showed polymorphism. On average, 7.2 polymorphic bands were amplified by each primer, indicating that hawthorn had high genetic diversity. In the study of the genetic relationship of different species and strains of hawthorn. Wu et al.[19,20] analyzed the genetic relationships of Hawthorns (*Crataegus* spp.) in China derived from cp DNA PCR-RFLP, observed the complexity of the origin and evolution of Hawthorn plants. *C. brettschneideri* could be considered a new species, not a variation of *C. pinnatifida*. Interspecies of the genus *Crataegus* had higher cp DNA variations, but no visual bands were detected in different genotypes of *C. pinnatifida*. Four species of Sect. *Sanguineae* (*C. kansuensis*, *C. dahurica*, *C. maximowiczii*, *C. sanguinea*) and *C. brettschneideri* of Sect. *Pinnatifidae* had the same cp DNA PCR-RFLP markers, but *C. altaica* of Sect. *Sanguineae* was distinct with others. Han et al.[21] used ISSR markers to analyze the genetic diversity of 59 hawthorn materials and the genetic relationship of hawthorn plants distributed in northern China. It is considered that *C. brettschneideri* is an independent species of hawthorn and has a close relationship with hawthorn. Feng et al.[22] used RAPD markers to analyze the genetic relationship of different hawthorn varieties, and the results showed that the genetic relationship between seedling hawthorn and other hawthorn varieties was far. Zhang et al. [23] used apple SSR primers to analyze the genetic relationship of hawthorn plants. It was found that SSR primers were universal among different species. Ten pairs of SSR primers suitable for *Crataegus* spp. were screened from 141 pairs of SSR primers of *Malus* spp., and analyzed the genetic relationship of 37 hawthorn germplasm resources of 8 species, the analysis results were consistent with the traditional morphological classification. Han et al. [24] also used apple primers for SSR analysis of different hawthorn varieties. 12 pairs of polymorphic primers were screened from 37 pairs of apple SSR primers, and 36 hawthorn varieties collected from Xinglong County, Hebei Province were analyzed. A total of 52 bands were amplified, including 44 polymorphic bands. Each pair of primers could detect 2-6 alleles. The polymorphic rate was between 50 % -100 %, and the genetic distance ranged from 0.0000 to 0.8440, indicating that there was abundant genetic diversity among varieties. The study also found that the genetic relationship between Wulinghong collected from Xinglong County and other varieties was far from each other, which may be a new species. Lo et al.[25-29] evaluated the genetic relationship between *Crataegus* and *Mespilus* (*Rosaceae*, *Pyreae*), studied the geographical parthenogenesis of hawthorn in the Pacific Northwest, and used microsatellite markers and cp DNA sequencing to analyze the population genetic structure of diploid sexual reproduction and polyploid asexual reproduction hawthorn in the Northwest Pacific. In addition, in other studies, Dai et al.[30,31] studied the *in vitro* regeneration of hawthorn cotyledons. The results showed that the regeneration ability of immature cotyledons, mature cotyledons and cotyledon leaves of hawthorn was significantly different. Among them, cotyledon leaves had higher regeneration ability, and the regeneration frequency of adventitious buds was more than 6 times that of mature cotyledons. Wang et al.[32] evaluated the main cultivated varieties of hawthorn in Shandong, which played a positive guiding role in the development of hawthorn production. The above studies provide important genetic reference data for the protection of local hawthorn resources.

#### 4. Strategy of molecular marker development in Hawthorn

Reduced-representation sequencing is a sequencing technology developed on the basis of second-generation sequencing, which uses enzyme digestion technology, sequence capture technology or other methods to reduce the complexity of species genomes, sequence specific fragments of genomes, and then reflect part of genome sequence information. The most widely used is RAD-seq, it uses restriction enzymes to digest the genome to produce a certain size of fragments, construct a sequencing library, and perform high-throughput sequencing on the RAD markers produced after digestion. Because RAD markers are small fragments of DNA tags near specific restriction sites on a genome-wide scale, representing the sequence characteristics of the entire genome, thousands of SNP and a large number of SSR can be obtained in most organisms by sequencing RAD markers[33-36]. The advantage of this technology is its high throughput, and the number of RAD markers developed by one sequencing is 10 times that of traditional molecular marker development technology. In addition, it is not limited by the genome sequence, and large-scale screening of SNP sites can also be performed for species without reference genomes. RAD-seq has been successfully applied to the development of SNP and SSR markers, genetic diversity analysis, construction of ultra-high density genetic map, QTL mapping of important economic traits and other research fields[37-43]. Therefore, a large number of SNP and SSR markers specific to wild hawthorn in western Tianshan Mountains were obtained by RAD-seq, which not only provided DNA molecular tools for genetic diversity and population structure analysis, genetic relationship identification, genetic map construction, but also laid a foundation for the rational development and utilization of wild plant resources and genetic breeding in Xinjiang.

#### Acknowledgment

This research work was supported by the Key Research and Development Program in Shandong Province (2019YYSP026). We also acknowledge the support of the High-Level Talent Introduction Project of Linyi University (LYDX2018BS032).

#### References

- [1] Yan, G. R., & Xu, Z. Wild fruit tree resources in the Tian Shan mountain area. *Northern Horticulture*, 2001,136(1), 24-27.
- [2] Yang, H. J., Cui, D. F., Xu, Z., et al. Analysis of seed plant composition and resource status of wild fruit forests in the Tianshan Mountains, China. *Journal of Plant Resources and Environment*, 2003,12(2), 39-45.
- [3] Hou, B., & Xu, Z. Research on the resources of rare wild fruit forests in the Ili Valley of the Tianshan Mountains. *Acta Botanica Boreali-Occidentalia Sinica*, 2005,25(11), 2266-2271.
- [4] Hou, B., & Xu, Z. Study on wild fruit trees and related species in Ili, China. *Arid Zone Research*, 2006,23(3), 453-458.
- [5] Zhao, H. Z., & Feng, B. T. (Eds.). *Chinese Fruit Monograph: Hawthorn Volume*. Beijing, China Forestry Press, 1996.
- [6] Lv, D. G., & Li, Z. X. (Eds.). *Specification for Germplasm Resource Description and Data Standard of Hawthorn*. Beijing, China Agriculture Press, 2006.
- [7] Xu, Z. Wild fruit trees and related species resources and distribution in Ili, Xinjiang. In *Proceedings of the First National Symposium on Wild Fruit Tree Resources and Development*, 2004.
- [8] Liu, H., Liao, K., Sun, Q., et al. Study on morphological characteristics of floral organs of three wild hawthorn species in Xinjiang. *Journal of Xinjiang Agricultural University*, 2014,37(4), 293-297.
- [9] Yan, G. R. Research and conservation of wild fruit trees in the Tian Shan of Xinjiang. *Chinese Wild Plant Resources*, 2001,20(4), 13-14.
- [10] Liu, T. J., Zhang, K., Liu, G. Y., et al. Current research and utilization of *Crataegus* species resources in China. *Journal of Southern Agriculture*, 2011,42(8), 847-852.

- [11] Fineschi S, Salvini D, Turchini D, et al. *Crataegus monogyna* Jacq. and *C. laevigata* (Poir.) DC. (Rosaceae, Maloideae) display low level of genetic diversity assessed by chloroplast markers. *Plant Systematics and Evolution*. 2005, 250: 187-196.
- [12] Albarouki E, Peterson A. Molecular and morphological characterization of *Crataegus* L. species (Rosaceae) in southern Syria. *Botanical Journal of the Linnean Society*, 2007, 153: 255-263.
- [13] Ferrazzini D, Monteleone I, Belletti P. Small-scale genetic diversity in oneseed hawthorn (*Crataegus monogyna* Jacq.). *European Journal of Forest Research*, 2008, 127(5):407-414.
- [14] Beigmohamadi M, Rahmani F. Genetic variation in hawthorn (*Crataegus* spp.) using RAPD markers[J]. *African Journal of Biotechnology*, 2011, 10(37): 7131-7135.
- [15] Li, Y. Y., Dai, H. Y., Guo, X. W., et al. A comparison of total DNA extraction methods in hawthorn. *Journal of Fruit Science*, 2007,24(1), 115-118.
- [16] Dai, H. Y. Molecular identification and innovative research on hawthorn germplasm resources (Doctoral dissertation). Shenyang Agricultural University, 2007.
- [17] Dai, H. Y., Guo, X. W., Zhang, Y., et al. Genetic diversity analysis of *Crataegus pinnatifida* Bge. using RAPD and ISSR markers. *Journal of Horticultural Science*, 2008, 35(8), 1117-1124.
- [18] Dai, H. Y., Zhang, Z. H., Zhou, C. S., et al. Establishment and optimization of ISSR analysis system in hawthorn. *Journal of Fruit Science*, 2007, 24(3), 313-318.
- [19] Wu, F. F., Zhang, Z. H., Dai, H. Y., et al. Analysis of genetic relationship of *Crataegus* species using cpDNA PCR-RFLP. *Journal of Shenyang Agricultural University*, 2008,(6), 664-668.
- [20] Wu, F. F. Study on chloroplast DNA genetic diversity of *Crataegus* species (Doctoral dissertation). Shenyang Agricultural University, 2009.
- [21] Han, X. Y., Liang, Y. H., Wang, Y. J., et al. Origin and classification status of *Crataegus oxyacantha* based on ISSR markers. *Journal of Jilin Agricultural University*, 2009, (2), 164-167.
- [22] Feng, H. X., Guo, S. J., Meng, Q. J., et al. RAPD analysis of genetic relationships among different hawthorn varieties. *Journal of Fruit Science*, 2009, 26(5), 729-732.
- [23] Zhang, Y., Dai, H. Y., Zhang, Q. J., et al. Analysis of genetic relationships of *Crataegus* species using apple SSR primers. *Journal of Fruit Science*, 2008, (4), 521-525.
- [24] Han, Z. X., Bi, Z. L., Lu, F. Q., et al. SSR analysis of different *Crataegus* varieties using apple primers. *Hebei Journal of Forestry and Orchard Research*, 2014,29(2), 165-168.
- [25] Lo EYY, Stefanovic S, Dickinson TA. Molecular reappraisal of relationships between *Crataegus* and *Mespilus* (Rosaceae, Pyreae)-two genera or one? *Systematic Botany*, 2007, 32: 596-616.
- [26] Lo EYY, Stefanovic S, Dickinson TA. Population genetic structure of diploid sexual and polyploid apomictic hawthorns (*Crataegus*; Rosaceae) in the Pacific Northwest. *Molecular Ecology*, 2009a, 18, 1145-1160.
- [27] Lo EYY, Stefanovic S, Christensen KI, et al. Evidence for genetic association between East Asian and western North American *Crataegus* L. (Rosaceae) and rapid divergence of the eastern North American lineages based on multiple DNA sequences. *Molecular Phylogenetics And Evolution*, 2009b, 51: 157-168.
- [28] Lo EYY, Stefanovic S, Dickinson TA. Reconstructing reticulation history in a phylogenetic framework and the potential of allopatric speciation driven by polyploidy in an agamic complex in *Crataegus* (Rosaceae). *Evolution*, 2010, 64: 3593-3608.
- [29] Lo EYY, Stefanovic S, Dickinson TA. Geographical parthenogenesis in Pacific Northwest hawthorns (*Crataegus*; Rosaceae). *Botany*, 2013, 91: 107-116.
- [30] Dai H, Zhang Z, Guo X. Adventitious bud regeneration from leaf and cotyledon explants of Chinese hawthorn (*Crataegus pinnatifida* Bge. Var. major NE Br.) *In Vitro Cellular & Developmental Biology-Plant*, 2007, 47(1): 2-8.
- [31] Dai H, Guo X, Zhang Y et al. Application of random amplified polymorphic DNA and inter-simple sequence repeat markers in the genus *Crataegus*. *Annals Of Applied Biology*, 2009, 154: 175-181.
- [32] Wang, G. Q., Huang, Y., & Meng, Q. J. Research on germplasm resources and their evaluation and utilization of *Crataegus* in Shandong. *Seed*, 2009,28(9), 56-58.

- [33] Miller MR, Dunham JP, Amores A, et al. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, 2007, 17(2): 240-248.
- [34] Van Tassell CP, Smith TP, Matukumalli LK, et al. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods*, 2008, 5(3): 247-252.
- [35] Barchi L, Lanteri S, Portis E, et al. Identification of SNP and SSR markers in eggplant using RAD tag sequencing. *BMC Genomics*, 2011, 12: 304.
- [36] Pegadaraju V, Nipper R, Hulke B, et al. De novo sequencing of sunflower genome for SNP discovery using RAD (Restriction site Associated DNA) approach. *BMC Genomics*, 2013, 14(1):556.
- [37] Baird NA, Etter PD, Atwood TS, et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, 2008, 3(10): e3376.
- [38] Hohenlohe PA, Bassham S, Etter PD, et al. Population genomics of parallel adaptation in three spine stickleback using sequenced RAD tags. *PLoS Genetics*, 2010, 6(2): e1000862.
- [39] Emerson KJ, Merz CR, Catchen JM, et al. Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 2010, 107(37): 16196-16200.
- [40] Hohenlohe PA, Amish JS, Catchen MJ, et al. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and west slope cutthroat trout. *Molecular Ecology Resources*, 2011, 11(Suppl. 1): 117-122.
- [41] Pfender WF, Saha MC, Johnson EA, et al. Mapping with RAD (restriction site associated DNA) markers to rapidly identify QTL for stem rust resistance in *Lolium perenne*. *Theoretical and Applied Genetics*, 2011, 122(8): 1467-1480.
- [42] Poland JA, Brown PJ, Sorrells ME, et al. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PloS ONE*, 2012, 7(2): e32253.
- [43] Richards PM, Liu MM, Lowe N, et al. RAD-Seq derived markers flank the shell colour and banding loci of the *Cepaea nemoralis* supergene. *Molecular Ecology*, 2013, 22(11): 3077-3089.