Molecular Analysis of Evolution and Origins of Cultivated Hawthorn (Crataegus spp.) and Related Species in China

Xiao Du¹, Xiao Zhang¹, Haidong Bu¹,², Ticao Zhang³, Yongchun Lao¹ and Wenxuan Dong¹*

¹ College of Horticulture, Shenyang Agricultural University, Shenyang, China, ² Mudanjiang Branch of Heilongjiang Academy of Agricultural Sciences, Mudanjiang, China, ³ College of Chinese Material Medica, Yunnan University of Traditional Chinese Medicine, Kunming, China

Hawthorn is of high economic value owing to its medicinal properties and health benefits. Crataegus is a member of the Rosaceae family; the genus has a complicated taxonomic history, and several theories on its origin have been proposed. In this study, 53 accessions from seven Crataegus taxa native to China and accessions of exotic Crataegus species (two from Europe and one from North America) were analyzed by specific locus amplified fragment sequencing (SLAF-seq). In total, 933,450 single-nucleotide polymorphisms were identified after filtering and used to investigate the species’ genomic evolution. Phylogenetic trees derived from nuclear simple sequence repeats (SSRs) and SLAF-seq data showed the same topology, in which Crataegus maximowiczii and Crataegus sanguineae formed a closely related cluster that was clearly separated from the cluster composed of Crataegus hupehensis, Crataegus pinnatifida, Crataegus pinnatifida var. major, Crataegus bretschneideri and Crataegus scabrifolia. Phylogenetic and structure analysis indicated that the seven Chinese Crataegus taxa had two separate speciation events. Plants that evolved the southwestern route shared the genepool with the European species, whereas plants along the northeastern route shared the genepool with the North American species. TreeMix genetic analysis revealed that C. bretschneideri may have a hybrid origin. This study provides valuable information on the origins of Chinese Crataegus and suggests an evolutionary model for the main Crataegus species that native to China.

Keywords: hawthorn, Crataegus, SLAF-seq, nSSR marker, molecular evolution

INTRODUCTION

The genus Crataegus (hawthorn), a member of the Rosaceae family, ranges from small shrubs to trees distributed in Eurasia and America (Phipps et al., 1990). Hawthorns are among the most economically important plant species in China, owing to their pleasant flavor, attractive color, and nutrient-rich fruit (Xu et al., 2016). Hawthorn use in preventive medicine dates to the late 1800s. Hawthorns contain biologically active compounds, such as flavonoids, phenols, and oligomeric procyanidins, which have therapeutic benefits (Dickinson et al., 2014; Dahmer and Scott, 2018). Previous laboratory tests and clinical trials have demonstrated the efficacy of hawthorn in the treatment and prevention of cardiovascular disease (Edwards et al., 2012).
On the basis of cladistics analyses of morphological data, Phipps (1990) suggested southwest China and Mexico were ancestral areas for the genus and that trans-Beringian migration of Asian and American \textit{Crataegus} had occurred. \textit{Crataegus} was postulated to have migrated westward from southwest China to Europe, and eastward from East Asia to North America. However, other authors hold conflicting opinions. Evans and Campbell (2002) treated \textit{Crataegus} as subtribe Pyrinae, and based on molecular and non-molecular characters, suggested that \textit{Crataegus} originated in North America. Lo et al. (2009) used sequences for the internal transcribed spacer region, chloroplast DNA regions, LEAFY intron2 to infer relationships among species from eastern Asia, western North America, eastern North America, and Europe; their findings indicated that eastern North America and Europe are probably the most recent common areas of origin for \textit{Crataegus}.

China is the center of \textit{Crataegus} cultivation, and the place of origin of both cultivated and some wild \textit{Crataegus} species. Based on morphological characters, some researchers have proposed that 18 species and six varieties of \textit{Crataegus} are widely distributed across China (Zhao and Feng, 1996; Xin and Zhang, 1997), other authors recognize 20 species of Chinese \textit{Crataegus}, with seven varieties (Dong and Li, 2015). Among these species, \textit{C. hupehensis}, \textit{C. pinnatifida} var. major, \textit{C. bretschneideri}, and \textit{C. scabrifolia} are cultivated. Previous efforts to understand the phylogenetic and biogeographic history of \textit{Crataegus} have included representatives of Chinese \textit{Crataegus}. Phipps (1990) suggested that \textit{C. scabrifolia} evolved into European \textit{Crataegus} and other Chinese \textit{Crataegus} species (\textit{C. pinnatifida}, \textit{C. hupehensis}, and \textit{C. sanguineae}). Based on sequence data from 14 plastid loci, Zarrei et al. (2015) treated \textit{C. maximowiczii} as section \textit{Sanguineae} and suggested that the origin of the section involved east-to-west trans-Beringian migration from western North America into eastern Asia. Lo et al. (2009) suggested that ancestors of \textit{C. hupehensis}, \textit{C. songorica}, and \textit{C. pinnatifida} dispersed from Europe into Asia. However, a consensus is lacking on the migration direction of Chinese and European \textit{Crataegus}.

Recent studies have examined intraspecific (Wu et al., 2008; Zhang et al., 2008; Sheng et al., 2017) and interspecific relationships (Su et al., 2015; Ma and Lu, 2016) of Chinese \textit{Crataegus}. Previous investigations have used morphological data analyses and limited molecular data, but no study has explored the origin and evolution of cultivated \textit{Crataegus} and related species that are native to China at the genomic level. Plant DNA contains abundant genetic information, and an increasing number of researchers have explored interspecific relationships and diversification of plants using molecular marker information. Molecular markers are used to determine genetic relationships within plant populations with almost 100% reliability (Güney et al., 2018), and random amplified repeats (Erfani-Moghadam et al., 2016; Zarrei et al., 2017), inter-simple sequence repeats (Sheng et al., 2017; Emami et al., 2018), and SSRs (Lo et al., 2009; Khiari et al., 2015; Brown et al., 2016) have been widely used for genetic characterization of \textit{Crataegus}, and to analyze genetic diversity among and within accessions of \textit{Crataegus}. Of these molecular markers types, SSR markers have attained considerable popularity in genetic research because they are highly polymorphic, convenient, and codominant. Based on double barcode genotyping systems and sequencing, specific-locus amplified fragment sequencing (SLAF-seq) was developed for \textit{de novo} single nucleotide polymorphism (SNP) discovery and genotyping using reduced representation library sequencing. SLAF-seq is a high-throughput, high-accuracy, and low-cost method with short cycles, and has been used for molecular breeding and analysis of germplasm resources (Huang et al., 2016). SLAF-seq does not depend on a reference genome sequence and is particularly useful for species the lack an assembled reference genome (Sun et al., 2013), because it is possible to perform polymorphism analysis and develop molecular markers directly from the sequence data provided by SLAF-seq (Zheng et al., 2016). Given these advantages, SLAF-seq has been used for rapid mass discovery of SNP markers for polymorphism analysis, system evolution, and germplasm resource identification (Chen et al., 2013; Zhang et al., 2013; Xu et al., 2015).

In the present study, we used SLAF-seq to gain insight into evolutionary relationships among seven \textit{Crataegus} taxa native to China, namely \textit{C. maximowiczii}, \textit{C. sanguineae}, \textit{C. hupehensis}, \textit{C. pinnatifida}, \textit{C. pinnatifida} var. major, \textit{C. bretschneideri}, and \textit{C. scabrifolia}. These taxa are widely distributed in China and cover most of the different climatic regions in the country. The sampled taxa include the four taxa cultivated in China (\textit{C. hupehensis}, \textit{C. pinnatifida} var. major, \textit{C. bretschneideri}, and \textit{C. scabrifolia}) and three species distributed in close proximity to cultivated \textit{Crataegus}. \textit{C. maximowiczii} and \textit{C. sanguineae} are typically distributed in northeastern China. Of the cultivated taxa, \textit{C. pinnatifida} var. major is endemic to China and has the longest history in cultivation, and \textit{C. scabrifolia} is considered to be the ancestral \textit{Crataegus} species. \textit{C. pinnatifida} is a species that is widespread throughout China. The seven taxa were selected by other researchers as representatives of Chinese \textit{Crataegus} in previous phylogenetic studies (e.g., \textit{C. pinnatifida}, \textit{C. hupehensis}, and \textit{C. sanguineae}: Phipps, 1990; \textit{C. hupehensis}, \textit{C. sanguineae}, \textit{C. maximowiczii}, and \textit{C. pinnatifida}: Lo et al., 2009; \textit{C. maximowiczii} and \textit{C. sanguineae}: Zarrei et al., 2015). Previous studies provided useful information that partially resolved the phylogenetic history of cultivated \textit{Crataegus} in China. However, the interspecific relationships and evolution of cultivated \textit{Crataegus} in China remain unclear. In the present study, SLAF-seq was used to analyze phylogenetic relationships among 53 accessions of cultivated \textit{Crataegus} and three related species in China based on SSRs and SNPs.

**MATERIALS AND METHODS**

**Plant Material**

In total, 53 accessions of Chinese \textit{Crataegus} were sampled, which consisted of \textit{C. maximowiczii} (5 accessions), \textit{C. sanguineae} (4 accessions), \textit{C. hupehensis} (6 accessions), \textit{C. pinnatifida} (14 accessions), \textit{C. pinnatifida} var. major (12 accessions), \textit{C. bretschneideri} (10 accessions) and \textit{C. scabrifolia} (2 accessions). The outgroup comprised accessions of three exotic taxa collected from abroad (\textit{C. monogyna} and \textit{C. laevigata} from Europe, and
Du et al. Evolution of Chinese Originated Hawthorn C. cruss-galli from North America). The 53 Chinese accessions (Table 1) were broadly distributed across a range of geographic and climatic conditions, and thus were considered to be representative of Chinese hawthorn diversity and to encompass all possible introductory sources in China.

All sampled trees were maintained at the National Hawthorn Germplasm Repository at Shenyang Agricultural University, China (Supplementary Table S1).

**Morphological Measurements**

Fresh leaves and fruits were sampled for morphological measurements when mature (from September to early October). Twenty-five leaves and fruits per accession were sampled. The measurement methods followed the technical code for evaluation crop germplasm resources for hawthorn (Crataegus L.) (Dong, 2013). Twenty-one characters, including leaf shape, leaf blade lobes, leaf blade margin, and leaf color, were measured. The qualitative trait characteristics and classifications are summarized in Supplementary Table S2.

**DNA Extraction and PCR Amplification**

Samples of young leaves were collected, labeled, frozen with liquid nitrogen, and stored at −80°C until DNA extraction. One gram of frozen leaf material was ground for genomic DNA extraction using cetyl-tri-methylammonium bromide in accordance with the protocol of Doyle and Doyle (1990). The DNA quality was checked using a Nanodrop-2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, United States).

Fifty-six samples were analyzed using nuclear SSR (nSSR) markers. PCR amplification was performed in 20 µl reaction mixture consisting of 1 µl template DNA (20–30 ng), 7 µl of 2× Es MasterMix buffer (CWBL, Beijing, China), 2 µl primers (20 ng/µl), and 10 µl sterile nuclease-free distilled H2O. The PCR protocol was as follows: initial denaturation at 94°C for 5 min; 30 cycles at 94°C for 1 min, 56–59°C for 1 min, and 72°C for 1 min; and a final extension at 72°C for 7 min. PCR amplification was carried out in a Thermal Cycler (Applied Biosystems, Foster City, CA, United States). Genotyping was performed using 20 SSR primers (Supplementary Table S3) that were arbitrarily selected from hawthorn transcriptional data (Xu et al., 2016).

The PCR products were separated on a 6% polyacrylamide gel in 0.5× TBE buffer. After electrophoresis, the gel was stained as previously described (Cairns and Murray, 1994). The resulting bands were recorded in a presence–absence matrix in which values represented the presence (1) or absence (0) of a band. Genetic similarities based on Jaccard's coefficients were

**TABLE 1** Details of geographic and sampling information for Crataegus investigated in this study.

| Taxon ID | Biogeographic regions | Taxon ID | Biogeographic regions |
|----------|-----------------------|----------|-----------------------|
| FLH      | Liaoning, China       | C. pinnatifida | NMGS LH           |
| LNDG     | Liaoning, China       |          |          | Inner Mongolia, China |
| 82015    | Jilin, China          | C. pinnatifida var. major | HLJM DFS LH |
| ZF1H     | Jilin, China          |          |          | Heilongjiang, China |
| JF1H     | Jilin, China          | C. maximowiczii | MSZ1H          |
| CH       | Jilin, China          |          |          | Heilongjiang, China |
| ZF2H     | Jilin, China          | C. sanguineae | LNSZ1H          |
| 555      | Jilin, China          |          |          | Liaoning, China    |
| JF2H     | Jilin, China          | C. monogyna | LNSZ2H          |
| FSZ1H    | Liaoning, China       |          |          | Liaoning, China    |
| HQ       | Shandong, China       |          |          | Liaoning, China    |
| XHMZ     | Shandong, China       | C. scabridula | YNSZ1H          |
| MYDJK    | Shandong, China       |          |          | Yunnan, China      |
| BRM      | Shandong, China       | C. laevigata | YNSZ2H          |
| CK       | Shandong, China       |          |          | Yunnan, China      |
| DMQ      | Shandong, China       | C. maximowiczii | DJSZ1H      |
| XLZ2R    | Hebei, China          |          |          | Britain            |
| JD1H     | Beijing, China        | C. sanguineae | LNSZ1H          |
| DW       | Jilin, China          |          |          | Liaoning, China    |
| QJK      | Liaoning, China       | C. sanguineae | LNSZ2H          |
| QYMP     | Liaoning, China       |          |          | Liaoning, China    |
| KFZ      | Liaoning, China       | C. sanguineae | LNSZ3H          |
| C. hupehensis |          |          |          | Liaoning, China    |
| HBSZ1H   | Hubei, China          |          |          | Liaoning, China    |
| HBSZ2H   | Hubei, China          | C. scabrifolia | YNSZ1H          |
| HBSZ3H   | Hubei, China          |          |          | Yunnan, China      |
| MHL      | Shandong, China       | C. monogyna | YNSZ2H          |
| XP2M     | Shandong, China       |          |          | Yunnan, China      |
| TASS     | Shandong, China       | C. cruss-galli | YNSZ3H          |

Frontiers in Plant Science | www.frontiersin.org 3 April 2019 | Volume 10 | Article 443
calculated using the SIMQUAL program with the Numerical Taxonomy Multivariate Analysis System (NTSYS-PC) v.2.0. The matrix of genetic similarity coefficients was used to generate a dendrogram with the unweighted pair group method with arithmetic mean (UPGMA) method implemented in NTSYS-PC v.2.0 (Rohlf, 1997).

**SLAF-Seq and Analysis**

Library preparation and sequencing of SLAF-markers from genomic DNA were performed by the Biomarker Technology Company (Beijing, China). Owing to the lack of a reference genome, the *Malus* genome 1 (GCA_002114115.1) was used for enzyme cutting site prediction. The following criteria were applied to determine the enzyme cutting scheme: (i) the smallest fragments in the repeat sequence, (ii) the fragments are distributed even in the genome, (iii) the length of fragments must be consistent with the scheme, and (iv) the number of SLAFs must meet expectations. After filtering, we used *RsaI* + *HaeIII* (NEB, Ipswich, MA, United States) for digestion, and the predicted target fragment length was 314–344 bp. Sequence data for the control, *Oryza sativa* subsp. *japonica*, was obtained from the Rice Annotation Project Database 2 for quality control and to ensure the effectiveness of the enzyme cutting scheme. In accordance with the enzyme cutting scheme, the restriction enzymes digested the qualified DNA to obtain the SLAFs; then, the adenine nucleotide (A) was added to the 3′ end of the SLAFs, the dual-index adapter was ligated, and the extract was purified and submitted for pair-end sequencing using an Illumina HiSeq™ 2500 platform (Illumina, Inc., San Diego, CA, United States). Dual-indexing was used to identify the original data from the sequence and obtain reads for the 56 samples. SNPs were called using the Burrows-Wheeler Alignment tool (BWA; Li and Durbin, 2009). Population polymorphism analysis was conducted using high-quality SNPs. Reads with clear index information were clustered based on sequence similarity.

**Phylogenetic Inference and Divergence Time Estimation**

A phylogenetic tree was constructed based on the SNPs using maximum likelihood (ML) analyses (Kobert et al., 2016). The ML analyses were performed using RaxML v.8.2.0 (Stamatakis, 2014). Principal component analysis (PCA) was performed using EIGENSOFT (Price et al., 2006). Population structure was investigated using STRUCTURE (Hubisz et al., 2009). To determine the most likely number of ancestral kinships (K) in the population, STRUCTURE was run 20 times for each K value from 3 to 10. We calculated ΔK, which indicates the change in likelihood of different numbers of clusters, and determined the cluster number with the highest ΔK value, which represents the most likely number of clusters in the population. We inferred admixture graphs using TreeMix v.1.12 (Pickrell and Pritchard, 2012). Divergence time estimation was conducted using PAML (Stadler and Yang, 2013) based on the SNPs data and fossil age. Fossils for calibration were selected from the TimeTree database 3. The divergence times between species from the TimeTree database were extracted from all peer-reviewed publications in molecular evolution and phylogenetic that reported estimates of the time of divergence among species. A hierarchical average linkage method was used to estimate divergence times (Ts) of clade pairs to build a Super Time tree, together with a procedure for testing and updating topological partitions to ensure the highest degree of consistency with individual time trees in every study (Hedges et al., 2015).

The split time between *C. pinnatifida*-1541SLH and *C. monogyna*-DZ1H was searched in the database. The estimated divergence data (14.7 Ma) was set as the ancestral node time, which was then used to estimate species divergence times.

**RESULTS**

**Plant Morphological Descriptions**

The leaf and fruit characteristics recorded are shown in **Supplementary Table S2**. The most obvious differences between species were in the leaf blade lobes and fruit size. Based on morphological characteristics (Figure 1), the seven Chinese *Crataegus* taxa were separated into the following four groups: (i) plants with shallowly dissected leaves and small fruit, which contained *C. maximowiczii* and *C. sanguineae*; (ii) plants with moderately lobed leaves and medium-sized fruit, which comprised *C. pinnatifida*; (iii) plants with deeply dissected leaves and large fruit, which consisted of *C. pinnatifida* var. *major* and *C. bretschneideri*; and (iv) plants with non-dissected or shallowly lobed leaves and large fruit, which comprised *C. hupehensis* and *C. scabrifolia*. The *C. maximowiczii* is the only species with have pubescence on the leaf.

**SSR Analysis**

The genetic relationships among the *Crataegus* taxa were determined based on a cluster analysis that used loci generated by the nSSR markers. Twenty microsatellite primer pairs were used for the nSSR analysis. Similarity values ranged from 0.49 to 0.96. Genotyping data for 48 alleles were obtained with the 20 primer pairs and used to construct a dendrogram with the UPGMA method (**Supplementary Figure S1**). The 56 accessions were well separated using the SSR markers.

The nSSR phylogenetic tree showed that *C. monogyna* and *C. laevigata*, which are from Europe, were entirely distinct from the other *Crataegus* species, and *C. crus-galli*, which is from North America, formed a sister group with *Crataegus* from northeastern China. The seven Chinese *Crataegus* taxa were grouped into two major clusters; one cluster contained *C. maximowiczii* and *C. sanguineae*, whereas the other contained *C. hupehensis*, *C. pinnatifida*, *C. pinnatifida* var. *major*, *C. bretschneideri*, and *C. scabrifolia*. These two clusters were separated at a similarity value of approximately 0.65. Four accessions (ZWSLH, GSSZ, RR5H, and RR3H) were considered to belong to *C. pinnatifida* based on their

---

1https://www.ncbi.nlm.nih.gov/assembly/
2http://rapdb.dna.affrc.go.jp/
3http://timetree.org
accessions of the southwestern Chinese species *C. scabrifolia* were indicated to be an early divergence in the cluster. The central *C. pinnatifida* var. *northeastern China* (*C. hupehensis*, *C. pinnatifida*, and *C. sanguineae*) ranged from southwestern to central, northern, eastern, and northeastern China (*C. maximowiczii*). The central accessions; cluster B was a widely distributed group, with species *C. sanguineae* and *C. pinnatifida*. Figure 2A) showed strong evidence for two clusters: cluster A was a northeastern group (*C. maximowiczii*, *C. hupehensis*, *C. pinnatifida*, *C. pinnatifida* var. *major*, and *C. bretschneideri*). To further understand the evolutionary history of Chinese *Crataegus*, we used a Bayesian clustering algorithm with admixed models (Hubisz et al., 2009) to estimate the ancestral proportions for each sample (Figure 2C). The AK analysis revealed that five populations (*K* = 5) represented the best model for these 56 samples (Supplementary Figure S2). When *K* = 4, the *C. bretschneideri* genepool was indicated to be derived from *C. maximowiczii* (yellow) and *C. pinnatifida* (red); when *K* = 5 to 8, the *C. bretschneideri* genepool predominantly (more than half) was suggested to be derived from *C. maximowiczii* and *C. pinnatifida*. The STRUCTURE analysis (Figure 2C) showed that the southwestern Chinese species *C. scabrifolia* harbored the genepool from the European species *C. monogyna* and *C. laevigata*, and showed evidence for introgression from *C. scabrifolia* and the two European species. The genepool of the North American species *C. crass-galli* was similar to that of the northeastern Chinese species *C. sanguineae*. Similar to the SSR dendrogram, the genepool of the accessions ZWSLH, GSSZ, RR5H, and RR3H shared with *C. maximowiczii* and *C. sanguineae*; these findings demonstrated that four accessions did not belong to *C. pinnatifida*. When *K* = 4 to 8, the accessions RR3H and RR5H shared the genepool of *C. pinnatifida* and *C. bretschneideri*. To explore the true identity of these four accessions, we considered them to be “unknown” accessions in the following analysis.

To understand the history of divergence and admixtures, we applied TreeMix to the 10 groups, six species and one variety,
and the groups unknown1 (GSSZ and ZWSLH) and unknown2 (RR3H and RR5H). *C. crass-galli*, *C. monogyna*, and *C. laevigata* were used as the outgroup taxa. TreeMix uses genome-wide allele frequency data and a Gaussian approximation to genetic drift to analyze whether the migration events between species is a pattern of population splits and mixtures in multiple populations. The result contains the population splits and gene flow between species. The arrow in the figure corresponds to the migration events, the darkness of the arrow color indicates the migration edge weight. In the TreeMix result (Figure 3), introgression occurred among *C. hupehensis*, *C. pinnatifida var. major*, and *C. pinnatifida*, which indicated that extensive gene flow had occurred in central, northern, and northeastern China. The *C. maximowiczii*, *C. sanguineae*, and two unknown groups formed a cluster that differed markedly from the other groups. Between the two clusters, gene flow from *C. maximowiczii* to *C. bretschneideri* was observed, with an arrow weight of 0.43, which indicated that *C. bretschneideri* is the result of admixture between *C. maximowiczii* and *C. pinnatifida*, with about 43% of the genome derived from *C. maximowiczii*. We did not observe gene flow between the four unknown accessions and other species. Based on the STRUCTURE and TreeMix results, we propose that *C. bretschneideri* may have been of hybrid origin, and shared the genepool with *C. pinnatifida* and *C. maximowiczii*. 

### Divergence Time Estimation

Based on the SLAF-seq data and fossil calibration, the major diversification events of Chinese *Crataegus* species were estimated to have occurred in the late Miocene and Pliocene (Figure 4). The split between clusters A and B was estimated to be ~10.8 Ma. The earliest Chinese *Crataegus* to diverge was the southwestern species *C. scabrifolia*, which split at ~8.81 Ma, followed by the central species *C. hupehensis*, which diverged at ~6.85 Ma. Subsequent diversification events occurred during the Pliocene. The northeastern species *C. sanguineae* and *C. maximowiczii* split at ~4.12 Ma; *C. pinnatifida var. major*, which is distributed in northern, eastern, and northeastern China, diverged at ~3.25 Ma; the northern and northeastern species *C. pinnatifida* was diverged at ~1.23 Ma; and the northeastern species *C. bretschneideri* was the most recent species to diverge and arose at ~0.22 Ma.

### DISCUSSION

#### Phylogenetic Hypothesis for Chinese *Crataegus*

We presented a robust phylogenetic construction for seven Chinese *Crataegus* species based on SLAF-seq data. The ML
phylogenetic tree showed evidence for two main clusters. Cluster A includes *C. maximowiczii* and *C. sanguinea*, whereas cluster B comprised *C. huphensis*, *C. pinnatifida*, *C. pinnatifida* var. *major*, *C. bretschneideri*, and *C. scabrifolia*. From a geographically perspective, cluster A consists of the species in northeastern China, whereas cluster B includes the widespread species that extend from southwestern to northeastern China. Combined the molecular data and the morphological characters, fruit size is an important character for *Crataegus* classification. Consistent with the morphological classification, molecular data showed strong support for the grouping of the two small-fruited species, namely *C. maximowiczii* and *C. sanguinea*. These two species are closely related, and the only differentiating character is pubescence on the leaf of *C. maximowiczii*. In addition to fruit size, leaf dissection is also an important character for discrimination of *Crataegus* species. *C. sanguinea* and *C. maximowiczii* are sympatric with *C. bretschneideri*, *C. pinnatifida*, and *C. pinnatifida* var. *major*, which are distributed in northeastern China. These five species from the
FIGURE 3 | TreeMix analysis of 53 hawthorn samples divided into 9 groups. With C. cruss-galli, C. monogyna and C. laevigata serving as the outgroup population, the arrow corresponds to the direction of migration.

northeast were divided into two groups, of which the group that contained C. bretschneideri, C. pinnatifida, and C. pinnatifida var. major was more closely related to the southwestern species. The STRUCTURE analysis showed that C. scabridolia, which indicated to be the earliest species to diverge in cluster A, belonged to the same lineage as the European species C. laevigata and the North American species C. cruss-galli. When K = 4 to 7, all C. scabridolia accessions clustered with C. laevigata and C. cruss-galli.

There are two contrasting perspectives on the migration direction of Chinese and European Crataegus. Based on the nSSR dendrogram and STRUCTURE analysis (Figure 2C), we suggest that the seven Chinese hawthorn species may have experienced two different speciation events. The first speciation event occurred in northeastern China, the northeastern species shared the genepool with the North American species. The second speciation event originated in southwestern China and then progressed northward and eastward to northeastern China. The taxa in southwestern route shared genepool with Europe Crataegus. The southwestern routed is consistent with Wang's (1992) description of the plants of other genus and families stretched from the Hengduan Mountains toward the northeast through the Qinling Range, the eastern fringe of the Loess Plateau including the Taihang Range, the Yinshan Range, the Changbai Mountains, and the Xiao Hinggan Mountains of Siberia or the adjacent regions.

Population structure is the result of both present and historical processes, and many factors may change the geographical distributions of plant species (Comes and Kadereit, 1998). Similar to the Quaternary climatic events that caused vegetation changes between southern and northern China (Sun et al., 2000), geological events are critical to the formation and development of the regional flora (Zhe and Jian, 2017; Zhe et al., 2017). In the late Miocene, interspecific divergence events occurred within Chinese Crataegus species. Since the late Pliocene, numerous intraspecific differentiation events have occurred. According to previous studies, the third intense uplift of the Qinghai–Tibet Plateau (QTP) and formation of the Hengduan Mountains began at this time (Li and Fang, 1998, 1999; Shi, 1998). The QTP uplift is one of the most important geological events of the Cenozoic; it changed the geography and climate in Asia and led to the Asian monsoon system (Li, 2006; Liu and Dong, 2013). The Asian monsoon system resulted in uneven distribution of precipitation in Yunnan Province and caused the migration of plants (Jacques et al., 2011; Su et al., 2013). Consequently, these events also may have caused the Crataegus species in southwestern China to migrate toward the northeast.

In this study, we observed that four accessions were misidentified based on morphological characteristics.
FIGURE 4 | Time-calibrated phylogeny inferred from SLAF-seq data in BEAST V.1.7.5 using node age based on fossil data. Blue arrows designate calibrated node (14.7 Ma). The blue words represent the fossils for calibration. The dots represents the split time.
The molecular data support the conclusion that these accessions may belong to species other than *C. pinnatifida*; however, additional research is needed to elucidate the true identity of these accessions.

**Hybridization of Chinese Crataegus Species**

Hybridization is recognized to be an important driving force in plant evolution (Mallet, 2007; Paun et al., 2009) that creates new species, or ecotypes, and results in reticulate evolution (Whitham et al., 1994; Linder and Rieseberg, 2004). Changes in geographical distribution provide opportunities for speciation through hybridization. Hybrids are commonly found in regions where different species overlap. Where species’ ranges overlap, hawthorns show introgressive hybridization, which results in a variety of morphological variants (Talent and Dickinson, 2007). Radford et al. (1968) and Phipps (2015) posited that hybridization is a potential explanatory factor for speciation in *Crataegus*. Furthermore, molecular data have provided evidence for hybridization in *Crataegus* (Lo et al., 2010a, b). *C. bretscheideri* is morphologically very similar to *C. pinnatifida*, and some researchers consider the former to be a variant of the latter species (Dai, 2007). Based on peroxidase isozymograms, Guo and Jiao (1995) suggested that *C. bretscheideri* is closely related to *C. pinnatifida*. On the basis of the present results, we propose that *C. bretscheideri* has a hybrid origin. The STRUCTURE analysis (Figure 2C) indicated that *bretscheideri* shared a genepool with *C. pinnatifida* and *C. maximowiczii*, when *K* = 4, the genepool of *bretscheideri* was derived from *C. maximowiczii* (yellow) and *C. pinnatifida* (red); when *K* = 5 to 8, the *bretscheideri* genepool dominated (more than half) was derived from *C. maximowiczii* and *C. pinnatifida*. The TreeMix results (Figure 3) indicated that gene flow had occurred from *C. maximowiczii* to *bretscheideri*. *bretscheideri* is the most recently divergent species, and *C. pinnatifida* was diverged before *bretscheideri*. Geographically, *bretscheideri* is distributed at the border of *C. maximowiczii* and *C. pinnatifida*. Based on these results, we hypothesize that *bretscheideri* arose through hybridization between *C. pinnatifida* and *C. maximowiczii*. The northern species migrated northeast, and hybridization with native species gave rise to new species.

**CONCLUSION**

This study is the first to use SLAF-seq to investigate the evolution and phylogenetic relationships of Chinese *Crataegus*. We hypothesize that the seven *Crataegus* species analyzed in this study experienced two speciation events. The southwestern species, *C. scabrifolia*, was indicated to be the earliest-diverging Chinese species, and shares a genepool with European *Crataegus* species. The northeastern *Crataegus* species share a genepool with the North American species *C. crus-galli*. Overall, the present results provide valuable information on the origin of *Crataegus* in China.

**AUTHOR CONTRIBUTIONS**

XD and WD conceived this project and designed the work. XD, XZ, HB, and YZ performed the research. XD and TZ analyzed the data. XD, WD, and XZ wrote the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

**FUNDING**

This work was supported by the Program (project) of “The Conservation and Utilization of Crop Germplasm Resource-Hawthorn (2015–2017).”

**ACKNOWLEDGMENTS**

We thank Dr. Zhang Qijing and Dr. Hou Yali for assisting with the experiments and commenting on the manuscript.

**SUPPLEMENTARY MATERIAL**

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.00443/full#supplementary-material
Du et al.  

Evolution of Chinese Originated Hawthorn

The Qinghai - Tibet Plateau Uplifting and Environmental Evolution

Li, J., and Fang, X. (1999). Uplift of the Tibetan Plateau and environmental changes. *Am. J. Bot.* 86, 227–234.

Jacques, F. M. B., Guo, S. X., Su, T., Xing, Y. W., and Huang, Y. J. (2011). *Molecular Phylogenetics and Evolution.*

Hubisz, M. J., Falush, D., Stephens, M., and Pritchard, J. K. (2009). Inferring weak selection using second-generation sequencing data. *PloS Genet.* 5, e1003776.

Huang, H. R., Wu, W., Zhang, J. X., Wang, L. J., Yuan, Y. M., and Ge, T. J., and Jiao, P. J. (1995). Hawthorn (Crataegus) resources in China.

Huang, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 25, 1754–1760.

Emami, A., Shabanian, N., Rahmani, M. S., Khadivi, A., and Mohammad-Panah, N. (2018). Genetic characterization of the *Crataegus* genus: implications for in situ conservation. *Sci. Hort.* 231, 56–65.

Evans, R. C., and Campbell, C. S. (2002). The origin of the apple subfamily Rosoideae: Article Collection of Academician Li Ji-Jun. *Can. J. Bot.* 80, 507–518.

Evans, R. C., Forest, F., Fay, M. F., and Chae, M. W. (2009). Hybrid speciation in angiosperms: parental divergence drives ploidy. *New Phytol.* 182, 507–518.

Evans, R. C., Forest, F., Fay, M. F., and Chae, M. W. (2009). Hybrid speciation in angiosperms: parental divergence drives ploidy. *New Phytol.* 182, 507–518.

Evans, R. C., Forest, F., Fay, M. F., and Chae, M. W. (2009). Hybrid speciation in angiosperms: parental divergence drives ploidy. *New Phytol.* 182, 507–518.

Evans, R. C., Forest, F., Fay, M. F., and Chae, M. W. (2009). Hybrid speciation in angiosperms: parental divergence drives ploidy. *New Phytol.* 182, 507–518.
Du et al. Evolution of Chinese Originated Hawthorn

Wu, F. F., Zhang, Z. H., Dai, H. Y., Zhang, Y., and Zhang, L. L. (2008). Genetic relationship of some Crataegus spp. (Hawthorn) revealed by chloroplast DNA PCR-RFLP. J. Biotechnol. 136:S103. doi: 10.1016/j.jbiotec.2008.07.235

Xin, X., and Zhang, Y. M. (1997). Chinese Hawthorn Germplasm Resources and Utilization. Beijing: China Agricultural Press.

Xu, X., Xu, R., Zhu, B., Yu, T., Qu, W., Lu, L., et al. (2015). A high density genetic map of cucumber derived from specific length amplified fragment sequencing (SLAF-seq). Front. Plant Sci. 5, 768. doi: 10.3389/fpls.2014.00768

Xu, J. Y., Zhao, Y. H., Zhang, X., Zhang, L. J., and Dong, W. X. (2016). Transcriptome Analysis and ultrastructure observation reveal that hawthorn fruit softening is due to cellulose/hemicellulose degradation. Front. Plant Sci. 7:1524. doi: 10.3389/fpls.2016.01524

Zarei, A., Erfanimoghadam, J., and Mozaffari, M. (2017). Phylogenetic analysis among some pome fruit trees of Rosaceae family using RAPD markers. Biotechnol. Biotechnol. Equip. 31:10. doi: 10.1080/13102818.2016.1276414

Zarei, M., Talent, N., Kuzmina, M., Lee, J., Lund, J., Shipley, P. R., et al. (2015). DNA barcodes from four loci provide poor resolution of taxonomic groups in the genus Crataegus. AoB Plants 7:v045. doi: 10.1093/aobpla/plv045

Zhang, Y., Dai, H. Y., Zhang, Q. J., Li, H., and Zhang, Z. H. (2008). Assessment of genetic relationship in Crataegus genus by the apple SSR primers. J. Fruit Sci. 25, 521–525.

Zheng, W. X., Li, Z. Q., Zhao, J. M., Zhang, Y. Z., Wang, C. H., Lu, X. C., et al. (2016). Study of the long-distance migration of small brown plant hoppers Laodelphax striatellus in China using next-generation sequencing. Pest Manag. Sci. 72, 298–305. doi: 10.1002/ps.3992

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Du, Zhang, Bu, Zhang, Lao and Dong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.