The draft genome of the endangered, relictual plant *Kingdonia uniflora* (Circaeasteraceae, Ranunculales) reveals potential mechanisms and perils of evolutionary specialization

Yanxia Sun[#,1,2], Tao Deng[#,3], Aidi Zhang[#,1,4], Michael J. Moore[5], Jacob B. Landis[6], Nan Lin[1], Huajie Zhang[1], Xu Zhang[1], Jinling Huang[7], Xiujun Zhang[*,1,4], Hang Sun[*,3], Hengchang Wang[*,1,2]

[1]CAS Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, Hubei, China

[2]Center of Conservation Biology, Core Botanical Gardens, Chinese Academy of Sciences, Wuhan, Hubei, China

[3]Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan, China

[4]Center of Economic Botany, Core Botanical Gardens, Chinese Academy of Sciences, Wuhan, Hubei, China

[5]Department of Biology, Oberlin College, Oberlin, OH, USA

[6]Department of Botany and Plant Sciences, University of California Riverside, Riverside, CA, USA

[7]Department of Biology, East Carolina University, Greenville, NC, USA

[#]These authors contributed equally to this article.

***Corresponding author:** E-mail: hcwang@wbgcas.cn, sunhang@mail.kib.ac.cn, zhangxj@wbgcas.cn

Abstract

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 evolutionary long periods of time. We assembled a 1,004.7-Mb draft genome (encoding 43,301 genes) and investigated the evolutionary history of _K. uniflora_, along with mechanisms related to its endangered status. Phylogenomic analyses based on 497 single copy genes confirmed the sister relationship between _K. uniflora_ and _Circaeaster agrestis_, which were estimated to have diverged around 52 Mya. Proliferation of LTR retrotransposons in _K. uniflora_ is estimated to occur around 2.7 Mya, coinciding with one recent uplift of the Hengduan Mountains between the late Miocene and late Pliocene. Across 12 species of monocots, early-diverging eudicots and core eudicots, _K. uniflora_ showed significant overrepresentation in gene families associated with DNA repair and underrepresentation in gene families associated with stress response. Most of the plastid _ndh_ genes were found to be lost not only in the plastome but also in the nuclear genome of _K. uniflora._ 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 at the same time, reducing genetic diversity which is important in responding to environment 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.

Key words: _Kingdonia uniflora_, evolutionary mechanisms, draft genome, LTR retrotransposons, gene families, DNA repair, stress response

Introduction

Habitat destruction caused by changing climate and human activities has driven numerous plant species to endangered status (Yang et al., 2018). Recently there has been a push to establish natural reserves to reduce the chance of extinction in vulnerable species. Nevertheless, as generally recognized, there is a need to apply genomics in conservation, employing genomic analyses to preserve plant diversity (Shafer et al., 2015). Genomics provides new opportunities for a detailed understanding of genetic diversity across the genome and the process involved in generating or losing this diversity (Windig and Engelsma, 2011), which provides aid to those making crucial policy decisions for conservation (Garner et al., 2016).

Plant lineages vary widely in their geographic distributions due to numerous factors, 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 due 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.

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 (~2800-4000 m), cold, damp climates with deep humus, and usually under species of *Abies*. *Kingdonia uniflora* and *Circaeaster agrestis* Maxim. together constitute the

early-diverging eudicot family Circaeasteraceae (Ranunculales) (APG IV, 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). All above indicate an ancient relictual character for *K. uniflora.* Additionally, *K. uniflora* typically reproduces asexually, relying on rhizome systems to produce new individuals. Hence, *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.

A previous study (Sun et al., 2017) found rampant loss and pseudogenization of *ndh* genes in the *Kingdonia uniflora* plastome. The possibility of the transfer of the lost plastid segments to the nuclear genome at that time could not be determined. 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 size estimation using flow cytometry suggested a haploid genome size of 1150 Mb for *K. uniflora* (Figure S3), while *k*-mer statistics indicated a similar genome size of 1170 Mb, with very low heterozygosity (Figure S4, Table S1). In the present study, we generated 236 Gb of Illumina reads and 106 Gb Pacbio reads (Table S2). A total assembly of 1,004.7 Mb (representing ~86% of the estimated genome size), consisting of 2,932 scaffolds (scaffold N50 length, 2.09 Mb; longest scaffold, 11.5 Mb) was achieved (Table 1). A total of 43,301 protein-coding genes were predicted

(Table 1), among which 35,953 genes (83.03%) were functionally annotated (Table S3). In addition to protein-coding genes, various noncoding RNA sequences were identified and annotated (Table S4), including 1,124 transfer RNAs, 715 ribosomal RNAs, 125 microRNAs, and 1,751 small nuclear RNAs. The completeness of gene 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 *Kingdonia 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*; while 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 to five other draft genomes of Ranunculales taxa, representing three of the seven Ranunculales families that have reported genome sequences available (Table S6); the quality of our assembly is comparable to 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), the (Table S6).

Repeat elements

Through a combination of approaches, we annotated 66.83% of the assembly as repetitive elements, among which LTRs were the most abundant, occupying 40.62% of the genome assembly length; DNA elements and long interspersed nuclear

145 elements occupied 5.0% and 3.0% of the genome, respectively (Table S7). The
146 proliferation of LTR retrotransposons in *K. uniflora* was estimated to peak around 2.7
147 Mya (Figure 2A). Analyses of age distributions built from synonymous substitutions
148 per synonymous site (*Ks*) indicated that *K. uniflora* has undergone one recent WGD
149 event, which occurred after its divergence from *C. agrestis* (Figure 2B). The inferred
150 WGD event in the *K. uniflora* genome was further supported by dot-plot analysis of
151 representative scaffolds, in which numerous paralogs derived from this event were
152 identified (Figure 2C).
153

154 Phylogenetic tree construction and estimation of divergence times

155 Applying OrthoFinder (Emms and Kelly, 2015) to eight whole-genome and four
156 transcriptome sequences including monocots, basal eudicots and core eudicots, we
157 identified a total of 18,742 orthogroups, among which 6,883 were shared by
158 *Kingdonia* and four other Ranunculales species (Figure 2D). Among these
159 orthogroups, 497 were identified as putative single-copy gene families. To further
160 investigate the phylogenetic relationships within Ranunculales, we conducted both
161 concatenated and coalescent analyses using the sequences of 497 single-copy genes in
162 12 species. The topologies from two analyses were identical, confirming the sister
163 relationship between *K. uniflora* and *C. agrestis*, and resolving Circaeasteraceae as
164 sister to the clade formed by Ranunculaceae and Berberidaceae (Figure 3);
165 Papaveraceae + Eupteleaceae was placed as the earliest-diverging clade (Figure 3). *K.*
166 *uniflora* and *C. agrestis* were estimated to have diverged ~51.8 Mya in our analyses
167 using MCMCtree with two calibration points (Figure 3). In addition, the phylogenetic
168 analysis with an expanded group of taxa, which correspond to a larger taxonomic
169 sampling but fewer loci indicated a similar phylogeny of Ranunculales, except the
170 placement of Eupteleaceae (Figure S6).
171

172 Gene family overrepresentation and underrepresentation

173 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; while 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, 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 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 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*, while 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 the proliferation of LTR retrotransposons in *K. uniflora* was likely triggered around one rapid uplift of the Hengduan Mountains 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). Whole genome duplications (WGD) 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 to 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 Himalaya orogeny. Previous DNA based studies commonly recognized that *Kingdonia uniflora* is closely related to *Circaeaster agrestis* (Kim et al., 2004; Wang et al., 2009; Sun et al., 2017); but the relationships within Ranunculales have remained unstable (e.g., Kim et al., 2004; Wang et al., 2009; Sun et al., 2017; Lane et al., 2018). Our phylogenetic analyses confirmed the sister relationship between *K. uniflora* and *C. agrestis*, while resolving them as sister to the clade formed by Ranunculaceae and Berberidaceae, which is not congruent with previous placements of Circaeasteraceae and Lardizabalaceae as sister pairs (Kim et al., 2004; Wang et al., 2009; Sun et al., 2017). The close relationship between Ranunculaceae and Berberidaceae have been recognized in both previous and current phylogenies (e.g., Kim et al., 2004; Wang et al., 2009; Sun et al., 2017; Lane et al., 2018). However, in previous studies, either Papaveraceae (Hoot et al., 1999; Soltis et al., 2000) or Eupteleaceae (Kim et al., 2004; Wang et al., 2009; Sun et al., 2017; Lane et al., 2018) have been identified as the early-diverging clade sister to

other Ranunculales lineages. Although our phylogenetic analysis based on 264 genes agree with the placement of Eupteleacea as the early-diverging clade of Ranunculales (*Figure S6*), our 497 gene set analyses resolved Papaveraceae + Eupteleaceae as the early-diverging clade (*Figure 3*). With the accumulation of molecular evidence, conflict between gene trees is ubiquitous. Conflicting gene trees could be caused by multiple factors, such as hybridization/introgression, incomplete lineage sorting, gene duplication/loss and horizontal gene transfer (Zou and Ge, 2008; Lu et al., 2016). While considering the relatively far relationships among different Ranunculales families, gene duplication/loss is possibly a major contributor to the conflicting gene trees within Ranunculales. 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* also has a narrow distribution confined to China and the Himalayas. However, the distribution range of *Circaeaster* is relatively larger; in places where *Kingdonia* occur *Circaeaster* can always be found; while 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.

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 by the reproductive systems of species (Stephenson et al., 2000; Babará et al., 2009; Wornik and Grube, 2010; Beatty and Provan 2011; Otálora et al., 2013). *Kingdonia 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 and Angers, 2012). In addition, without segregation and recombination, the obligate asexual multiplication may push a species into extinction due 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 it to reduce the accumulation of deleterious mutations. While, this DNA-repair system might also reduce genetic diversity produced by mutations. Having high genetic diversity is very important for plants to respond to environmental changes. We thus speculate 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 zero degrees centigrade. 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 towards biotic and abiotic stimuli (Vogt, 2010). As concluded by *La Camera et al.* (*2004*) when plants suffer environmental stress or pest diseases, phenylpropanoids could evoke relevant response mechanisms to protect plants from damage. Similarly, secondary metabolites also play a role in plant defense against environmental stresses and pest diseases (Bennett and Wallsgrove, 1994). Given the relatively stable 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 leaded to the underrepresentation of stress response genes, which finally resulted in loss of ability to adapt to other environments; while the asexual reproductive strategy promoted overrepresentation of DNA repair genes, which reduced genetic diversity associated with adaptive capacity to environmental changes. Hence, both of the underrepresentation of stress response genes and overrepresentation of DNA repair genes are responsible for the low adaptive ability of *K. uniflora.*

Equable habitats probably promoted dispensability of most *ndh* genes in *K. uniflora*. The *ndh* genes encode subunits of the thylakoid NADPH complex that mediates cyclic electron flow around Photosystem I and facilitates chlororespiration (Martín and Sabater, 2010). 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 *K. uniflora* plastome (Sun et al., 2017). All 11 plastid *ndh* genes are intact in *C. agrestis* (Sun et al., 2017), indicating 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; Ni 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 (Martín and Sabater, 2010). *Kingdonia uniflora* is extremely selective in habitat preference, and is known as the indicator for natural ecological environment without disturbance. A series of studies suggests that the *ndh* genes can be dispensable under mild non-stressing environments (e.g., Casano et al., 2001; Martín et al., 2004; Rumeau et al., 2007; Martín and Sabater, 2010). 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 anabolism processes (Antal et al., 2015). The underrepresentation of gene families related to metabolic pathways, as detected from our CAFÉ 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 dominate mode of reproduction in *K. uniflora*, while at the same time, 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 are 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 lead it to a current status of endangerment.

Methods

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 cetyltriethylammonium bromide (CTAB) method (Doyle and Doyle, 1987). For 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.

De novo assembly

The PacBio long reads were first corrected and *de novo* assembled using Canu v1.8 (Koren et al., 2017) with default parameters 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.

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 *de novo* repetitive sequences constructed by RepeatModeler. Intact LTR (long terminal repeat) retrotransposons were identified by searching the genome of *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, 2017), 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) .

Structural and functional annotation of genes

Putative protein-coding gene structures in the *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 *de novo* assembled transcripts of *K. uniflora* transcriptome data generated in this study, and *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 (Jones et al., 2014). Completeness of the genome was assessed by performing gene annotation using the BUSCO (v3.0.2) methods (Simão et al, 2015) by searching the Embryophyta library..

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

410 further detect whole-genome duplication events.

411

412 ## Gene family and phylogenomic analysis

413 Orthogroups were constructed using eight genome sequences and four transcriptome
414 sequences (Table S10). CD-HIT (Huang et al., 2010) was employed to remove
415 redundancy caused by alternative splicing variations (-c 0.8 -aS 0.8). To exclude
416 putative fragmented genes, genes encoding protein sequences shorter than 50 aa
417 (amino acids) were filtered out. All filtered protein sequences of 12 species were
418 compared with each other using BLASTP (E value = 1E-5) and clustered into
419 orthologous groups by OrthoFinder (Emms and Kelly, 2015). Protein sequences of
420 single-copy gene families identified by OrthoFinder were used for phylogenetic tree
421 construction. MAFFT version 7.0 (Katoh and Standley, 2013) was used to generate
422 multiple sequence alignment for protein sequences in each single-copy family. Poorly
423 aligned regions were further trimmed using the Gblocks (Castresana, 2000; Talavera
424 and Castresana, 2007). The alignments of each gene family were concatenated to a
425 super alignment matrix, which was then used for phylogenetic tree reconstruction
426 through the PROTCATJTT model in RAxML version 8.1.2 (Stamatakis, 2014). To
427 assess species tree clade support, a coalescent-based analysis was also conducted
428 using RAxML bootstrap gene trees as input for ASTRAL v. 4.7.6 (Mirarab et al.,
429 2015). A second data set consisting of 17 taxa was also used following the same steps
430 which consisted of increased taxonomic sampling with the tradeoff of fewer loci.
431 Divergence time between 12 species was estimated using MCMCtree in PAML
432 (Yang, 2007) with the options “independent rates” and “HKY85” model. A Markov
433 chain Monte Carlo analysis was run for 100,000,000 generations, using a burn-in of
434 1,000 iterations. Two constraints were used for time calibrations: (1) 140–150 Mya
435 for the monocot-dicot split (Gaut et al., 1996; Yang et al., 2018); 112-124 Mya for the
436 Ranunculales crown group (Magallón et al., 2015; Sun et al., 2018).

437

438 ## 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). 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.

Plastid *ndh* gene searching

Intact sequences of all (11) plastid *ndh* genes, including *ndhA*, *ndhB*, *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ* and *ndhK*, were extracted from the plastome of *C. agrestis*[10]. Then BLASTN analyses (E value = 1E-5) between the 11 gene sequences and assembled *K. uniflora* genome sequences was conducted.

Acknowledgements

This work was supported by the Strategic Priority Research Program of Chinese Academy of Sciences (XDA20050203), the Programme 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).

References

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). Multiple regulatory mechanisms in the chloroplast of green algae: relation to hydrogen production. Photosynth. Res. 125:357-381.

Bais, H.P., Vepachedu, R., Gilroy, S., Callaway, R.M., and Vivanco, J.M. (2003). Allelopathy and exotic plant invasion: from molecules and genes to species interactions. Science 301:1377-1380.

468 Barbará, T., Martinelli, G., Palma-Silva, C., Fay, M. F., Mayo, S., and Lexer, C.
469 (2009). Genetic relationships and variation in reproductive strategies in four
470 closely related bromeliads adapted to neotropical ‘inselbergs’: *Alcantarea*
471 *glaziouana*, *A. regina*, *A. geniculata* and *A. imperialis* (Bromeliaceae). Ann. Bot.
472 103:65-77.
473 Beatty, G.E., and Provan, J. (2011). High clonal diversity in threatened peripheral
474 populations of the yellow bird’s nest (*Hypopitys monotropa*; syn. *Monotropa*
475 *hypopitys*). Ann. Bot. 107:663-670.
476 Bell, G., and Collins, S. (2008). Adaptation, Extinction and Global Change. Evol.
477 Appl. 1:3-16.
478 Bennett, R.N. and Wallsgrove, R.M. (1994). Secondary Metabolites in Plant Defense
479 Mechanisms. New Phytology 127:617-633.
480 Casano, L.M., Martín, M., and Sabater, B. (2001). Hydrogen peroxide mediates the
481 induction of chloroplast Ndh complex under photooxidative stress in barley.
482 Plant Physiol. 125:1450e1458.
483 Castonguay, E., and Angers, B. (2012). The Key Role of Epigenetics in the
484 Persistence of Asexual Lineages. *Genetics Research International* 2012, 1-9.
485 Castresana, J. (2000). Selection of conserved blocks from multiple alignments
486 for their use in phylogenetic analysis. Mol. Biol. Evol. 17:540-552.
487 Chen, C., Chen, H., He, Y.H., and Xia, R. TBtools, a Toolkit for Biologists integrating
488 various biological data handling tools with a user-friendly interface. DOI:
489 https://doi.org/10.1101/289660 (2018)
490 Clark, M.K., House, M. A., Royden, L. H., Whipple, K., Burchfiel, B. C., Zhang, X.,
491 and Tang, W. (2005). Late Cenozoic uplift of southeastern Tibet. Geology 33:
492 525-528.
493 Crane, P.R., Friis, E.M., and Pedersen, K.R. (1994). Paleobotanical evidence on the
494 early radiation of magnoliid angiosperms. Plant Syst. Evol. 8:51-72.
495 De Bie, Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFÉ: a computational
496 tool for the study of gene family evolution. Bioinformatics 22:1269-1271.
497 Doyle, J.J., and Doyle, J.L. (1987). A rapid DNA isolation procedure for small

498 quantities of fresh leaf tissue. Phytochem Bull 19:11-15.
499 Drinnan, A.N., Crane, P.R., and Hoot, S.B. (1994). Patterns of floral evolution in the
500 early diversification of non-magnoliid dicotyledons (eudicots). Plant Syst. Evol.
501 8:93-122.
502 Eckert, C.G. (2001). The loss of sex in clonal plants. Evol. Ecol. 15:501-520.
503 El Baidouri, M., and Panaud, O. (2013). Comparative genomic paleontology across
504 plant kingdom reveals the dynamics of TE-driven genome evolution. Genome
505 Biol. Evol. 5:954-65.
506 Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and
507 flexible software for de novo detection of LTR retrotransposons. BMC
508 Bioinformatics 9:18.
509 Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole
510 genome comparisons dramatically improves orthogroup inference accuracy.
511 Genome Biol. 16:157.
512 Favre, A., Päckert, M., Pauls, S. U., Jähnig, S. C., Uhl, D., Michalak, I., and
513 Muellner-Riehl, A. N. (2015). The role of the uplift of the Qinghai-Tibetan
514 Plateau for the evolution of Tibetan biotas. Biol. Rev. 90:236-253.
515 Garner, B.A., Hand, B. K., Amish, S.J., Bernatchez, L., Foster, J.T., Miller, K.M.,
516 Morin, P.A., Narum, S.R., O'Brien, S.J., Roffler, G., et al. (2016). Genomics in
517 Conservation: Case studies and bridging the gap between data and application.
518 Trends Ecol. Evol. 31:81-83.
519 Gaut, B.S., Morton, B.R., McCaig, B.C., and Clegg, M.T. (1996). Substitution rate
520 comparisons between grasses and palms: synonymous rate differences at the
521 nuclear gene Adh parallel rate differences at the plastid gene rbcL. Proc. Natl.
522 Acad. Sci. 93:10274-10279.
523 Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I.,
524 Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length
525 transcriptome without a genome from RNA-Seq data. Nat. Biotechnol.
526 29:644-652.
527 Holt, C., and Yandell, M. (2011). MAKER2: an annotation pipeline and

genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491.

Hoorn, C., Mosbrugger, V., Mulch, A., and Antonelli, A. (2013). Biodiversity from mountain building. Nat. Geosci. 6:154-154.

Hoot, S.B., Magallón, S., and Crane, P.R. (1999). Phylogeny of basal eudicots based on three molecular data sets: *atpB*, *rbcL*, and 18S nuclear ribosomal DNA sequences. Ann. Missouri. Bot. Gard. 86:1-32.

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.

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.

Kim, H.T., Kim, J.S., Moore, M.J., Neubig, K.M., Williams, N.H., Whitten, 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/small single-copy region boundaries. PLoS ONE 10: e0142215.

Kim, S., Soltis, D.E., Soltis, 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.

558 Koren, S. Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M.
559 (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer
560 weighting and repeat separation. Genome Res. 27:722-736.
561 La Camera, S., Gouzerh, G., Dhondt, S., Hoffmann, L., Fritig, B., Legrand, M., and
562 Heitz, T. (2004). Metabolic reprogramming in plant innate immunity: the
563 contributions of phenylpropanoid and oxylipin pathways. Immunol. Rev. 198,
564 267-284.
565 Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H.-H., Rognes, T., and Ussery, D.W.
566 (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes.
567 Nucleic Acids Res. 35:3100-3108.
568 Landis, J. B., Soltis, D. E., Li, Z., Marx, H. E., Barker, M. S., Tank, D. C., and Soltis,
569 P. S. (2018). Impact of whole-genome duplication events on diversification rates
570 in angiosperms. Am. J. Bot. 105:348-363.
571 Lane, A.K., Augustin, M.M., Ayyampalayam, S., Plant A., Gleissberg, S., Di Stilio,
572 V.S., Depamphilis, C.W., Wong, G.K., Kutchan, T.M., and Leebens-Mack, J.H.
573 (2018). Phylogenomic analysis of Ranunculales resolves branching events across
574 the order. Bot. J. Linn. Soc. 187:157-166.
575 Lin, C.-S., Chen, J. J. W., Chiu, C.-C., Hsiao, H. C. W., Yang, C.-J., Jin, X.-H.,
576 Leebens-Mack, J., de Pamphilis, C.W., Huang, Y-T., Yang, L-H., et al. (2017).
577 Concomitant loss of NDH complex-related genes within chloroplast and nuclear
578 genomes in some orchids. Plant J. 90:994-1006.
579 Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection
580 of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25:955-964.
581 Lu, L.M., Chen, Z.D., and Lu, A.M. (2016). Will there ever be a tree of life that
582 systematists can agree on? Chinese Sci. Bull. 61:958-963.
583 Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L.L., and Hernández-Hernández,
584 T. (2015). A metacalibrated time-tree documents the early rise of flowering plant
585 phylogenetic diversity. New Phytol. 207:437-453.
586 Martín, M., and Sabater, B. (2010). Plastid *ndh* genes in plant evolution. Plant Physiol.
587 Bioch. 48:636e645.

588 Martin, M., Casano, L.M., Zapata, J.M., Guera, A., del Campo, E.M.,
589 Schmitz-Linneweber, C., Maier, R.M., and Sabater, B. (2004). Role of thylakoid
590 Ndh complex and peroxidase in the protection against photo-oxidative stress:
591 fluorescence and enzyme activities in wild-type and ndhF-deficient tobacco.
592 Physiol. Plant. 122:443e452.
593 Meng, K., Wang, E., and Wang, G. (2016). Uplift of the Emei Shan, western Sichuan
594 Basin: Implication for eastward propagation of the Tibetan Plateau in Early
595 Miocene. J. Asian Earth. Sci. 115:29-39.
596 Michael, T.P. (2014). Plant genome size variation: bloating and purging DNA. Brie.
597 Funct. Genomics 13:308-317.
598 Mirarab, S., and Warnow, T. (2015). ASTRAL-II: coalescent-based species tree
599 estimation with many hundreds of taxa and thousands of genes. Bioinformatics
600 31:44–52.
601 Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology
602 searches. Bioinformatics 29:2933-2935.
603 Nie, Z.X., Ye, Y.J., Bai, T.D., Xu, M., and Xu, L-A. (2017). Complete Chloroplast
604 Genome of *Pinus massoniana*(Pinaceae): Gene Rearrangements, Loss
605 of ndhGenes, and Short Inverted Repeats Contraction, Expansion. Molecules 22:
606 1528.
607 Otálora, M.A.G., Salvador, C., Martínez, I., and Aragón, G. (2013). Does the
608 reproductive strategy affect the transmission and genetic diversity of bionts in
609 cyanolichens? A case study using two closely related species. Microb. Ecol. 65:
610 517-530.
611 Qu, S.J., and Jiang, N. (2018). LTR_retriever: A Highly Accurate and Sensitive
612 Program for Identification of Long Terminal Repeat Retrotransposons. Plant
613 Physiol. 176:1410-1422.
614 Rowley, D.B. (1996). Age of initiation of collision between India and Asia: a review
615 of stratigraphic data. Earth Planet. Sc. Lett. 145:1-13.
616 Ruiz-Sanchez, E., Rodriguez-Gomez, F., and Sosa, V. (2012). Refugia and geographic
617 barriers of populations of the desert poppy, *Hunnemannia fumariifolia*

618 (Papaveraceae). Org. Divers. Evol. 12:133-143.
619 Rumeau, D., Peltier, G., and Cornac, L. (2007). Chlororespiration and cyclic electron
620 flow around PSI during photosynthesis and plant stress response. Plant Cell
621 Environ. 30:1041e1051.
622 Shafer, A.B.A., Wolf, J.B.W., Alves, P.C., Bergström, L., Bruford, M.W., Brännström,
623 I., Colling, Guy., Dalén, L., De Meester, L., Ekblom, R., et al. (2015). Genomics
624 and the challenging translation into conservation practice. Trends Ecol. Evol.
625 30:78-87.
626 Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M.
627 (2015). BUSCO: Assessing genome assembly and annotation completeness with
628 single-copy orthologs. Bioinformatics 31:3210-3212.
629 Soltis, D.E., Soltis, P.S., and Chase, M.W. (2000). Angiosperm phylogeny inferred
630 from 18S rDNA and atpB sequences. Bot. J. Linn. Soc. 133:381-461.
631 Stamatakis, A., Ludwig, T., and Meier, H. (2004). RAxML-III: A fast program
632 for maximum likelihood-based inference of large phylogenetic trees.
633 Bioinformatics 21:456-463.
634 Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006).
635 AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res.
636 34:W435-W439.
637 Stephenson, A.G., Good, S.V., and Vogler, D.W. (2000). Interrelationships among
638 inbreeding depression, plasticity in the self-incompatibility system, and the
639 breeding system of *Campanula rapunculoides* L. (Campanulaceae). Ann. Bot. 85:
640 211-219.
641 Sun, B.-N., Wu, J.-Y., Liu, Y.-S. (Christopher), Ding, S.-T., Li, X.-C., Xie, S.-P., Yan,
642 D-F., and Lin, Z.-C. (2011). Reconstructing Neogene vegetation and climates to
643 infer tectonic uplift in western Yunnan, China. Palaeogeogr. Palaeoclimatol.
644 Palaeoecol. 304:328-336.
645 Sun, Y., Moore, M. J., Lin, N., Adelalu, K. F., Meng, A., Jian, S., Yang, L.S., Li, J.Q.,
646 and Wang, H. (2017). Complete plastome sequencing of both living species of
647 Circaeasteraceae (Ranunculales) reveals unusual rearrangements and the loss of

the *ndh* gene family. BMC Genomics 18:592.

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.

Tenaillon, M.I., Hollister, J.D., and Gaut, B.S. (2010). A triptych of the evolution of plant transposable elements. Trends Plant Sci. 15:471-8.

Thomas, G. E., Geetha, K. A., Augustine, L., Mamiyil, S., and Thomas, G. (2016). Analyses between reproductive behavior, genetic diversity and pythium responsiveness in Zingiber spp. reveal an adaptive significance for hemiclonality. Front. Plant Sci. 7:1913.

Vogt, T. (2010). Phenylpropanoid biosynthesis. Mol. Plant 3:2-20.

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.

Wang, P., Scherler, D., Liu-Zeng, J., Mey, J., Avouac, J.-P., Zhang, Y., and Shi, D. (2014). Tectonic control of Yarlung Tsangpo Gorge revealed by a buried canyon in Southern Tibet. Science 346:978-981.

Wang, W., Lu, A.-M., Ren, Y., Endress, M. E., and Chen, Z.-D. (2009). Phylogeny and classification of Ranunculales: Evidence from four molecular loci and morphological data. Perspect. Plant Ecol. 11:81-110.

Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., Lee, T., Jin, H., Marler, B., Guo, H., et al. (2012). MCScanX: A toolkit for detection and evolutionary

678 analysis of gene synteny and collinearity. Nucleic Acids Res. 40:e49.
679 Wen, J., Zhang, J.-Q., Nie, Z.-L., Zhong, Y., and Sun, H. (2014). Evolutionary
680 diversifications of plants on the Qinghai-Tibetan Plateau. *Front. Genet.* 5, 1-16.
681 Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell,
682 A., Leroy, P., Morgante, M., Panaud, O., et al. (2007). A unified classification
683 system for eukaryotic transposable elements. Nat. Rev. Genet. 8:973-82.
684 Windig, J.J., and Engelsma, K.A. (2010). Perspectives of genomics for genetic
685 conservation of livestock. Conserv. Genet. 11:635-641.
686 Wornik, S., and Grube, M. (2010). Joint dispersal does not imply maintenance of
687 partnerships in lichen symbioses. Microb. Ecol. 59:150-157.
688 Xing, Y., and Ree, R.H. (2017). Uplift-driven diversification in the Hengduan
689 Mountains, a temperate biodiversity hotspot. Proc. Natl. Acad. Sci.
690 114:E3444-E3451 .
691 Xu, Z., and Wang, H. (2007). LTR_FINDER: An efficient tool for the prediction of
692 full-length LTR retrotransposons. Nucleic Acids Res. 35:W265-W268.
693 Yang, Y., Ma, T., Wang, Z., Lu, Z., Li, Y., Fu, C., Chen, X., Zhao, M., Olson, M.S.,
694 and Liu, J. (2018). Genomic effects of population collapse in a critically
695 endangered ironwood tree *Ostrya rehderiana*. Nature Communications 9:5449.
696 Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol.
697 Evol. 24:1586-1591.
698 Zou, X., and Ge, S. (2008). Conflicting gene trees and phylogenomics. J. Syst. Evol.
699 46:795-807.
700
701
702

703 Table 1 Genome assembly of *Kingdonia uniflora.*

Genome features	Contigs/Scaffolds
Total length, bp	1,004,656,313
Total number of contigs	2932
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

704

705

706

707

708

709

710

711

712

713

714 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	5.28E-07	1.59E-05
	Homologous recombination	17	56	2.87E-06	6.74E-05
	Mismatch repair	12	39	7.31E-05	0.000968079
	Sphingolipid metabolism	9	26	0.000282866	0.003082136
	DNA replication	12	50	0.000516123	0.005446878
	Nucleotide excision repair	14	69	0.000789672	0.007779277
	Peroxisome	16	87	0.000878589	0.008505414
	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	2.06E-09	1.57E-07
	Phenylpropanoid biosynthesis	5	157	2.06E-07	6.24E-06
	Starch and sucrose metabolism	5	202	6.95E-07	1.96E-05
	Biosynthesis of secondary metabolites	6	1,076	0.00020813	0.001073645

	Metabolic pathways	6	1,910	0.004094269	0.015173186

715

716

717 **Figure 1.** Comparative analyses of genomic features between *Kingdonia uniflora* and
718 *Aquilegia coerulea*. Tracks from inside to outside are collinearity between both
719 genomes, number of chromosomes/scaffolds, gene density, GC content and TE
720 density.
721 **Figure 2A**. Insertion time distribution of LTR retrotransposons.
722 **Figure 2B**. Distribution of synonymous substitution rates (*Ks*) for pairs of
723 paralogs/orthologs in/between *K. uniflora* and *C. agrestis.*
724 **Figure 2C.** Dot plots of paralogs identified across contigs in the *K. uniflora* genome.
725 **Figure 2D.** Venn diagram showing unique and shared gene families between
726 genomes of *K. uniflora* and four other Ranunculales species.
727 **Figure 3.** Dated phylogeny for 12 plant species with *Oryza* as an outgroup. A time
728 scale is shown at the bottom, and red points in some nodes indicate fossil
729 calibration points.
730 **Figure 4**. Length comparison of *ndh* genes between *K. uniflora* and *C. agrestis.*
731
732

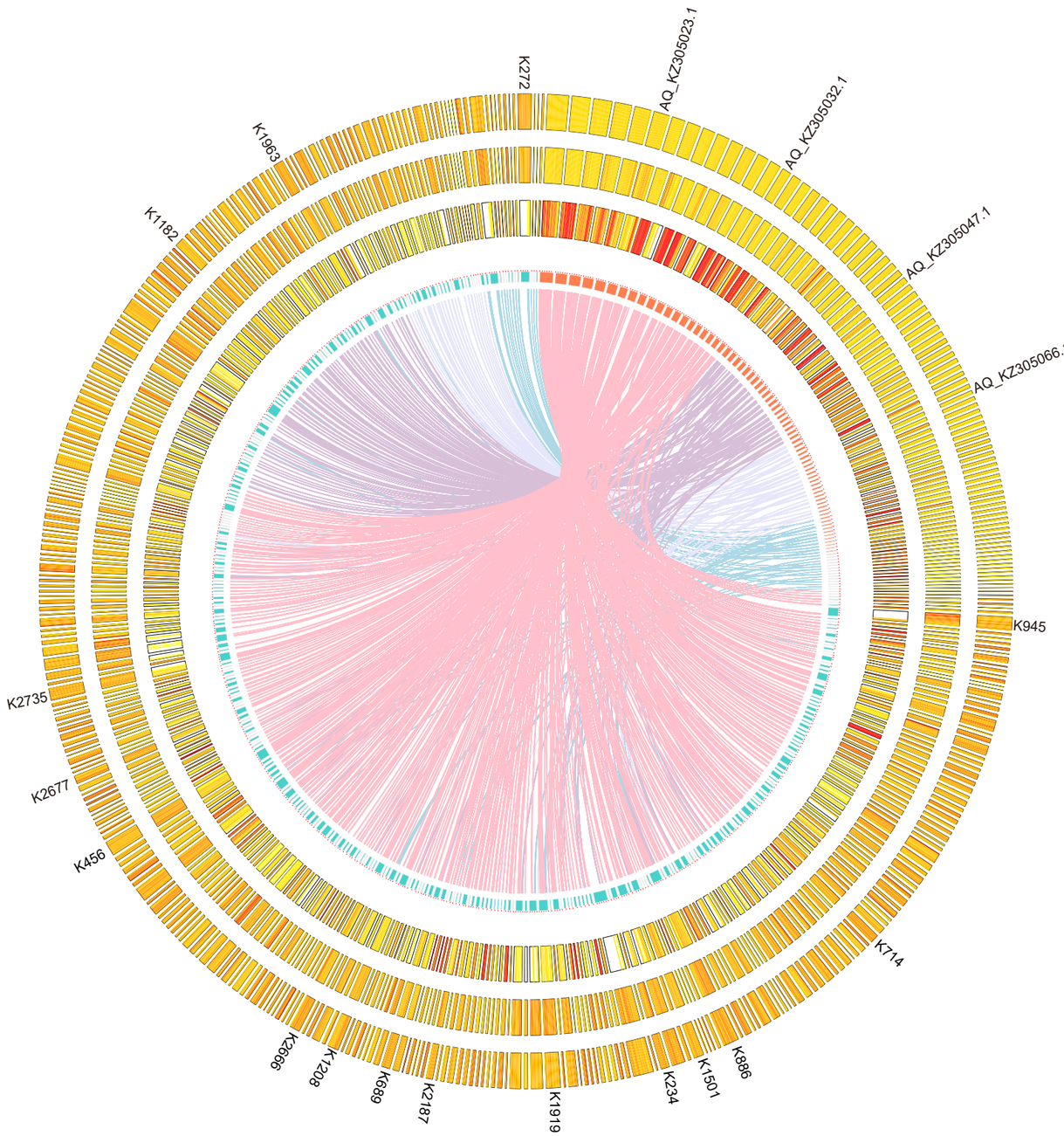
K272
AQ_KZ305023.1
AQ_KZ305032.1
AQ_KZ305047.1
AQ_KZ305066.1
K945
K714
K886
K1501
K234
K1919
K2187
K689
K1208
K2666
K456
K2677
K2735
K1182
K1963

Figure 1

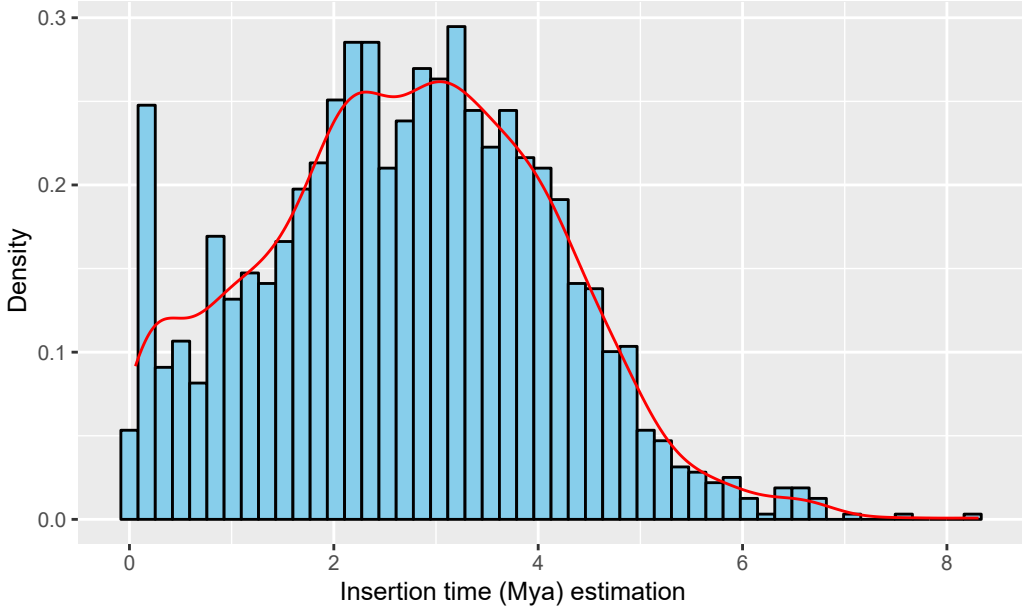

Figure 2A

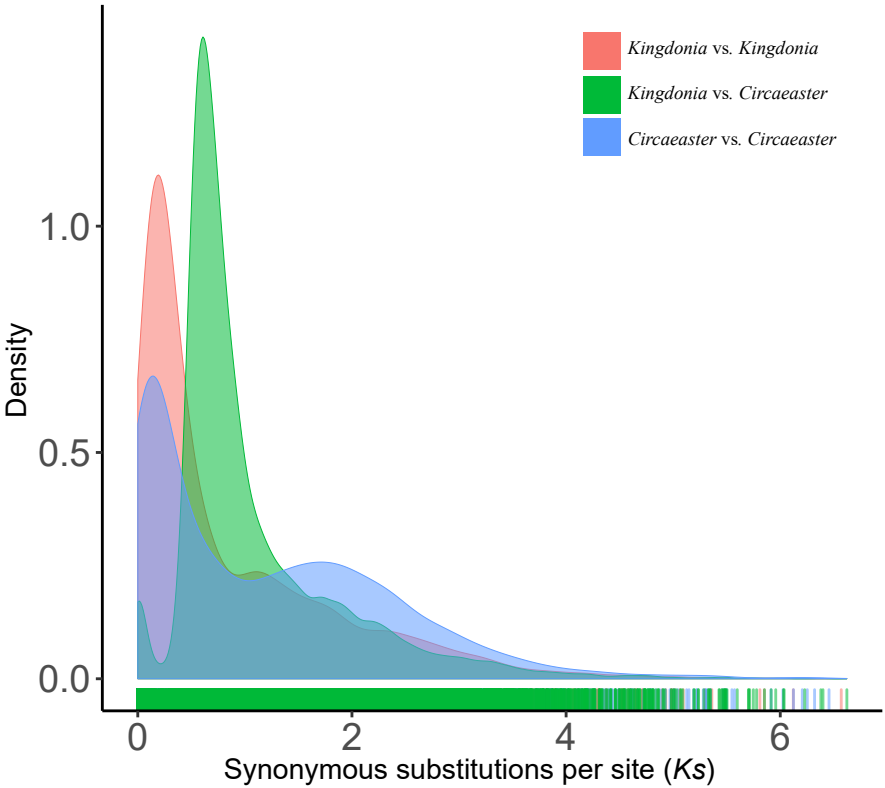
Kingdonia vs. Kingdonia
Kingdonia vs. Circaeaster
Circaeaster vs. Circaeaster
Density
1.0
0.5
0.0
0
2
4
6
Synonymous substitutions per site (Ks)

Figure 2B

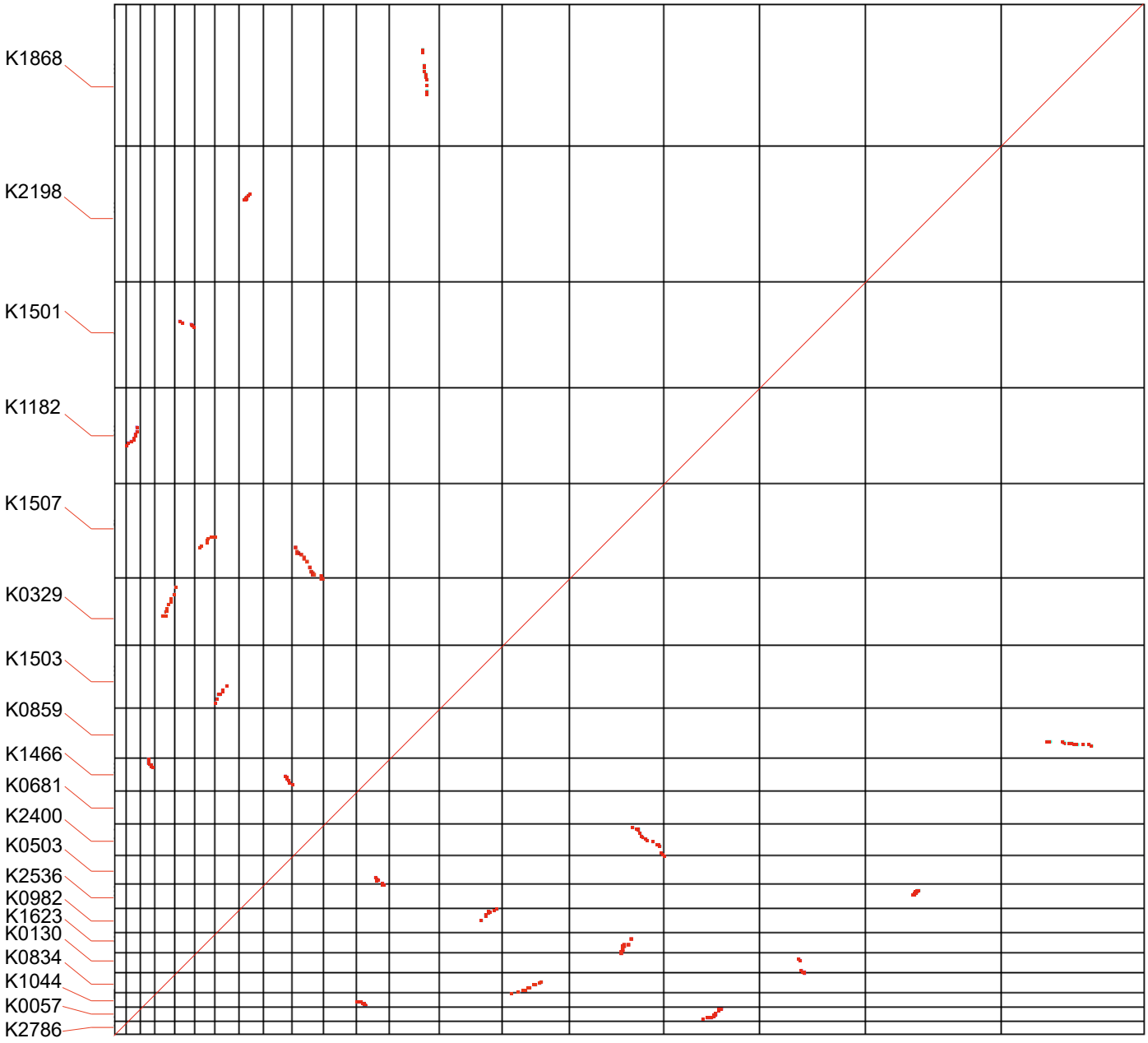

Figure 2C

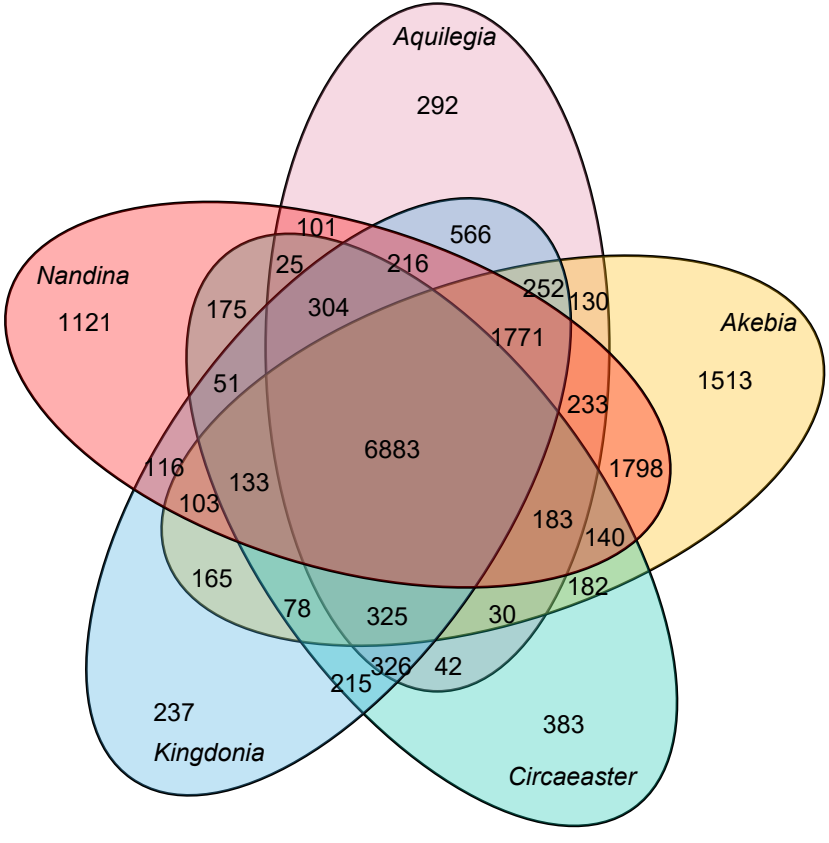
Aquilegia
292
Nandina
1121
Akebia
1513
Kingdonia
237
Circaeaster
383
101
566
25
216
252
130
175
304
1771
51
233
6883
116
133
1798
103
183
140
165
182
78
325
30
326
42
215

Figure 2D

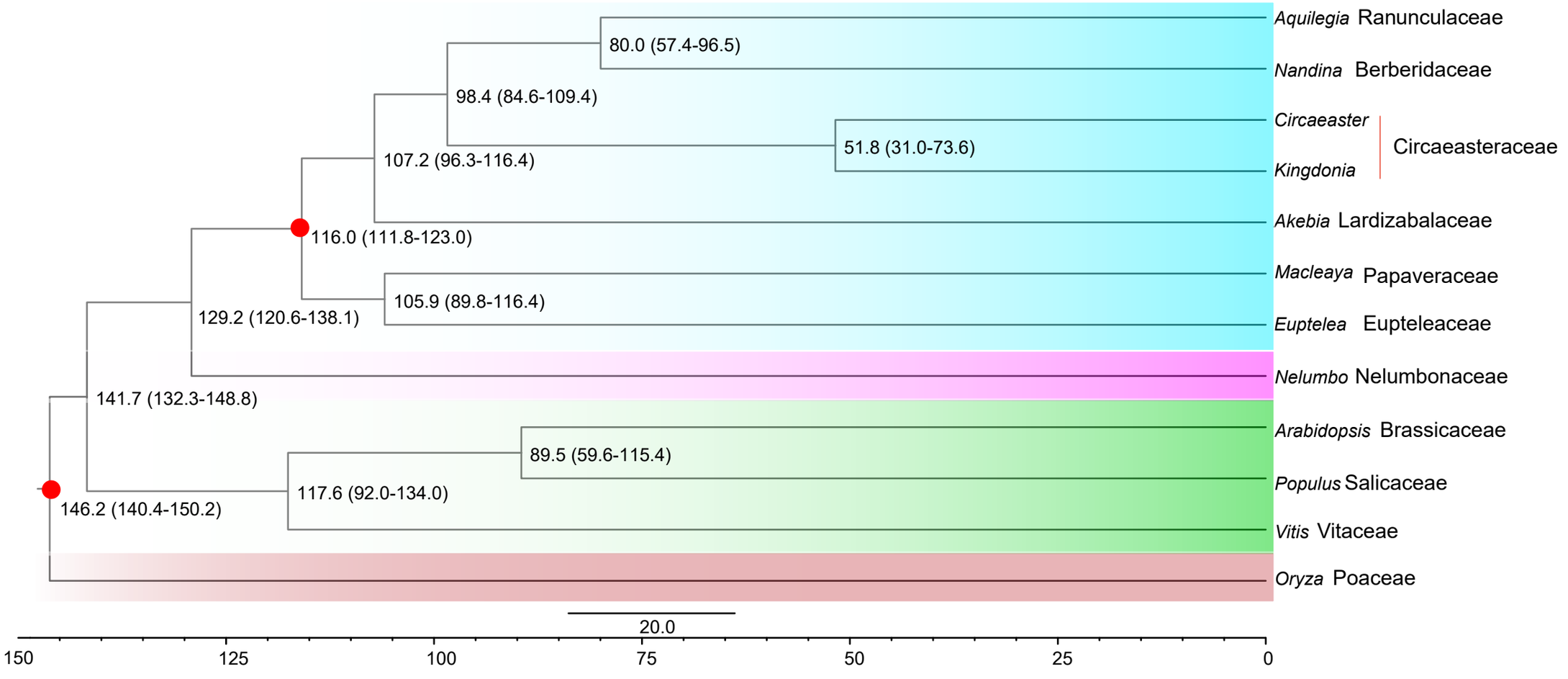
80.0 (57.4-96.5)
98.4 (84.6-109.4)
51.8 (31.0-73.6)
107.2 (96.3-116.4)
116.0 (111.8-123.0)
105.9 (89.8-116.4)
129.2 (120.6-138.1)
141.7 (132.3-148.8)
89.5 (59.6-115.4)
117.6 (92.0-134.0)
146.2 (140.4-150.2)
Aquilegia Ranunculaceae
Nandina Berberidaceae
Circaeaster
Kingdonia
Circaeasteraceae
Akebia Lardizabalaceae
Macleaya Papaveraceae
Euptelea Eupteleaceae
Nelumbo Nelumbonaceae
Arabidopsis Brassicaceae
Populus Salicaceae
Vitis Vitaceae
Oryza Poaceae
20.0
150
125
100
75
50
25
0

Figure 3

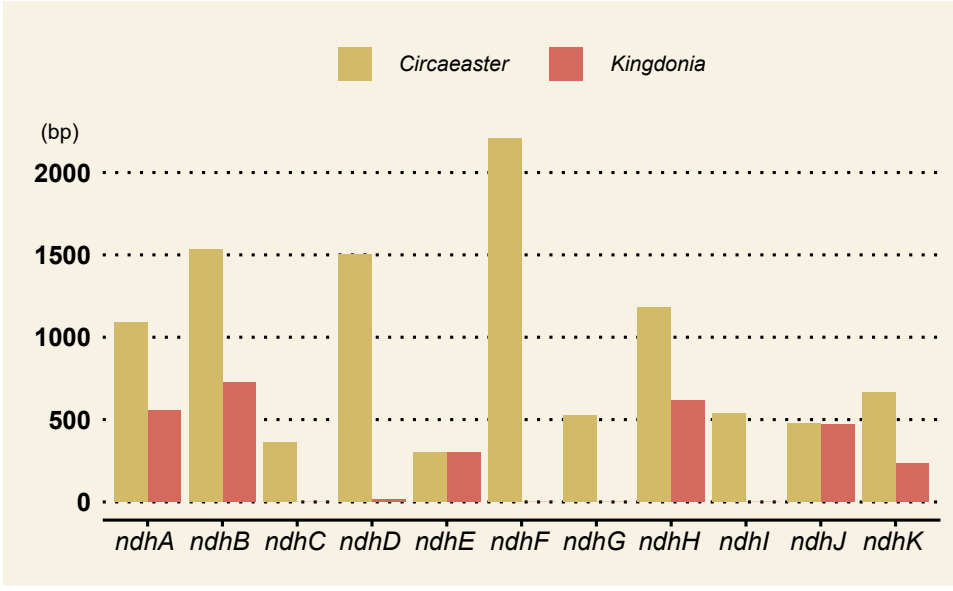
Circaeaster
Kingdonia
(bp)
2000
1500
1000
500
0
ndhA
ndhB
ndhC
ndhD
ndhE
ndhF
ndhG
ndhH
ndhI
ndhJ
ndhK

Figure 4