Genome Sequencing of the Endangered *Kingdonia uniflora* (Circaeasteraceae, Ranunculales) Reveals Potential Mechanisms of Evolutionary Specialization

Yanxia Sun, Tao Deng, Aidi Zhang, ..., Xiujun Zhang, Hang Sun, Hengchang Wang

*zhangxj@wbgcas.cn* (X.Z.)
*sunhang@mail.kib.ac.cn* (H.S.)
*hcwang@wbgcas.cn* (H.W.)

**HIGHLIGHTS**

- Provides a high-quality assembly for *K. uniflora* genome
- *K. uniflora* shows overrepresentation in gene families associated with DNA repair
- *K. uniflora* shows underrepresentation in gene families associated with stress response
- The current endangered status of *K. uniflora* is related with adaptive degeneration

Sun et al., iScience 23, 101124
May 22, 2020 © 2020 The Author(s).
https://doi.org/10.1016/j.isci.2020.101124
Genome Sequencing of the Endangered Kingdonia uniflora (Circaeasteraceae, Ranunculales) Reveals Potential Mechanisms of Evolutionary Specialization

Yanxia Sun,1,2,9,10 Tao Deng,3,9 Aidi Zhang,1,4,9 Michael J. Moore,5 Jacob B. Landis,6,7 Nan Lin,1,2 Huajie Zhang,1,2 Xu Zhang,1,2 Jinling Huang,3 Xiujun Zhang,1,4,* Hang Sun,3,* and Hengchang Wang1,2,*

SUMMARY

Kingdonia uniflora, an alpine herb, has an extremely narrow distribution and represents a model for studying evolutionary mechanisms of species that have adapted to undisturbed environments for evolutionarily long periods of time. We assembled a 1,004.7-Mb draft genome (encoding 43,301 genes) of K. uniflora and found significant overrepresentation in gene families associated with DNA repair, underrepresentation in gene families associated with stress response, and loss of most plastid ndh genes. During the evolutionary process, the overrepresentation of gene families involved in DNA repair could help asexual K. uniflora reduce the accumulation of deleterious mutations, while reducing genetic diversity, which is important in responding to environmental fluctuations. The underrepresentation of gene families related to stress response and functional loss of ndh genes could be due to lack or loss of ability to respond to environmental changes caused by long-term adaptation to a relatively stable ecological environment.

INTRODUCTION

Habitat destruction caused by changing climate and human activities has driven numerous plant species to endangered status (Yang et al., 2018). Current management of endangered species includes in situ and ex situ measurements (Nuijten et al., 2016). Owing to the rapidly growing number of threatened species, optimizing conservation practices based on evolutionary history and endangered mechanisms is imperative. In recent years, the development of genomic techniques has opened up new opportunities for species conservation. The main improvement is the enormous increase in loci allowing for a more detailed study of previous research questions (Nuijten et al., 2016). Specifically, genomic approaches provide the possibility to detect adaptive genetic variation across the entire genome (Kohn et al., 2006; Ouborg et al., 2010; Angeloni et al., 2012; Nuijten et al., 2016), which may help to address long-standing questions in conservation biology not fully resolved with traditional methods (Primmer, 2009; Avise, 2010; Frankham, 2010; Ouborg et al., 2010; Allendorf et al., 2010; Angeloni et al., 2012; Funk et al., 2012; Steiner et al., 2013; McMahon et al., 2014; Shafer et al., 2015; Garner et al., 2016; Benestan et al., 2016; Fuentes-Pardo and Ruzzante, 2017).

Plant lineages vary widely in their geographic distributions owing to numerous factors, e.g., ecological niche range, dispersal capacity, reproductive investments, and time since origin (Debussche and Thompson, 2003; Lavergne et al., 2004; Becker, 2010; Youssef et al., 2011), among which one crucial factor is adaptive capacity to respond to environmental changes. Lineages maintaining a small distribution are likely to possess low adaptive capacity to respond to geological and climatic changes at large scales, as well as habitat changes at small scales. For example, some asexual lineages, generating genetically and phenotypically identical individuals, are limited in their capacity to respond quickly to environmental fluctuation owing to low genetic diversity, which can lead to a status of endangerment or even extinction (Bell and Collins, 2008). In addition, species living in an equable environment for long periods might lack or lose the ability to defend against rapidly fluctuating environmental stress. Once the habitat is altered or destroyed, these species are incapable of colonizing new habitats. Shrinking habitats and low adaptive ability to new environments together lead some plants to endangered status. A recent study has shown that endangered species have more vulnerable genomes than that of congeneric non-endangered species (Hamabata et al., 2019); in fact, the endangered species exhibited significantly lower genetic diversity.
and proportion of duplicated genes and accumulated more deleterious variations than non-endangered species. Owing to limited genomic studies, genomic changes of endangered plants are largely unknown. To clarify which genetic factors contribute to vulnerability in endangered species, genome-wide genetic changes in functionality should be assessed (Hamabata et al., 2019). Kingdonia uniflora Balf. f. and W.W. Sm. (Circaeasteraceae, Ranunculales), an alpine herb (diploid, 2n = 18), has a very narrow distribution (Figure S1). The habitat of K. uniflora represents an ecological environment of primeval forest with few disturbances. Specifically, K. uniflora is restricted to growing in high altitudes (~2,800–4,000 m); cold, damp climates with deep humus; and usually under species of Abies. K. uniflora and Circaeaster agrestis Maxim. together constitute the early-diverging eudicot family Circaeasteraceae (Ranunculales) (Angiosperm Phylogeny Group, 2016). Previous estimates showed K. uniflora diverged from C. agrestis around 52 mya (Ruiz-Sanchez et al., 2012). Although no fossil record is known for K. uniflora, fossil fruits similar to those of C. agrestis have been reported from the mid-Albian of Virginia, USA (Crane et al., 1994; Drinna et al., 1994; Sun et al., 2017). Remarkably, different from all other angiosperms, K. uniflora and C. agrestis possess an unusual dichotomous venation (Figure S2) similar to that found in ferns and Ginkgo (Sun et al., 2017). Old age for the lineage and dichotomous veins indicate an ancient relictual character for K. uniflora. Additionally, K. uniflora typically reproduces asexually, relying on rhizome systems to produce new individuals. These indications raise multiple questions, including: As an ancient relictual lineage, what history has K. uniflora experienced? As an endangered and habitat-dependent species, what genomic traits does K. uniflora possess? K. uniflora provides an ideal model to study the evolutionary mechanisms of ancient plant lineages that have an extremely narrow distribution and rely on a highly specialized habitat.

The ndh genes encode subunits of the thylakoid NADPH complex that mediate cyclic electron flow around Photosystem I and facilitates chlororespiration (Martin and Sabater, 2010). A series of studies suggests that the ndh genes can be dispensable under mild non-stressing environments (e.g., Casano et al., 2001; Martin et al., 2004; Rumeau et al., 2007; Martin and Sabater, 2010). A previous study (Sun et al., 2017) found rampant loss and pseudogenization of ndh genes in the K. uniflora plastome, which may indicate an association between the vulnerability of K. uniflora and genomic changes. In the present study, we provide a de novo genome sequence of K. uniflora using both Illumina and Pacbio sequencing technologies. We aim to investigate the evolutionary history of K. uniflora and reveal the potential mechanisms of its evolutionary specialization.

### RESULTS

#### Genome Assembly and Annotation

| Genome Features               | Contigs/Scaffolds |
|-------------------------------|-------------------|
| Total length, bp              | 1,004,656,313     |
| Total number of contigs       | 2,932             |
| Longest length, bp            | 11,531,354        |
| Length of N50, bp             | 2,099,369         |
| Length of N90, bp             | 292,588           |
| GC content, %                 | 38.04%            |
| No. of genes                  | 43,301            |

Table 1. Genome Assembly of Kingdonia uniflora

and proportion of duplicated genes and accumulated more deleterious variations than non-endangered species. Owing to limited genomic studies, genomic changes of endangered plants are largely unknown. To clarify which genetic factors contribute to vulnerability in endangered species, genome-wide genetic changes in functionality should be assessed (Hamabata et al., 2019). Kingdonia uniflora Balf. f. and W.W. Sm. (Circaeasteraceae, Ranunculales), an alpine herb (diploid, 2n = 18), has a very narrow distribution (Figure S1). The habitat of K. uniflora represents an ecological environment of primeval forest with few disturbances. Specifically, K. uniflora is restricted to growing in high altitudes (~2,800–4,000 m); cold, damp climates with deep humus; and usually under species of Abies. K. uniflora and Circaeaster agrestis Maxim. together constitute the early-diverging eudicot family Circaeasteraceae (Ranunculales) (Angiosperm Phylogeny Group, 2016). Previous estimates showed K. uniflora diverged from C. agrestis around 52 mya (Ruiz-Sanchez et al., 2012). Although no fossil record is known for K. uniflora, fossil fruits similar to those of C. agrestis have been reported from the mid-Albian of Virginia, USA (Crane et al., 1994; Drinna et al., 1994; Sun et al., 2017). Remarkably, different from all other angiosperms, K. uniflora and C. agrestis possess an unusual dichotomous venation (Figure S2) similar to that found in ferns and Ginkgo (Sun et al., 2017). Old age for the lineage and dichotomous veins indicate an ancient relictual character for K. uniflora. Additionally, K. uniflora typically reproduces asexually, relying on rhizome systems to produce new individuals. These indications raise multiple questions, including: As an ancient relictual lineage, what history has K. uniflora experienced? As an endangered and habitat-dependent species, what genomic traits does K. uniflora possess? K. uniflora provides an ideal model to study the evolutionary mechanisms of ancient plant lineages that have an extremely narrow distribution and rely on a highly specialized habitat.

The ndh genes encode subunits of the thylakoid NADPH complex that mediate cyclic electron flow around Photosystem I and facilitates chlororespiration (Martin and Sabater, 2010). A series of studies suggests that the ndh genes can be dispensable under mild non-stressing environments (e.g., Casano et al., 2001; Martin et al., 2004; Rumeau et al., 2007; Martin and Sabater, 2010). A previous study (Sun et al., 2017) found rampant loss and pseudogenization of ndh genes in the K. uniflora plastome, which may indicate an association between the vulnerability of K. uniflora and genomic changes. In the present study, we provide a de novo genome sequence of K. uniflora using both Illumina and Pacbio sequencing technologies. We aim to investigate the evolutionary history of K. uniflora and reveal the potential mechanisms of its evolutionary specialization.
regions assessed by BUSCO (Benchmarking Universal Single Copy Orthologs) showed that 90.6% of the green plant single-copy orthologs were complete (Table S5).

We compared the draft genome of *K. uniflora* with the well-annotated genomes of the model plant *Arabidopsis thaliana* (Brassicaceae) and the Ranunculales species *Aquilegia coerulea* (Ranunculaceae). The genome size of *K. uniflora* is much larger than that of both references. The *K. uniflora* genome showed strong synteny with the genome of *Aq. coerulea* (Figure 1), but weak synteny with that of *A. thaliana* (Figure S5), which is not surprising given their placement in the angiosperm Tree of Life. The gene density in *K. uniflora* is lower than that in *Aq. coerulea* and *A. thaliana*, whereas the density of TEs (transposable elements) in *K. uniflora* was higher than that in *Aq. coerulea* and *A. thaliana* (Figures 1 and S5).

We also compared our draft genome sequence with five other draft genomes of Ranunculales taxa, representing three of the seven Ranunculales families (Table S6); the quality of our assembly is comparable with that of all the five species, generating the longest N50 length and relatively fewer scaffolds. Comparatively, the genome of *K. uniflora* is larger than other Ranunculales species with sequenced genomes, such as *Aq. coerulea* (293.08 Mb), *Eschscholzia californica* (489.065 Mb), and *Macleaya cordata* (377.83 Mb) (Table S6).
Repeat Elements

Through a combination of approaches, we annotated 66.83% of the assembly as repetitive elements, among which long terminal repeats (LTRs) were the most abundant, occupying 40.62% of the genome assembly length; DNA elements and long interspersed nuclear elements occupied 5.0% and 3.0% of the genome, respectively (Table S7). The proliferation of LTR retrotransposons in *K. uniflora* was estimated to peak around 2.7 mya (Figure 2A). Analyses of age distributions built from synonymous substitutions per synonymous site (*Ks*) indicated that *K. uniflora* has undergone one recent whole-genome duplication (WGD) event, which occurred after its divergence from *C. agrestis* (Figure 2B). The inferred WGD event in the *K. uniflora* genome was further supported by dot-plot analysis of representative scaffolds, in which numerous paralogs derived from this event were identified (Figure 2C).

Phylogenetic Tree Construction and Estimation of Divergence Times

Applying OrthoFinder (Emms and Kelly, 2015) to eight whole-genome and four transcriptome sequences including monocots, basal eudicots, and core eudicots, we identified a total of 18,742 orthogroups, among which 6,883 were shared by *Kingdonia* and four other Ranunculales species (Figure 2D). Among these orthogroups, 497 were identified as putative single-copy gene families. To further investigate the phylogenetic relationships within Ranunculales, we conducted both concatenated and coalescent analyses using the sequences of 497 single-copy genes in 12 species. The topologies from two analyses were identical, confirming the sister relationship between *K. uniflora* and *C. agrestis* and resolving Circaeasteraceae as sister to the clade formed by Ranunculaceae and Berberidaceae (Figure 3); Papaveraceae + Eupteleaceae was placed as the earliest-diverging clade (Figure 3). *K. uniflora* and *C. agrestis* were estimated to have diverged ~51.8 mya in our analyses.
using MCMCtree with two calibration points (Figure 3). In addition, the phylogenetic analysis with an expanded group of taxa, which correspond to a larger taxonomic sampling but fewer loci indicated a similar phylogeny of Ranunculales, except the placement of Eupteleaceae (Figure S6).

Gene Family Overrepresentation and Underrepresentation
Comparisons of the genomes among 12 species identified a total of 111 gene families that are significantly (p < 0.01) overrepresented in K. uniflora and 22 gene families that are significantly underrepresented (Table S8). The results from Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) annotations showed that overrepresented gene families were considerably enriched in DNA repair pathways, such as homologous recombination, mismatch repair, DNA replication and nucleotide excision repair, whereas gene families showing significant underrepresentation in the K. uniflora genome were found to be involved in pathways related to stress or pest responses, such as the phenylpropanoid biosynthesis and secondary metabolites biosynthesis (Table 2).

Dispensability of Plastid ndh Genes
To detect whether the ndh genes/segments lost from K. uniflora plastome were transferred to the nuclear genome, we conducted a BLASTN search using 11 intact ndh sequences extracted from C. agrestis as the query, using the assembled K. uniflora genome sequences as target. The result showed that no intact sequences for ndh plastid genes were discovered with the exception of ndhE and ndhJ, indicating functional copies of these genes likely have been lost (Table S9 and Figure 4).

DISCUSSION
Species that live in stable habitats face less stress, which can cause lack or loss of ability to respond to environmental changes. In the present study, we de novo assembled the genome of K. uniflora, an ancient relictual species exclusively found in China. Given that K. uniflora has a larger genome than many of its close relatives, we hypothesize that the proliferation of LTR retrotransposons and the WGD event together are likely responsible for the increased genome size of K. uniflora (Michael, 2014). Several studies (Tenaillon et al., 2010; Michael, 2014) have suggested that the proliferation of TEs and specifically long terminal repeat retrotransposons (LTRs) in genomes is the primary driver of genome size differences in plants. This is because LTRs are expressed as RNA and reverse-transcribed into a new DNA element that can be inserted in every replication cycle (Wicker et al., 2007). A comparative study using high-quality genomes detected a correlation between intact LTR retrotransposons and genome size (El Baidouri et al., 2013). Abundant LTR retrotransposons were detected in the genome of K. uniflora, whereas considerably fewer LTR retrotransposons were identified in other members of the order with sequenced genomes (Figure 1).
To investigate the evolutionary dynamics of the LTR retrotransposons, we estimated their insertion dates. The results indicate that the proliferation of LTR retrotransposons in *K. uniflora* was likely triggered around one rapid uplift of the Hengduan Mountain region, occurring between the late Miocene and late Pliocene (Kirby et al., 2002; Clark et al., 2005; Sun et al., 2011; Wang et al., 2012, 2014; Meng et al., 2016; Xing and Ree, 2017). WGDs have been shown to pervade the evolutionary history of angiosperms (Landis et al., 2018), and *K. uniflora* is no different (Figures 2B and 2C). Therefore, the relatively larger genome size of *K. uniflora* compared with close relatives might be promoted by both LTRs proliferation and WGD events.

Based on phylogenetic inference and estimated divergence times, we speculate that the speciation of *K. uniflora* was promoted by the Himalayan orogeny. Previous studies based on DNA commonly recognize *K. uniflora* being closely related to *C. agrestis* (Kim et al., 2004; Wang et al., 2009; Sun et al., 2017). Our phylogenetic analyses confirmed the sister relationship between *K. uniflora* and *C. agrestis*. These species were resolved as sister to the clade formed by Ranunculaceae and Berberidaceae, a result incongruent with previous placements of Circaeasteraceae and Lardizabalaceae as sister pairs (Kim et al., 2004; Wang et al., 2009; Sun et al., 2017). The divergence estimation between *K. uniflora* and *C. agrestis* (~51.8 mya) is consistent with a previous estimate of ~52 mya (Ruiz-Sanchez et al., 2012) and also coincides with the timing of the first stage of Himalayan orogeny (Rowley, 1996; Huang et al., 2015). Similar to *K. uniflora*, *C. agrestis* has a narrow distribution confined to China and the Himalayas. However, the distribution range of Circaeaster is relatively larger; in places where *Kingdonia* occurs Circaeaster can always be found, whereas in most regions that Circaeaster is distributed *Kingdonia* is absent. Previous studies have shown that orogeny could create conditions favoring speciation of resident lineages (e.g., Hoorn et al., 2013; Wen et al., 2014; Favre et al., 2015). Hence, we hypothesize that the divergence between *K. uniflora* and *C. agrestis* was likely driven by the rapid uplift of contemporary Himalayan orogeny.

An asexual reproductive system and overrepresentation of DNA repair genes together reduce genetic diversity of *K. uniflora*. Levels of genetic diversity are often associated with reproductive strategies of species (Otálora et al., 2013). Colonization, population persistence, and extinction probabilities are all influenced

### Table 2. Functional Annotation of the Significantly Overrepresented and Underrepresented Gene Families in *Kingdonia uniflora*

| Gene Families               | KEGG Terms                        | Input No. | Background No. | p Value       | Corrected p Value |
|-----------------------------|-----------------------------------|-----------|----------------|---------------|-------------------|
| Overrepresented gene families | Glycosphingolipid biosynthesis - globo series | 9         | 9              | $5.28 \times 10^{-7}$ | $1.59 \times 10^{-5}$ |
|                             | Homologous recombination          | 17        | 56             | $2.87 \times 10^{-6}$ | $6.74 \times 10^{-5}$ |
|                             | Mismatch repair                   | 12        | 39             | $7.31 \times 10^{-5}$ | 0.000968079       |
|                             | Sphingolipid metabolism           | 9         | 26             | 0.000282866   | 0.003082136       |
|                             | DNA replication                   | 12        | 50             | 0.000516123   | 0.00546878       |
|                             | Nucleotide excision repair        | 14        | 69             | 0.000789672   | 0.00779277       |
|                             | Peroxisome                        | 16        | 87             | 0.000878589   | 0.00805414       |
|                             | Galactose metabolism              | 12        | 55             | 0.001064483   | 0.010034842      |
|                             | Plant hormone signal transduction | 33        | 271            | 0.002344688   | 0.020456082      |
| Underrepresented gene families | Cyanoamino acid metabolism        | 5         | 60             | $2.06 \times 10^{-9}$ | $1.57 \times 10^{-7}$ |
|                             | Phenylpropanoid biosynthesis      | 5         | 157            | $2.06 \times 10^{-7}$ | $6.24 \times 10^{-6}$ |
|                             | Starch and sucrose metabolism     | 5         | 202            | $6.95 \times 10^{-7}$ | $1.96 \times 10^{-5}$ |
|                             | Biosynthesis of secondary metabolites | 6     | 1,076         | 0.00020813   | 0.001073645      |
|                             | Metabolic pathways                | 6         | 1,910          | 0.004094269   | 0.015173186      |
by the reproductive systems of species (Stephenson et al., 2000; Babaré et al., 2009; Wornik and Grube, 2010; Beatty and Provan, 2011; Otálora et al., 2013). K. uniflora primarily reproduces asexually via rhizomes. This reproductive mode could produce identical individuals rapidly but lacks recombination and the possibility to create genetic variation in offspring, reducing the opportunities for adaptive evolution (Eckert, 2001; Castonguay; Angers, 2012). In addition, without segregation and recombination, obligate asexual multiplication may push a species into extinction owing to the steady accumulation of deleterious mutations (Thomas et al., 2016). During the long-term evolutionary history of K. uniflora, deleterious mutation accumulation cannot be ruled out; if something goes awry, such as the occurrence of a fatal mutation, whole clusters of clones can be wiped out. The integrated DNA-repair mechanism overrepresented in K. uniflora would allow for a reduction in the accumulation of deleterious mutations. This DNA-repair system might also reduce genetic diversity produced by mutations. Having high genetic diversity is important for plants to respond to environmental changes. We thus speculate that the extremely narrow distribution range of K. uniflora is associated with low genetic diversity, which restricted suitable environments to simplex, equable habitats rather than multiple, divergent habitats. Specifically, K. uniflora can only live in high elevations with minor human disturbances, while being characterized by perennial cold temperatures of below 0°C. The extreme temperature is likely to cause DNA, RNA, and protein damage. The overrepresentation of gene families for DNA repair might also be one kind of protection from extremely low temperature.

Underrepresentation of genes associated with stress response in K. uniflora leads to degeneration of adaptive ability to environmental changes. Phenylpropanoids are believed to contribute to all aspects of plant responses toward biotic and abiotic stimuli (Vogt, 2010). As concluded by La Camera et al. (2004), when plants suffer environmental stress, pests, or diseases, phenylpropanoids could evoke relevant response mechanisms to protect the plant from damage. Similarly, secondary metabolites also play a role in plant defense against environmental stresses, pests, and diseases (Bennett and Wallsgrove, 1994). Given the relatively undisturbed ecological environment of K. uniflora and lack of habitat stress during growth, adaptation to such conditions resulted in the functional degeneration of stress response systems. In addition, phenylpropanoids can also promote invasion of new habitats (Bais et al., 2003; Vogt, 2010). Our results indicate the underrepresentation of gene families involved in phenylpropanoid biosynthesis, which might be one reason causing low ability of K. uniflora to invade new habitats.

We conclude long-term living in highly equable habitats lead to the underrepresentation of stress response genes, which finally resulted in loss of ability to adapt to changing environments. The asexual reproductive strategy promoted overrepresentation of DNA repair genes, which reduced genetic diversity associated with adaptive capacity to environmental changes. Hence, both the underrepresentation of stress response genes and overrepresentation of DNA repair genes are responsible for the low adaptive ability of K. uniflora.
Equable habitats likely promoted the dispensability of most ndh genes in K. uniflora. Most angiosperms contain 11 plastid ndh genes, whereas all ndh genes, except for ndhE and ndhJ, were found to be either pseudogenized (ΨndhA, ΨndhB, ΨndhD, ΨndhH, and ΨndhK) or absent (ndhC, ndhF, ndhI, and ndhG) in the K. uniflora plastome (Sun et al., 2017). All 11 plastid ndh genes are intact in C. agrestis (Sun et al., 2017), indicating that the loss of ndh genes from K. uniflora occurred after the split between K. uniflora and C. agrestis, suggesting that within 52 million years most of the plastid ndh genes were lost from K. uniflora not only in the plastome but also in the nuclear genome. Among land plants, the plastid ndh loci have also been found absent in non-photosynthetic plants, epiphytes, Gnetales, conifers, and Erodium (Geraniaceae) (Kim et al., 2015; Lin et al., 2017; Nie et al., 2017). Evidence suggests that the thylakoid NADPH complex could optimize photosynthesis for plants under environmental stresses, while being found dispensable for plant growth under optimal growth conditions (Martin and Sabater, 2010). K. uniflora is extremely selective in habitat preference (Li et al., 2001) and is known as the indicator for natural ecological environment without disturbance. We hence speculate that the current habitats of K. uniflora might have promoted the dispensability of the plastid ndh genes. Additionally, within plants, NADPH supplies hydrogen for many anabolic processes (Antal et al., 2015). The underrepresentation of gene families related to metabolic pathways, as detected from our CAFE-based analyses, is likely related with the nonfunction state of plastid ndh genes.

Changing climate, shrinking habitats, and low adaptive ability to environmental changes together contributed to the extremely narrow distribution of K. uniflora. The overrepresentation of gene families involved in DNA repair could help reduce the accumulation of deleterious mutations during asexual reproduction, which is the dominant mode of reproduction in K. uniflora, while reducing genetic diversity, which is important in responding to environment fluctuations. The underrepresentation of gene families in charge of stress response and nonfunction of plastid ndh genes could be due to the adaptive degeneration caused by long-term adaptation to living in relatively stress-free environments. Considering the long evolutionary history of K. uniflora, and the fossil records from C. agrestis in the mid-Albian of Virginia, USA (Crane et al., 1994; Drinna et al., 1994), we speculate it should have been widespread around the world. Changing climate, shrinking habitats, asexual reproduction, and adaptive degeneration caused by relying on easeful environment together led it to a current status of endangerment.

Limitations of the Study
We reported the draft genome of the ancient and currently endangered K. uniflora. Although we revealed the evolutionary history and potential endangered mechanism of K. uniflora through primary analyses, population genomic analyses are required to estimate genetic diversity, reveal genetic structure and its causes, and identify loci associated with environmental adaption.

Resource Availability
Lead Contact
Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Yanxia Sun (sunyanxia@wbgcas.cn).

Materials Availability
This study did not generate new reagents.

Data and Code Availability
The accession number for the genome assembly and raw reads reported in this paper is GenBank: PRJNA587615 and PRJNA611722, respectively.

METHODS
All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION
Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101124.
ACKNOWLEDGMENTS

This work was supported by the Strategic Priority Research Program of Chinese Academy of Sciences (XDA20050203), the Program Foundation for the Backbone of Scientific Research by Wuhan Botanical Garden, Chinese Academy of Sciences (Y855241G01), the Major Program of National Natural Science Foundation of China (31590823), and the National Key R and D Program of China (2017YFC0505200).

AUTHOR CONTRIBUTIONS

H.W., H.S., and Xi. Zhang conceived the study Y.S., H.Z., and N.L. collected the materials. Y.S. and T.D.prepared DNA and RNA for sequencing. Y.S., A.Z., and Xu Zhang performed the genome assembly and genome annotation. Y.S. and A.Z. designed genomics analyses. Y.S., T.D., M.J.M., J.B.L., and J.H. wrote the manuscript with the input of all co-authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Allendorf, F.W., Hohenlohe, P.A., and Luikart, G. (2010). Genomics and the future of conservation genetics. Nat. Rev. Genet. 11, 697–709.

Angiosperm Phylogeny Group (2016). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. Bot. J. Linn. Soc. 181, 1–20.

Antal, T.K., Krendeleva, T.E., and Tyystjärvi, E. (2015). Comparative genomic paleontology across plant biotas. Biol. Rev. 90, 489–496.

Barbara, T., Martinek, G., Palma-Silva, C., Fay, M.F., Mayo, S., and Lexer, C. (2009). Genetic relationships and variation in reproductive strategies in four closely related bromeliads adapted to neotropical ‘inselbergs’. Alcantarea glazioiana, A. regina, A. peniculata and A. imperialis (Bromeliaceae). Ann. Bot. 103, 65–77.

Beatty, G.E., and Provan, J. (2011). High clonal diversity in threatened peripheral populations of the yellow bird’s nest (Hypopteryx monotropa: syn. Monotropa hypopitys). Ann. Bot. 107, 663–670.

Bell, G., and Collins, S. (2008). Adaptation, extinction and global change. Evol. Appl. 1, 3–16.

Benestan, L.M., Ferchaud, A.-L., Hohenlohe, P.A., Garner, B.A., Naylor, G.J.P., Baums, I.B., Schwartz, M.K., Kelley, J.L., and Luikart, G. (2016). Conservation genomics of natural and managed populations: building a conceptual and practical framework. Mol. Ecol. 25, 2967–2977.

Bennett, R.N., and Wallisgrove, R.M. (1994). Secondary metabolites in plant defense mechanisms. New Phytol. 127, 617–633.

Casano, L.M., Martin, M., and Sabater, B. (2001). Hydrogen peroxide mediates the induction of chloroplast Ndh complex under photooxidative stress in barley. Plant Physiol. 125, 1450e1458.

Castonguay, E., and Angers, B. (2012). The key role of epigenetics in the persistence of asexual lineages. Genet. Res. Int. 2012, 1–9.

Clark, M.K., House, M.A., Royden, L.H., Whipple, K., Burchfield, B.C., Zhang, X., and Tang, W. (2005). Late cenozoic uplift of southeastern tibet. Geology 33, 525–528.

Crane, P.R., Friis, E.M., and Pedersen, K.R. (1994). Paleobotanical evidence on the early radiation of magnoliid angiosperms. Plant Syst. Evol. 8, 51–72.

Debussche, M., and Thompson, J.D. (2003). Habitat differentiation between two closely related Mediterranean plant species, the endemic Cyclamen balearicum and the widespread C. repandum. Acta Oecol. 24, 35–45.

Drinnan, A.N., Crane, P.R., and Hoot, S.B. (1994). Patterns of floral evolution in the early diversification of non-magnoliid dicotyledons (eudicots). Plant Syst. Evol. 8, 93–122.

Eckert, C.G. (2001). The loss of sex in clonal plants. Evol. 15, 501–520.

El Baidouri, M., and Panaud, O. (2013). Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. Genome Biol. Evol. 5, 954–965.

Emma, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16, 157.

Fayre, A., Packert, M., Pauls, S.U., Jahnig, S.C., Uhl, D., Midhalak, I., and Mueller-Rohr, A.N. (2015). The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biota. Biol. Rev. 90, 236–253.

Frankham, R. (2010). Challenges and opportunities of genetic approaches to biological conservation. Biol. Conserv. 143, 1919–1927.

Fuentes-Pardo, A.P., and Ruzzante, D.E. (2017). Whole-genome sequencing approaches for conservation biology: advantages, limitations and practical recommendations. Mol. Ecol. 26, 5369–5406.

Funk, W.C., McKay, J.K., Hohenlohe, P.A., and Allendorf, F.W. (2012). Harnessing genomics for delineating conservation units. Trends Ecol. Evol. 27, 489–496.

Garner, B.A., Hand, B.K., Amish, S.J., Bernatchez, L, Foster, J.T., Miller, K.M., Morin, P.A., Narum, S.R., O’Brien, S.J., Roffler, G., et al. (2016). Genomics in Conservation: case studies and bridging the gap between data and application. Trends Ecol. Evol. 31, 81–83.

Hamabata, T., Kinoshita, G., Kurita, K., Cao, P.-L., Ito, M., Murata, J., Komaki, Y., Isagi, Y., and Makino, T. (2019). Endangered island endemic plants have vulnerable genomes. Commun. Biol. 2, 244.

Hoon, C., Mosbrugger, V., Mulch, A., and Antonelli, A. (2013). Biodiversity from mountain building. Nat. Geosci. 6, 154.
Huang, W., van Hinsbergen, D.J.J., Lippert, P.C., Guo, Z., and Dupont-Nivet, G. (2015). Paleomagnetic tests of tectonic reconstructions of the India-Asia collision zone. Geophys. Res. Lett. 42, 2642–2649.

Kim, H.T., Kim, J.S., Moore, M.J., Neubig, K.M., Williams, N.H., Whitien, W.M., and Kim, J.-H. (2015). Seven new complete plastome sequences reveal rampant independent loss of the ndh gene family across Orchids and associated instability of the inverted repeat/smaller single-copy region boundaries. PLoS One 10, e0142215.

Kim, S., Solts, D.E., Solts, P.S., Zanis, M.J., and Suh, Y. (2004). Phylogenetic relationships among early-diverging eudicots based on four genes: were the eudicots ancestrally woody? Mol. Phylogenet. Evol. 31, 16–30.

Kirby, E., Reiners, P.W., Krol, M.A., Whipple, K.X., Hodges, K.V., Farley, K.A., Tang, W.Q., and Chen, Z. (2002). Late Cenozoic evolution of the eastern margin of the Tibetan Plateau: inferences from 40Ar/39Ar and (U-Th)/He thermochronology. Tectonics 21, 1–20.

Kohn, M.H., Murphy, W.J., Ostrander, E.A., and Wayne, R.K. (2006). Genomics and conservation genetics. Trends Ecol. Evol. 21, 629–637.

La Camera, S., Gouzerh, G., Dhondt, S., Hoffmann, L., Fritig, B., Legrand, M., and Heitz, T. (2004). Metabolic reprogramming in plant innate immunity: the contributions of phenylpropanoid and oxylipin pathways. Immunol. Rev. 195, 267–284.

Landis, J.B., Solts, D.E., Li, Z., Marx, H.E., Barker, M.S., Tank, D.C., and Solts, P.S. (2018). Impact of whole-genome duplication events on diversification rates in angiosperms. Am. J. Bot. 105, 348–363.

Lavergne, S., Thompson, J.D., Garnier, E., and Debussche, M. (2004). The biology and ecology of narrow endemic and widespread plants: a review of narrow endemic and widespread plants. J. Northwest For. Univ. 16, 1–4.

Martin, M., and Sabater, B. (2010). Plastid ndh genes in plant evolution. Plant Physiol. Biochem. 48, 636e645.

Lin, C.-S., Chen, J.J.W., Chiu, C.-C., Hsiao, H.C.W., Yang, C.-J., Jin, X.-H., Leebens-Mack, J., de Pamphilis, C.W., Huang, Y.-T., Yang, L.-H., et al. (2017). Concomitant loss of NDH complex-related genes within chloroplast and nuclear genomes in some orchids. Plant J. 90, 994–1006.

Martin, M., Casano, L.M., Zapata, J.M., Guera, A., del Campo, E.M., Schmitz-Linneweber, C., Maier, R.M., and Sabater, B. (2004). Role of thylakoid Ndh complex and peroxidase in the protection against photo-oxidative stress: fluorescence and oxygen activity in wild-type and ndhF-deficient tobacco. Physiol. Plant 122, 443e452.

McMahon, B.J., Teeling, E.C., and Höglund, J. (2014). How and why should we implement genomics into conservation? Evol. Appl. 7, 999–1007.

Meng, K., Wang, E., and Wang, G. (2016). Uplift of the Enne Shan, western Sichuan basin: implication for eastward propagation of the Tibetan plateau in early Miocene. J. Asian Earth. Sci. 115, 29–39.

Michael, T.P. (2014). Plant genome size variation: bloating and purging DNA. Brief Funct. Genomics 13, 308–317.

Nie, Z.X., Ye, Y.J., Bai, T.D., Xu, M., and Xu, L.-A. (2017). Complete chloroplast genome of Pinus massoniana (Pinaceae): gene rearrangements, loss of ndh Genes, and short inverted repeats contraction. Molecules 22, 1528.

Nuijten, R.J.M., Bosse, M., Crooijmans, R.P.M.A., Madsen, O., Schaftenaar, W., Ryder, O.A., Groenen, M.A.M., and Megens, H.-J. (2016). The use of genomics in conservation management of the endangered Visayan warty pig (Sus cebifrons). Int. J. Genomics 2016, 1–9.

Ouborg, N.J., Pertoldi, C., Loeschke, V., Bijlsma, R. (Kule), and Hredick, P.W. (2010). Conservation genetics in transition to conservation genomics. Trends Genet. 26, 177–187.

Otaño, M.A.G., Salvador, C., Martínez, I., and Aragón, G. (2013). Does the reproductive strategy affect the transmission and genetic diversity of bionts in cyanolichens? A case study using two closely related species. Microp. Ecol. 65, 517–530.

Primmer, C.R. (2009). From conservation genetics to conservation genomics. Ann. NY Acad. Sci. 1162, 357–368.

Rowley, D.B. (1996). Age of initiation of collision between India and Asia: a review of stratigraphic evidence. Earth Planet. Sci. Lett. 145, 1–13.

Rúa-Sanchez, E., Rodriguez-Gomez, F. and Sosa, V. (2012). Refugia and geographic barriers of populations of the desert poppy, Hymennophila fumarolia (Papaveraceae). Org. Divers. Evol. 12, 133–143.

Rumeau, D., Pelser, G., and Cormac, L. (2007). Chlororespiration and cyclic electron flow around PSI during photosynthesis and plant stress response. Plant Cell Environ. 30, 1041e1051.

Shafer, A.B.A., Wolf, J.B.W., Alves, P.C., Bergström, L., Bruford, M.W., Bra˚nnstro´m, I., Ryder, O.A., Capy, P., Chalhoub, B., Flavell, A., LeRoy, P., Morgante, M., Panaud, O., et al. (2007). A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. 8, 973–982.

Wen, J., Zhang, J.-Q., Nie, Z.-L., Zhong, Y., and Sun, H. (2014). Evolutionary diversifications of plants on the Qinghai-Tibetan plateau. Front. Genet. 5, 1–16.

Wicker, T., Sabot, F., Huu-Van, A., BENNETZEN, J.L., Capy, P., Chalhoub, B., Flavell, A., LeRoy, P., MORGANTE, M., PANAUD, O., et al. (2007). A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. 8, 973–982.

An, X., and Ren, H. (2017). Uplift driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot. Proc. Natl. Acad. Sci. 114, E3444–E3451.

Yang, Y., Ma, T., Wang, Z., Lu, Z., Li, Y., Fu, C., Chen, X., Zhao, M., Olson, M.S., and Liu, J. (2018). Genomic effects of population collapse in a critically endangered ironwood tree Ostrya reideriana. Nat. Commun. 9, 5449.

Youssef, S., Baulem, A., Vela, E., Juin, M., Dumas, E., Affre, L., and Tatoni, T. (2011). Factors underlying the narrow distribution of the Mediterranean annual plant Arenaria provincialis (Caryophyllaceae). Folia Geobot. 46, 327–350.
Supplemental Information

Genome Sequencing of the Endangered *Kingdonia uniflora* (Circaeasteraceae, Ranunculales) Reveals Potential Mechanisms of Evolutionary Specialization

Yanxia Sun, Tao Deng, Aidi Zhang, Michael J. Moore, Jacob B. Landis, Nan Lin, Huajie Zhang, Xu Zhang, Jinling Huang, Xiujun Zhang, Hang Sun, and Hengchang Wang
Figure S1. Distribution range of *K. uniflora*. The black dot shows the individuals with previous occurrence record but no longer have extant populations; the red dots represent the current distribution range of *K. uniflora*. Related to Table 1
Figure S2. Morphological features of *K. uniflora*. Related to Table 1.
Figure S3. Estimation of *K. uniflora* genome size based on flow cytometer analysis. The above panel showing the 2C DNA of *Actinidia chinensis* (Hopping, 1994) at 32377.78, and the panel below indicating 2C DNA of *K. uniflora* at 47614.17. Related to Table 1.
Figure S4. 17 k-mer frequency distribution of sequencing reads. Related to Table 1.
Figure S5. Comparative analyses of genomic features between *Kingdonia uniflora* and *Arabidopsis thaliana*. Tracks from inside to outside are collinearity between both genomes, number of chromosomes/scaffolds, gene density, GC content and TE density. Related to Figure 1.
Figure S6. Dated phylogeny for 17 plant species with *Oryza* as an outgroup. A time scale is shown at the bottom. Related to Figure 3.
Table S1. Statistics of characteristics of *K. uniflora* genome (*K*-mer=17). Related to Table 1.

| Characteristics | Value         |
|-----------------|---------------|
| *K*-mer         | 17            |
| Peak_Depth      | 80            |
| N *K*-mer       | 58,788,492,58 |
| Genome size     | 1170 Mb       |
| Revised Genome size | 1005 Mb    |
| Heterozygous rate| nan           |
| Repeat rate     | 0.6242        |
Table S2. Sequencing and quality filtering statistics. Related to Table 1.

|                      | Total data (G) | Sequence coverage (X) |
|----------------------|----------------|-----------------------|
| Illumina sequencing   | 236            | 201.7                 |
| Pacbio Sequencing     | 106.5          | 90.6                  |
| Total                | 342.5          | 292.7                 |
Table S3. Information of function annotation in *K. uniflora* genes. Related to Table 1.

| Database    | Annotated Number | Annotated Percent (%) |
|-------------|------------------|-----------------------|
| NR          | 35818            | 82.7%                 |
| Swiss-Prot  | 27859            | 64.3%                 |
| KEGG        | 32104            | 74.1%                 |
| InterPro    | 25951            | 59.9%                 |
| Total       | 35953            | 83.03%                |
Table S4. Statistics of noncoding RNA in *K. uniflora* genome. Related to Table 1.

| Type     | Copy | Average length (bp) | Total length (bp) |
|----------|------|---------------------|-------------------|
| tRNA     | 1124 | 76                  | 85487             |
| rRNA     | 715  |                     |                   |
| 18S      | 81   | 1802                | 146008            |
| 28S      | 76   | 6732                | 511650            |
| 5.8S     | 0    | _                   | _                 |
| 5S       | 558  | 115                 | 63990             |
| snRNA    |      |                     |                   |
| CD-box   | 1447 | 104                 | 149886            |
| HACA-box | 97   | 124                 | 11991             |
| splicing | 207  | 154                 | 31895             |
| miRNA    | 125  | 126                 | 15757             |
Table S5. Scaffolds from the *K. uniflora* assembly were aligned to conserved genes using BUSCO method. Related to Table 1.

| Species | Genome Size | BUSCO annotation assessment results |
|---------|-------------|-------------------------------------|
| *K. uniflora* | 1004.7 Mb | C:90.6% [D:4.1%], F:3.3%, M:6.1%, n:1375 |

C: Complete Single-Copy BUSCOs  
D: Complete Duplicated BUSCOs  
F: Fragmented BUSCOs  
M: Missing BUSCOs  
n: Total BUSCO groups searched
Table S6. Comparison of genome assembly within Ranunculales. Related to Table 1.

| Species               | Family            | Genome size (Mb) | No. of scaffolds | No. of contigs | Length of N50, bp | References        |
|-----------------------|-------------------|------------------|------------------|----------------|------------------|-------------------|
| *Aquilegia coerulea*  | Ranunculaceae     | 301.98           | 970              | 7,189          | 121,821          | Filiault et al., 2018 |
| *Berberis thunbergii* | Berberidaceae     | 2240.74          | 11,815           | 11,815         | 397,058          | NCBI available    |
| *Kingdonia uniflora*  | Circaeasteraceae  | 1004.7           | 2,932            | 2932           | 2,099,369        |                   |
| *Eschscholzia californica* | Papaveraceae | 489.065          | 53,253           | 85,931         | 20,647           | Hori et al., 2018 |
| *Macleaya cordata*    | Papaveraceae      | 377.834          | 4,547            | 25,550         | 36,130           | Liu et al., 2017  |
| *Papaver somniferum*  | Papaveraceae      | 2715.53          | 34,381           | 65,344         | 1,773,300        | Guo et al., 2018  |
### Table S7. Statistics of repeat sequences and transposable elements in *K. uniflora* genome. Related to Table 1.

| Type          | Repeat size  | Percent of genome (%) |
|---------------|--------------|-----------------------|
| DNA           | 50812195     | 5.06                  |
| LINE          | 30109270     | 3.00                  |
| SINE          | 173746       | 0.02                  |
| LTR           | 408124656    | 40.62                 |
| Simple repeat | 5211550      | 0.52                  |
| Unknown       | 179205100    | 17.84                 |
| **Total**     | **671481635**| **66.83**             |
Table S9. Length comparasion of *ndh* genes between *K. uniflora* and *C. agrestis*. Related to Figure 4.

|       | *C. agrestis* (bp) | *K. uniflora* (bp) |
|-------|-------------------|--------------------|
| *ndhA* | 1089              | 553                |
| *ndhB* | 1530              | 723                |
| *ndhC* | 360               | 0                  |
| *ndhD* | 1500              | 15                 |
| *ndhE* | 303               | 303                |
| *ndhF* | 2205              | 0                  |
| *ndhG* | 528               | 0                  |
| *ndhH* | 1179              | 615                |
| *ndhI* | 540               | 0                  |
| *ndhJ* | 474               | 468                |
| *ndhK* | 666               | 234                |
Table S10. Information of species used for phylogenetic analyses. Colored characters showing the expanded taxa in Figure S5 compared with that in Figure 3. Related to Figure 3.

| Species                        | Family         | Source   |
|--------------------------------|----------------|----------|
| Oryza sativa L.                | Poaceae        | NCBI     |
| Vitis vinifera L.              | Vitaceae       | NCBI     |
| Populus trichocarpa Torr. and Gray | Salicaceae   | NCBI     |
| Arabidopsis thaliana L.         | Brassicaceae   | NCBI     |
| Nelumbo nucifera Gaertner      | Nelumbonaceae  | NCBI     |
| Euptelea pleiosperma J. D. Hooker and Thomson | Eupteleaceae | 1 kp     |
| Argemone mexicana L.            | Papaveraceae   | 1 kp     |
| Papaver bracteatum Lindl.       | Papaveraceae   | 1 kp     |
| Capnoides sempervirens (L.) Borkh. | Papaveraceae | 1 kp     |
| Macleaya cordata (Willd.) R. Br. | Papaveraceae   | NCBI     |
| Akebia trifoliata (Thunberg) Koidzumi | Lardizabala   | 1 kp     |
| Circaeaster agrestis Maxim.     | Circaeasteraceae | Current study |
| Kingdonia uniflora Balf. f. and W.W. Sm. | Circaeastera   | Current study |
| Hydrastis canadensis L.         | Ranunculaceae  | 1 kp     |
| Common Name                        | Family            | Source  |
|-----------------------------------|-------------------|---------|
| Aquilegia coerulea E. James       | Ranunculaceae     | NCBI    |
| Podophyllum peltatum L.            | Berberidaceae     | 1 kp    |
| Nandina domestica Thunberg         | Berberidaceae     | 1 kp    |
Transparent Methods

1 Genome sequencing and assembly

1.1 Plant materials and sequencing

Fresh *K. uniflora* leaves were collected from individuals growing from the same rhizome in the Taibai Mountains (altitude 2,844 m, N 34.038°, E107.715°), Shaanxi, China. Total genomic DNA (≥10 ug, ≥50 ng/ul) was isolated from fresh leaves using the conventional cetyltrimethylammonium bromide (CTAB) method (Doyle and Doyle, 1987). To help to estimate the genome size and polish genome assembly, we conducted Illumina sequencing; two paired-end sequencing libraries with insert sizes of 270 bp and 500 bp, respectively, were constructed and sequenced on the Illumina HiSeq X ten platform (Illumina Inc., CA, USA) at Beijing Genomics Institute (BGI) in Wuhan, Hubei, China. For PacBio single-molecule real-time sequencing, sequencing libraries with 20-kb DNA inserts were constructed and sequenced on the PacBio Sequel platform (Pacific Biosciences, CA, USA) at BGI. We also collected fresh leaves of *C. agrestis* in Taibai Mountains (altitude 2,837m, N34.038°, E107.68) for RNA extraction. Total RNA was extracted from young leaves (~100 mg) of both *K. uniflora* and *C. agrestis* using TRIzol Reagent RNA Purification (DSB, Guangdong, China). A cDNA library with insert sizes of 350-400 bp was prepared using NEBNext Ultra RNA Library Prep Kit for Illumina (NEB, MA, USA) and paired-end sequenced on the HiSeq X ten platform (Illumina Inc., CA, USA) at
BGI. The transcriptome data of *K. uniflora* was used for the prediction of protein-coding genes; and the transcriptome data of *C. agrestis* was used to detect single copy genes.

**1.2 De novo assembly**

The PacBio long reads were first error corrected and then *de novo* assembled using Canu v1.8 (Koren et al., 2017) with default parameters (rawErrorRate=0.300, correctedErrorRate=0.045, minReadLength=1000, minOverlapLength=500, canuIterationMax=2) except for setting the genome size to 1.2 G to obtain contigs. Then iterative polishing was conducted on the Canu derived contigs using Pilon v1.2.3 (Walker et al., 2014) in which adapter-trimmed paired-end Illumina reads from DNA sequencing were aligned with the raw assembly with default parameters to fix bases and correct local misassembles. RNA-seq reads were assembled into transcripts using Trinity v2.6.6 (Grabherr et al., 2011) with the paired-end option and remaining default parameters.

**2 Genome annotation**

**2.1 Annotation of repetitive sequences**

We identified *de novo* repetitive sequences in the *K. uniflora* genome using RepeatModeler (http://www.repeatmasker.org/RepeatModeler/) based on a self-blast search. We further used RepeatMasker (http://www.repeatmasker.org/) to search for known repetitive sequences using a
cross-match program with a Repbase-derived
RepeatMasker library and the \textit{de novo} repetitive sequences constructed by RepeatModeler. Intact LTR (long terminal repeat) retrotransposons were identified by searching the genome of \textit{K. uniflora} with LTRharvest (Ellinghaus et al., 2008) (-motif tgca -motifmis 1) and LTR_Finder (Xu and Wang, 2007) (-D 20000 -d 1000 -L 5000 -I 100). We combined results from both analyses and filtered false positives using LTR_retriever (Qu and Jiang, 2018), which also calculated the insertion date (\(t\)) for each LTR retrotransposons (\(t = K/2r\), \(K\): genetic distance) using a substitution rate (\(r\)) of 1.4*10−9 substitutions per site per year calculated by MCMCtree in PAML (Yang, 2007).

2.2 Structural and functional annotation of genes

Putative protein-coding gene structures in the \textit{K. uniflora} genome were homology predicted using the Maker package v2.31.10 (Holt and Yandell, 2011) with protein references from the published Ranunculales genomes and the \textit{de novo} assembled transcripts of \textit{K. uniflora} transcriptome data generated in this study, and \textit{de novo} predicted using Augustus v3.3.2 (Stanke et al., 2006). The rRNAs were predicted using RNAmmer v1.2 (Lagesen et al., 2007), tRNAs were predicted using tRNAscan-SE v1.4 (Lowe and Eddy, 1997), and other noncoding RNA sequences were identified using Rfam v12.0 by inner calling using Infernal v1.1.2 (Nawrocki and Eddy, 2013).

Functional annotation of the protein-coding genes was carried out by performing BLASTP analyses (e-value cut-off 1e-05) against the NCBI nonredundant protein sequence database and SwissProt. Searches for gene motifs and domains were performed using InterProScan v5.16.55.
Completeness of the genome was assessed by performing gene annotation using the BUSCO (v3.0.2) method (Simão et al., 2015) by searching the Embryophyta library.

3 Investigation of whole-genome duplication

We identified paralogs (within *K. uniflora* and *C. agrestis*, respectively) and orthologs (between *K. uniflora* and *C. agrestis*) using BLASTP (E value = 1E-07). For each gene pair, the number of synonymous substitutions per synonymous site (*Ks*) based on the NG method was calculated using TBtools (Chen et al., 2018); *Ks* values of all gene pairs were plotted to identify putative whole-genome duplication events. In addition, MCScanx (Wang et al., 2012) was used to identify syntenic blocks within the *K. uniflora* genome. Dot-plot analysis of syntenic blocks with at least five gene pairs was conducted using the dot plotter program within the MCScanX package to further detect whole-genome duplication events.

4 Gene family and phylogenomic analysis

Orthogroups were constructed with 11 other sequenced plant species (Table S10). To get reliable tree topology, all basal eudicot species (*Nandina domestica*, *Aquilegia coerulea*, *Circaeaster agrestis*, *Akebia trifoliata*, *Macleaya cordata*, *Euptelea pleiosperma*, and *Nelumbo nucifera*) with available genome sequences were included for the phylogenetic analyses. In addition, the frequently-used outgroups (*Arabidopsis thaliana*, *Populus trichocarpa*, *Vitis vinifera* and *Oryza sativa*) in previous studies (e.g., Chen et al., 2018; Song et al., 2018; Yang et al., 2019)
were also included in our analyses. CD-HIT (Huang et al., 2010) was employed to remove redundancy caused by alternative splicing variations (-c 0.8 -aS 0.8). To exclude putative fragmented genes, genes encoding protein sequences shorter than 50 aa (amino acids) were filtered out. All filtered protein sequences of the 12 species were compared with each other using BLASTP (E value = 1E-5) and clustered into orthologous groups by OrthoFinder (Emms and Kelly, 2015). Protein sequences of single-copy gene families identified by OrthoFinder were used for phylogenetic tree construction. MAFFT version 7.0 (Katoh and Standley, 2013) was used to generate multiple sequence alignment for protein sequences in each single-copy family. Poorly aligned regions were further trimmed using the Gblocks (Castresana, 2000; Talavera and Castresana, 2007). The alignments of each gene family were concatenated to a super alignment matrix, which was then used for phylogenetic tree reconstruction through the PROTCATJTT model in RAxML version 8.1.2 (Stamatakis, 2014). To assess species tree clade support, a coalescent-based analysis was also conducted using RAxML bootstrap gene trees as input for ASTRAL v. 4.7.6 (Mirarab et al., 2015). To test the accuracy of above phylogenetic analyses, a second data set consisting of 17 taxa was also used following the same steps which consisted of increased taxonomic sampling with the tradeoff of fewer loci.

To investigate the evolutionary history of *K. uniflora*, divergence time between 12 species was estimated using MCMCtree in PAML (Yang, 2007) with the options “independent rates” and “HKY85” model. A Markov chain Monte Carlo analysis was run for 100,000,000 generations, using a burn-in of 1,000 iterations. Two constraints were used for time calibrations: (1) 140–150 Mya for the monocot-dicot split (Gaut et al., 1996; Yang et al., 2018); 112-124 Mya for the Ranunculales crown group (Magallón et al., 2015; Sun et al., 2018).
5 Gene family overrepresentation and underrepresentation

Overrepresentation and underrepresentation of the OrthoFinder-derived orthologous gene families were determined using CAFÉ v. 4.1 (De Bie et al., 2006). The program uses a birth and death process to model gene gain and loss across a user-specified phylogenetic tree. The distribution of family sizes generated under this model can provide a basis for assessing the significance of the observed family size differences among taxa. For each significantly overrepresented and underrepresented gene family in *K. uniflora*, functional information was inferred via KOBAS (http://kobas.cbi.pku.edu.cn/anno_iden.php) using KEGG Pathway database.

6 Plastid *ndh* gene searching

To examine whether the plastid *ndh* genes have been completely lost from *K. uniflora* or transferred to the nuclear genome, intact sequences of all (11) plastid *ndh* genes (*ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ and ndhK*) were extracted from the plastome of *C. agrestis* (Sun et al., 2017). Then BLASTN analyses (E value = 1E-5) between the 11 gene sequences and assembled *K. uniflora* genome sequences was conducted.
Supplemental references

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540-552.

Chen, C., Chen, H., He, Y.H., and Xia, R. TBtools, a Toolkit for Biologists integrating various biological data handling tools with a user-friendly interface. DOI: https://doi.org/10.1101/289660 (2018)

Chen, J., Hao, Z., Guang, X., Zhao, C., Wang, P., Xue, L., Zhu, Q., Yang, L., Sheng, Y., Zhou, Y., et al. (2019). Liriodendron genome sheds light on angiosperm phylogeny and species-pair differentiation. Nat. Plants 5, 18-25.

De Bie, Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFÉ: a computational tool for the study of gene family evolution. Bioinformatics 22, 1269-1271.

Doyle, J.J., and Doyle, J.L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19, 11-15.

Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics 9, 18.

Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16, 157.

Gaut, B.S., Morton, B.R., McCaig, B.C., and Clegg, M.T. (1996). Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene Adh parallel rate differences at the plastid gene rbcL. Proc. Natl. Acad. Sci. 93, 10274-10279.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al.
(2011). Full-length transcriptome without a genome from RNA-Seq data. Nat. Biotechnol. 29, 644-652.

Holt, C., and Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12, 491.

Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics 26, 680.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics 30, 1236-1240.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772-780.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27, 722-736.

Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H.-H., Rognes, T., and Ussery, D.W. (2007). RNAmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35, 3100-3108.

Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955-964.

Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L.L., and Hernández-Hernández, T. (2015). A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. New Phytol. 207, 437-453.
Mirarab, S., and Warnow, T. (2015). ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31, 44–52.

Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29:, 2933-2935.

Qu, S.J., and Jiang, N. (2018). LTR_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons. Plant Physiol. 176, 1410-1422.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210-3212.

Song, C., Liu, Y., Song, A., Dong, G., Zhao, H., Sun, W., Ramakrishnan, S., Wang, Y., Wang, S., Li, T., et al. (2018). The Chrysanthemum nankingense Genome Provides Insights into the Evolution and Diversification of Chrysanthemum Flowers and Medicinal Traits. Mol Plant 11, 1482-1491.

Stamatakis, A., Ludwig, T., and Meier, H. (2004). RAxML-III: A fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics 21, 456-463.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 34, W435-W439.

Sun, Y., Moore, M. J., Landis, J. B., Lin, N., Chen, L., Deng, T., Zhang, J.W., Meng, A.P., Zhang, S.J., Tojibaev, O.S., et al. (2018). Plastome phylogenomics of the early-diverging eudicot family Berberidaceae. Mol. Phylogenet. Evol. 128, 203-211.

Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein
sequence alignments. Systematic Biol. 56, 564-577.

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9, e112963.

Wang, E., Kirby, E., Furlong, K. P., van Soest, M., Xu, G., Shi, X., Kamp, P.J.J., and Hodges, K. V. (2012). Two-phase growth of high topography in eastern Tibet during the Cenozoic. Nat. Geosci. 5, 640-645.

Xu, Z., and Wang, H. (2007). LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 35, W265-W268.

Yang, X., Yue, Y., Li, H., Ding, W., Chen, G., Shi, T., Chen, J., Park, M.S., Chen, F., and Wang, L. (2018). The chromosome-level quality genome provides insights into the evolution of the biosynthesis genes for aroma compounds of *Osmanthus fragrans*. Horticulture Research 5, 72.

Yang, Y., Ma, T., Wang, Z., Lu, Z., Li, Y., Fu, C., Chen, X., Zhao, M., Olson, M.S., and Liu, J. (2018). Genomic effects of population collapse in a critically endangered ironwood tree *Ostrya rehderiana*. Nature Communications 9, 5449.

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586-1591.