Population Transcriptomes Reveal the Interspecific Adaptive Genetic Differentiation of Liriodendron by Landscape Genetics

Yufang Shen
Nanjing Forestry University https://orcid.org/0000-0001-8070-3635
Hui Xia
Nanjing Forestry University
Zhonghua Tu
Nanjing Forestry University
Yaxian Zong
Nanjing Forestry University
Lichun Yang
Nanjing Forestry University
Huogen Li (hgli@njfu.edu.cn)
Nanjing Forestry University https://orcid.org/0000-0001-6455-9346

Research article

Keywords: Adaptive genetic differentiation, Population transcriptomes, Landscape genetics, Adaptive gene, Liriodendron, RNA-seq

Posted Date: September 8th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-62166/v1

Abstract

Background: Adaptive genetic differentiation is a hotspot in the research of speciation mechanisms in evolutionary biology. Genomic resources are important for detecting ecological adaptive evolution of non-model plants. Using RNA-seq for non-model plants is a good approach to obtain their genomic resources. The combination of population transcriptome resources and environmental data can provide insights into the genetic mechanism of adaptive genetic differentiation.

Results: Based on the population transcriptome data, we investigated the spatial distribution of genetic variations in *Liriodendron* to detect relationships between ecological factors and genetic differentiation. Environmental data and genetic variations from 17 populations were integrated to detect the population structure, adaptive genes and key environmental factors that shape the population genetic structure by landscape genetic approach. Here, we identified 16592 high-quality single nucleotide polymorphisms (SNPs). The population structure analysis results showed that 17 populations were divided into three groups: *L. tulipifera*, eastern group and western group of *L. chinense*. Redundancy analysis and latent factor mixed model analysis suggested that precipitation seasonality, precipitation in the driest quarter, diurnal temperature, and solar radiation in May were closely associated with the adaptive genetic differentiation of *Liriodendron*. Ecological niche differentiation analysis implied significant ecological niche divergence between *L. chinense* and *L. tulipifera* habitats. In total, 858 environment-related loci were identified, which were associated with 464 genes. Pathway enrichment analysis revealed that these genes were significantly enriched in multiple biological pathways. Related studies confirmed that these biological pathways play vital roles in plant growth, development, stress, immune response and photosynthesis.

Conclusions: Our research provided empirical evidence that environmental factors may play a key role in driving adaptive genetic differentiation of species. Furthermore, the combination of population transcriptome resources and environmental datasets provides new insights into the study of adaptive genetic differentiation of species.

Background

Adaptive genetic differentiation is a hotspot in the research of speciation mechanisms in evolutionary biology [1–7]. Ecological adaptive differentiation is a crucial mechanism of natural selection in the process of speciation, especially in plants [8–10]. Due to the limited propagation distance of plant pollen and seeds, it is impossible to seek advantages and prevent harm through rapid migration, similar to animals. Therefore, ecological adaptive differentiation is more common in plants than in animals [11–15]. Ecological processes are key to adaptive differentiation and speciation when the selection of environmental differences leads to the continuous evolution of barriers to gene flow between populations [1]. Adaptive differentiation refers to the differentiation of species from heterogeneous environments under selection pressure [16]. In response to these selective pressures, plants will undergo adaptive evolution. During the long-term adaptive evolution of plants, signal perception, signal transduction, and

gene expression eventually manifest phenotypic, stress resistance and phenological changes [17]. The essence behind these changes is the adaptive evolution of genes involved in the formation of these traits. Identifying these genes related to ecological adaptation and ecological factors that affect the local adaptation of plants will provide insight into the genetic mechanism of ecological adaptation [18].

In recent years, the complex evolutionary relationship and interaction mechanism between biological evolution and the environment have been widely confirmed in molecular ecology. The environment can affect the neutral evolutionary process (mutation, drift and gene flow) through various evolutionary factors (such as ecological isolation and chemical mutagenesis), thereby affecting the evolution of species or populations [19]. Furthermore, continuous environmental selection pressure can also act on the fixation process of specific genes, thus affecting the adaptive evolution of species or populations and even leading to the formation of new species [20–21]. Adaptive evolution is an evolutionary force that has received wide attention in recent years. According to relevant studies, it has become one of the most important factors affecting the evolution of species in addition to mutation, drift and gene flow, and its contribution to the evolution of species generally reaches 5–10% [22]. However, it is difficult for researchers to comprehensively understand the mechanism of biological evolution and its relationship with the environment due to the limitations of the current human cognitive level and technological means. In addition, there is no comprehensive and systematic method to reveal the genetic mechanism of adaptive genetic differentiation due to the short period of research on ecological adaptation and different research focuses [23].

With the development of high-throughput sequencing technologies, it is convenient to obtain genomic data from populations. Combining environmental factor data with genomic data and using population genetics statistical methods will provide researchers with rich information about the correlation between environmental characteristics and species adaptive evolution and details of their interactions [24–27]. Although genomic data cover a wealth of genetic information, for forest tree species and non-model species with large genomes, population genome sequencing also attracts a huge cost. For researchers, using RNA-seq strategy for forest tree species and non-model species is a good potential approach to obtain their genomic resources. Therefore, using population transcriptome sequencing may be a suitable strategy for adaptive evolution studies. First, transcriptome data have the advantage of moderate data size compared with genomic data and traditional genetic marker data, and the exon regions of genes are mainly obtained [23]. Second, transcriptome data can be used for various downstream analyses, such as genetic marker exploiting, candidate gene identification, QTL mapping, and comparative genomics. Third, transcriptome sequencing is more economical than whole genome sequencing. Currently, research on the ecologically adaptive evolution of species through population transcriptome sequencing strategies is increasingly favored by investigators. Studies have been conducted on the adaptive evolution of species using population transcriptome strategies. In these studies, the researchers used transcriptome sequencing to identify genes involved in adaptive evolution and even found that some species have undergone adaptive differentiation [28–32]. These studies further confirm that it is feasible to study the adaptive evolution of species using transcriptome sequencing strategies. At present, although the transcriptome method has been recognized and developed to study the adaptive evolution of species,

most of the studies are mainly on animals or model plants, whereas few studies have been conducted on non-model plants, especially tall forest trees.

Nearly half of the Earth's land is covered by forests, and many tree species play foundations or key roles [33]. Forest trees not only provide survival resources for other organisms but also play vital roles in maintaining ecological balance. Approximately 3/4 of the terrestrial biomass on earth comes from forests, and forests are closely related to the carbon budget in the atmosphere [34]. Therefore, the stability of forest ecosystems plays a key role in the survival of other organisms. Ecological adaptive differentiation is a very important mechanism of natural selection in the process of speciation, especially in the formation of forest trees [8–10]. Ecological adaptive differentiation affects the stability of forest ecosystems by changing the diversity of forest trees [34–37]. Therefore, it is important to understand the genetic mechanism of the ecological adaptation differentiation of forest trees.

Liriodendron is a deciduous broad-leaved tree of Magnoliaceae. The genus includes two extant sister species, namely, *Liriodendron chinense* and *Liriodendron tulipifera* [38–40]. These two sister species represent a well-known example of an eastern Asian and North American disjunct distribution of temperate deciduous forest species [41]. *L. chinense* occurs in small, widely scattered populations in subtropical China and northern Vietnam, while *L. tulipifera* is prolific throughout the southeastern United States [40–42]. Extensive fossil evidence indicates that *Liriodendron* was common throughout the Northern Hemisphere in the Tertiary period [39, 43–45], while the genus currently consists of only two sister species. *Liriodendron* is a shade-intolerant tree that requires adequate sunlight and precipitation during the growing season [46]. *Liriodendron* generally blooms in May. *L. tulipifera* is a fasting-growing tree indigenous to southeastern North America but is planted worldwide because acclimates to various environments and sequesters large amounts of carbon dioxide due to its developed root system [47]. Compared with *L. tulipifera*, *L. chinense* grows more slowly [48]. Due to endangered habitats, intense interspecific competition, low seed viability, and artificial interference, *L. chinense* is restricted to southern areas of the Yangtze River of China and has been listed as a secondary threatened species in China [40, 42]. Previous studies found that the adaptation of the two species is very different [49–51]; consequently, these are ideal materials for studying the evolution and adaptive differentiation of East Asia-North American discontinuously distributed plants. Because there is no research on the adaptive evolution of *Liriodendron*, the genetic mechanism of the interspecific adaptive genetic differentiation of this genus is unclear.

Although the genetic differentiation between and within species of *Liriodendron* has been identified using genome-wide markers, only one individual was used per population, and most of the SNP markers obtained were not in the exon regions of genes [52]. In this study, 80 *Liriodendron* individuals from 17 populations were sampled, with 2–5 individuals per population. Here, we report a population-level transcriptome study to investigate the relationship between interspecies genetic differentiation and the adaptive divergence of *Liriodendron* using a landscape genetics approach. First, we obtained at least 8 Gb of clean data for each of the 80 individuals and implemented SNP calling of cDNA. These single nucleotide polymorphisms (SNPs) were used to examine the population genetic structure. Then,

constrained linear ordination analysis, i.e., redundancy analysis (RDA), was performed to identify environmental factors related to adaptive genetic differentiation. In addition, spatial evolutionary and ecological vicariance analysis (SEEVA) was performed to detect ecological niche divergence between groups of *Liriodendron*. We identified the potential adaptive loci by performing correlation analysis between environmental factors and genetic variations in the background levels of population structure. Finally, the genes where these adaptive loci are located were identified, and these genes were subjected to pathway enrichment analysis to mine genes related to environmental adaptation. Furthermore, the characteristics of leaf phenotypes were also investigated and analyzed. The study aimed to (1) identify population genetic structure of *Liriodendron*, (2) evaluate the role of environmental factors on the adaptive genetic differentiation of *Liriodendron*, (3) exploit adaptive genes, and provide an empirical reference for studying the genetic basis underlying the adaptation. The combination of population transcriptome resources and environmental data is expected to provide new insights into the study of adaptive genetic differentiation of species.

Results

RNA-seq and SNP calling

The cDNA library of 80 individuals of *Liriodendron* was successfully constructed, and RNA sequencing generated a total of 832.16 Gb of clean data, with each individual having at least 8 Gb of clean data (Additional file 1: Table S1). With the genome sequences of *L. chinense* as the reference genome [52], 16592 high-quality SNPs were identified by SNP calling with strict filtering.

Population genetic structure

Based on the 13990 neutral SNPs (removing outlier loci identified by Arlequin and BayeScan from 16592 SNPs) and using Admixture, we observed three separate ancestry groups: *L. tulipifera* in North America (NA), *L. chinense* in eastern China (CE) except for the XN population, and *L. chinense* in western China (CW) (Fig. 1A). The Admixture plots and cross-validation error of each K (K = 2 to 10) are shown in Additional file 2: Fig. S1 and Additional file 3: Fig. S2, respectively. For the PCA (principal component analysis), the first principal component (PC1) separated *L. chinense* from *L. tulipifera* (P = 1.68 E-15), while the second principal component (PC2) separated CE from CW except for the XN population (P = 4.27 E-36) (Fig. 1B). The neighbor-joining tree further revealed that 17 populations were divided into three clusters, NA, CE, and CW, except for the XN population of CE clustering into CW (Fig. 1C). NA was genetically closer to CW than to CE. The AMOVA results also showed that genetic variation mainly existed among groups (68.78%, F_{CT} = 0.69, P < 0.00001; Table 1). The genetic diversity analysis results showed that the genetic diversity of *L. tulipifera* populations is higher than that of *L. chinense* populations as a whole (Table 2). The genetic differentiation statistics results showed a high degree of genetic differentiation between *L. tulipifera* and *L. chinense* (Fig. 1D) (*Fst* = 0.86111). In addition, obvious genetic divergence was observed between CE and CW (*Fst* = 0.16935). Furthermore, we found that the genetic

differentiation between NA and CE was slightly higher (*Fst* = 0.76841) than that between NA and CW (*Fst* = 0.75842).

Table 1 Hierarchical AMOVAs for SNP variations surveyed in *Liriodendron*.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation	Fixation indices	*P*-value
Among groups	2	142505.59	1356.35	68.78	FCT = 0.69	P < 0.00001
Among populations within groups	14	19645.82	94.26	4.78	FSC = 0.15	P < 0.00001
within populations	143	74569.45	521.46	26.44	FST = 0.74	P < 0.00001

d.f.: degree of freedom.

Table 2 Details of the population locations, sample sizes, genetic diversity, voucher numbers and depositary of 17 populations of *Liriodendron*.

Pop ID	locations	Lat	Lng	Alt	*N*	N_A	*PPA*	H_E	Voucher numbers & depositary
NA	United States	-	-	-	25	6535	39.4	0.084	-
MSL	Missouri	38.87	-92.23	245	5	2772	16.7	0.060	00081690, PE
BK	North Carolina	36.07	-77.15	8	5	3876	23.3	0.078	00318635, NAS
LYS	Louisiana	30.42	-91.02	18	5	3284	19.8	0.071	01899157, PE
NK	South Carolina	33.83	-80.82	65	5	3961	23.9	0.079	00318637, NAS
ZZY	Georgia	34.64	-84.76	230	5	3628	21.8	0.075	00318634, NAS
CHN	China	-	-	-	55	10477	63.1	0.073	-
CE	Eastern China	-	-	-	24	4324	26	0.058	-
HS	Huangshan, Anhui	30.17	116.1	1250	5	2329	14.0	0.050	00318611, NAS
WYS	Wuyishan, Jiangxi	27.92	117.8	1000	5	1474	8.9	0.036	00081629, PE
SY	Songyang, Zhejiang	28.50	119.6	138	5	2465	14.8	0.052	00081550, PE
LS	Lushan, Jiangxi	29.53	116	1167	4	2345	14.1	0.054	20010020006, NJFU
XN	Xianning, Hubei	29.8	114.2	68	5	3023	18.2	0.062	20010020089, NJFU
CW	Western China	-	-	-	31	8457	50.9	0.072	-
XY	Xuyong, Sichuan	28.2	105.5	800	5	2148	12.9	0.047	0028959, CDBI
SN	Suining, Hunan	26.33	110.2	1500	4	1904	11.5	0.049	00081582, PE
EX	Exi, Hubei	30.3	109	1180	5	3448	20.2	0.064	00081593, PE
SZ	Sangzhi, Hunan	29.15	110.2	1200	5	3313	19.9	0.064	00318629, NAS
YY	Youyang, Sichuan	28.82	108.8	890	5	2780	16.7	0.060	01859555, PE
ST	Songtao,	26.8	109.5	1050	2	1683	10.1	0.055	00875730, PE

	Guizhou								
LP	Liping, Guizhou	26.34	109.19	421	5	2130	12.8	0.048	20010020014, NJFU

Pop ID: population code, Lat: latitude, Lng: longitude, Alt : altitude, *N*: number of individuals, N_A: number of polymorphic alleles, *PPA*: percentage of polymorphic alleles, H_E: Nei's 1987 measure of nucleotide diversity, NA: *L. tulipifera* populations, CHN: *L. chinense* populations, CE: eastern populations of *L. chinense*, CW: western populations of *L. chinense*; PE: Herbarium code of Beijing Institute of Botany; NAS: Herbarium code of Jiangsu Institute of Botany; NJFU: Herbarium code of Nanjing Forestry University; CDBI: Herbarium code of Chengdu Institute of Biology.

Fig. 1. Population genetic structure of *Liriodendron*. (A) The Admixture plots (K = 2 and 3) based on 13990 neutral loci. (B) The PCA result based on 16592 SNPs identified from *Liriodendron*. The 17 colors correspond to 17 populations from 3 groups. (C) Neighbor-joining phylogenetic tree of *Liriodendron* based on 16592 SNPs, with the evolutionary distances measure by *p*-distances with phylip. The three colors correspond to the 3 groups. (D) Matrix of the pairwise *Fst* of 17 populations in *Liriodendron*.

To provide insight into the roles of environmental factors in population genetic structure and their contribution to this population structure, RDA was performed. The RDA results are shown in Table 3 and Fig. 2. First, the interactive-forward-selection analysis detected the top 16 most influential environmental variables from 55 environmental variables. The 16 environmental variables included two temperature variables (bio2 and bio7), three precipitation variables (bio13, bio15 and bio17), 6 solar radiation variables (srad1, srad2, srad5, srad6, srad7 and srad10) and 5 wind speed variables (wind2, wind7, wind8, wind9 and wind10). The RDA results showed that the correlation between 16592 alleles and 16 environmental variables was 1.000 on axes 1 and 2. The eigenvalue ratios of axes 1 and 2 were 32.10% and 8.38%, respectively. Sixteen environmental factors divided the 17 populations of *Liriodendron* into three groups. Axis 1 (RDA 1) separated *L. chinense* from *L. tulipifera*, while axis 2 (RDA 2) separated CE from CW except for the XN population (Fig. 2). Among the 16 environmental factors, 8 environmental factors had a significant influence on the population genetic structure of *Liriodendron* (Table 3). The eight environmental factors included one temperature variable (bio2), three precipitation variables (bio13, bio15 and bio17), three solar radiation variables (srad2, srad5 and srad6) and one wind speed variable (wind2). The results suggested that precipitation and light play key roles in shaping the population structure of *Liriodendron*.

Table 3 Effects of 16 environmental variables on and their explained contributions to the population genetic structure of *Liriodendron*.

variables	Explains%	pseudo-*F*	*P*-value	*P*-value correction
bio15	30.4	6.6	0.002	0.004
srad6	30.0	6.4	0.002	0.004
srad5	29.8	6.4	0.002	0.004
bio2	28.9	6.1	0.002	0.004
bio13	21.7	4.2	0.002	0.004
bio17	21.5	4.1	0.002	0.004
srad2	20.1	3.8	0.002	0.004
wind2	19.9	3.7	0.002	0.004
srad7	17.6	3.2	0.006	0.01067
bio7	15.2	2.7	0.008	0.0128
srad10	10.7	1.8	0.054	0.07733
wind10	10.1	1.7	0.066	0.08123
wind9	9.8	1.6	0.058	0.07733
srad1	7.4	1.2	0.208	0.23771
wind8	6.3	1.0	0.402	0.402
wind7	6.0	1.0	0.396	0.402

Explains%: The contributions of variables to the population genetic structure.

Fig. 2. Redundancy analysis of *Liriodendron* showing the relative contributions of 16 environmental variables to the population genetic structure. The biplot depicts the eigenvalues and the lengths of eigenvectors for the RDA, and the color gradient corresponds to genetic diversity.

Ecological niche differentiation analysis

The SEEVA results showed that 22 ecological factors undergo significant divergence (D > 0.75, P < 0.0016) between *L. chinense* and *L. tulipifera* habitats (Fig. 3; Additional file 4: Table S2). In addition, 6 ecological factors have diverged significantly (D > 0.75, P < 0.0016) in the eastern and western habitats of *L. chinense* (Fig. 3; Additional file 4: Table S2).

Fig. 3. The divergence index and impact index shown as bar diagrams for elevation and 55 climatic variables between sister species and sister groups. (A) The divergence index and impact index shown as bar diagrams for elevation and 55 climatic variables between *L. chinense* and *L. tulipifera*. (B) The

divergence index and impact index shown as bar diagrams for elevation and 55 climatic variables between east and west groups.

Leaf shape analysis

The results of leaf shape analysis showed that the leaf shape variation was the most abundant in CW, and the variation range covered almost all the leaf variation in CE and NA (Fig. 4). However, CE and NA have common leaf variations and unique leaf variations. Furthermore, the statistical results of the number of cracks in leaves showed that the number for *L. chinense* ranged from 3 to 6, and 92.14% of the leaves had 3 cracks. However, the number of cracks in the leaves of *L. tulipifera* ranged from 3 to 9, among which 5 and 3 were common, accounting for 63.10% and 16.67%, respectively (Additional file 5: Table S3).

Fig. 4. Leaf shape variation distributions in three groups of *Liriodendron* (X-axis: The ratio of x1 and x2; Y-axis: The ratio of y1 and y2; x1: The vertical distance from the right lateral sinus to the primary vein. x2: The vertical distance from the right lateral lobe to the primary vein. y1: The vertical distance from the right lateral sinus to the leaf base. y2: The vertical distance from the right apical lobe to the leaf base) (Confidence interval: 95%).

Characterization of environment-associated loci

There were 1917 and 1808 outlier sites identified by Arlequin and BayeScan, respectively (Additional file 6: Table S4 and Additional file 7: Table S5). In total, 1123 outlier loci were identified by both Arlequin and BayeScan. The results of the environment-associated analysis showed that 858 EAL associated with at least one environmental factor were identified by LFMM (Additional file 8: Table S6). Furthermore, LFMM analysis showed that precipitation seasonality (bio15), elevation, precipitation in the driest quarter (bio17), precipitation in the warmest quarter (bio18), the mean diurnal range of temperature (bio2), solar radiation in May (Srad5) and water vapor pressure in July (vapr7) were associated with the highest numbers (> 70) of EAL (Additional file 8: Table S6).

GO and KEGG enrichment analysis

The genomic contexts of the 858 EAL were determined based on the *L. chinense* reference genome [52]. We found that 667 EAL were associated with 464 genes, and 191 EAL corresponded with the intergenomic regions (Additional file 8: Table S6). The results of GO annotation showed that these genes represented a broad range of biological processes, molecular functions and cellular components (Additional file 9: Fig. S3). Furthermore, some genes are considered to be involved in the response to biotic and abiotic stresses, phenology, disease resistance and hormonal responses (Additional file 10: Table S7), such as *ARF1* (ethylene and cytokinin responses) [53], *RAP2-7* (ethylene-responsive transcription factor), *RP/EB family member 1B-like* (response to wounding) [54], *FRIGIDA-like* (flowering time) [55], *UVR8* (photomorphogenesis and stress acclimation) [56] and *RGA2-like* (disease resistance protein) [57], were identified in other plants. In addition, we found some genes involved in the response to

biotic and abiotic stresses, stimuli, immune responses, growth and development, and programmed cell death by Gene Ontology annotation (Additional file 11: Table S8).

The enriched GO terms biological process, cellular component and molecular function are shown in Additional file 9: Fig. S3. Genes of GO terms such as ‘inorganic diphosphatase activity’, ‘protein transport’, ‘deoxyribonucleotide metabolic process’, ‘protein secretion’, and ‘ubiquitin-like protein-specific protease activity’ were highly enriched. Pathway enrichment analysis revealed that genes involved in plant-pathogen interactions, phosphatidylinositol signaling systems, ubiquitin-mediated proteolysis, carotenoid biosynthesis, alpha-linolenic acid metabolism and phagosomes were significantly enriched (Fig. 5). Related studies have shown that these biological pathways play vital roles in plant growth, development, stress, immune response and photosynthesis.

Fig. 5. The plots of the KEGG enrichment analysis results.

Discussion

Origin and population genetic differentiation of *Liriodendron*

Based on allozyme polymorphisms, cpDNA, and fossil evidence, Parks and Wendel first proposed genetic differentiation between *L. tulipifera* and *L. chinense* [39] and suggested that the separation between *L. chinense* and *L. tulipifera* occurred 10-16 Ma ago [39]. In our study, significant genetic differentiation was also detected based on SNPs from transcriptome data in *Liriodendron*. Furthermore, we also found an obvious east-west genetic divergence in *L. chinense* (Figs. 4A-C). This result is consistent with previous reports [52, 58-60]. Additionally, Yang used cpDNA and inferred that the divergence time of the east-west lineage was 0.443 Ma ago [60]. In this study, we found that the genetic differentiation between NA and CE was slightly higher (*Fst* = 0.76841) than that between NA and CW (*Fst* = 0.75842). The neighbor-joining tree further revealed that NA is genetically closer to CW than to CE (Fig. 1C). Interestingly, the results of leaf shape analysis showed that the leaf shape variation was the most abundant in CW, and the variation range covered almost all of the leaf variations in CE and NA (Fig. 4). The results suggested that CW may be an older ancestral group. Both CE and NA originated from CW.

In addition, geological paleontologists found *Liriodendron* pollen fossils of the early Cretaceous Aptian-Albian in the white crane cave in GuangZhou, China [61]. This is the oldest fossil record of *Liriodendron* ever found. Later, researchers found more *Liriodendron* fossils of the late Cretaceous in Asia, Europe and North America [62, 63]. Fossil evidence has shown that *Liriodendron* is widely distributed across the Northern Hemisphere and reached its maximum distribution during the Tertiary period [39, 43-45]. Therefore, we inferred that *Liriodendron* probably originated in southern China and moved westwards via West Asia, Europe, and the North Atlantic Road Bridge to North America and eastwards via Japan, Russia, and the Bering Land Bridge to North America. The genus was widespread in the Northern Hemisphere in the Late Cretaceous and reached its maximum distribution in the Tertiary period. With mid-Cenozoic global cooling [64], *Liriodendron* species migrated to low and middle latitudes, which caused the extinction of populations in Europe and Siberia. The populations that were continuously distributed in the

Northern Hemisphere were divided into two groups and retreated to eastern Asia and North America. This resulted in the interruption of gene flow and the genetic differentiation of *Liriodendron* between East-Asia and North America. Subsequently, with the repeated glacial and interglacial cycles during the Quaternary period, North America, Europe and East Asia were covered by ice sheets at mid-high latitudes [65, 66]. Ultimately, the populations of *Liriodendron* mainly retreated to the subtropical areas of China and southeastern United States, forming the current geographical pattern. Furthermore, the statistical results of the number of cracks of leaves showed that leaves with three cracks are the primary leaf shape in *L. chinense*, compared with five cracks in *L. tulipifera* (Fig. 3; Additional file 5: Table S3). This result confirmed the obvious leaf shape differentiation between *L. chinense* and *L. tulipifera*.

For *L. chinense*, the long-term in situ refugia pattern may have caused genetic differentiation among the isolated populations [67]. In this study, the population structure analysis showed an obvious genetic divergence between CE and CW, but not the XN population (Figs. 4A-C), which is consistent with the pattern of their natural geographical distribution [40, 42, 68]. This result is consistent with our previous research showing that there are shared haplotypes between the XN population and CW based on cpDNA and adaptive gene sequences [67]. This may be related to the geographical location of the XN population, which is closer to CE between CE and CW and may have experienced recent historical gene flow with CW. Therefore, we guess that ecological processes may play key roles in adaptive differentiation and speciation when the selection of environmental differences leads to the continuous evolution of barriers to gene flow between populations.

Genetic differentiation and environmental adaptive divergence

In previous studies, it was found that the heterogeneity of the species' habitats would cause the species to undergo adaptive evolution under different environmental gradients and could even lead to adaptive genetic differentiation and the formation of new species [6, 28-32, 69-71]. In this study, the RDA results showed that 16 environmental factors divided 17 populations of *Liriodendron* into three groups. Axis 1 separated *L. chinense* from *L. tulipifera*, while axis 2 separated CE from CW (Fig. 2). Among the 16 environmental factors, 8 had a significant influence on the population genetic structure of *Liriodendron* (Table 3). Furthermore, LFMM analysis showed that seven environmental factors were associated with the highest number of EAL (Additional file 8: Table S6). Here, we found that four environmental factors (bio2, bio15, bio17 and srad5) play key roles in the environmental adaptive genetic differentiation of *Liriodendron* according to RDA and LFMM analyses. In addition, SEEVA results showed that these four environmental factors were significant differentiation between *L. chinense* and *L. tulipifera*, and had different levels of divergence between eastern and western groups of *L. chinense* (Additional file 4: Table S2). In conclusion, we speculate that these four environmental factors play the most important roles in the adaptive genetic differentiation of *Liriodendron*.

Diurnal temperature plays an important role in the growth and development of plants. Plants photosynthesize during the day to produce sugars and other organic matter. At night, there is no light, and plants consume organic matter through respiration. If the temperature is lower at night, the plant's

respiration is weakened, and organic matter accumulates, providing more nutrients for the growth and development of plants. Therefore, a large diurnal temperature will be more conducive to plant growth and development than a small diurnal temperature. In our study, we found a wide difference in the mean diurnal range of temperature (bio2) between southeastern America and subtropical China. Compared with subtropical China where *L. chinense* is located, the southeastern United States where *L. tulipifera* is located has a larger average diurnal temperature range (Additional file 4: Table S2). This also confirms why *L. tulipifera* grows faster than *L. chinens*e [47, 48].

Subtropical China has a subtropical monsoon climate with four distinct seasons: hydrothermal same season, dry in winter and rainy in summer. The southeastern United States has a subtropical monsoon humid climate and is warm and humid in winter and rainy in summer. Different climate types in the two regions cause discrepancies in precipitation seasonality (bio15) (Additional file 4: Table S2), which will inevitably have various effects on species evolution in the two regions. We also found a wide difference in precipitation of the driest quarter (bio17) between the southeastern United States and subtropical China. Compared with subtropical China, the southeastern United States has higher precipitation in the driest quarter (Additional file 4: Table S2). This suggests that *L. chinense* in subtropical China may have better drought resistance. Previous studies also found that *L. chinense* is sensitive to water conditions and is not resistant to flooding [49, 50].

Light plays vital roles in the growth, development and morphogenesis of plants. In this study, there was an obvious difference in solar radiation in May (srad5) between the southeastern United States and subtropical China. Compared with subtropical China, solar radiation in the southeastern United States is stronger in May (Additional file 4: Table S2). Interestingly, the flowering period of *Liriodendron* is also in May. This implies that different levels of solar radiation may have various effects on the flowering period of *L. chinense* and *L. tulipifera*. An investigation found that the flowering period of *L. tulipifera* begins in late spring to summer and lasts for 2-6 weeks [46], which may be related to the stronger solar radiation in the southeastern United States. In addition, the results of the ecological niche differentiation analysis indicated significant ecological niche divergence between *L. chinense* and *L. tulipifera* habitats. Conclusively, our research suggests that the divergent selection of environmental differences plays a key role in the interspecific genetic differentiation of *Liriodendron*.

Furthermore, the results of genetic diversity analysis showed that the genetic diversity of *L. tulipifera* populations is higher than that of *L. chinense* populations. This may be because *L. tulipifera* is widely and continuously distributed in the southeastern United States, with frequent gene flow between populations, while *L. chinense* is distributed in small and scattered populations in subtropical China, with isolation among populations, which suggests the reason why *L. tulipifera* has stronger adaptation. Greater genetic variations provide the genetic basis for the widespread adaptation of *L. tulipifera* to various environments, which is consistent with previous research [51].

Signatures of natural selection

Both the results of the outlier analysis and environmental association analysis identified 5.17% (858/16592) candidate loci that deviate from selective neutrality and relate to environmental adaptation. Among the 858 loci, 667 loci are associated with 464 genes, and 191 loci correspond to the intergenomic regions. We found that some genes are involved in the response to biotic and abiotic stresses, phenology, disease resistance and hormonal response in other plants (Additional file 10: Table S7). For example, *UVR8* is involved in photomorphogenesis and stress acclimation in *Arabidopsis* [56], and *RP/EB family member 1B-like* is involved in the response to biotic wounding in *Nicotiana attenuata* [54]. *FRIGIDA-like* is involved in flowering time in *Arabidopsis* [55], and *RGA2-like* encodes a disease resistance protein in *Hevea brasiliensis* [57]. In addition, *ARF1* plays an important role in ethylene and cytokinin responses in *Arabidopsis thaliana* [53]. *MPK6* is involved in ethylene biosynthesis, oxidative stress and osmotic stress in plants. Furthermore, we found that the loci of these genes were associated with specific environmental factors (Additional file 10: Table S7). For example, *FRIGIDA-like* (Scaffold580_7851536) is associated with solar radiation, water vapor pressure and wind speed. *UVR8* (Scaffold3419_3526874 and Scaffold3419_3527115) is associated with elevation, solar radiation and diurnal temperature. *MPK6* (Scaffold291_7154415) is associated with elevation, precipitation seasonality and water vapor pressure. The identification of these adaptive genes suggests that it is feasible to exploit adaptive genes using a landscape genetics approach.

Furthermore, pathway enrichment analysis revealed that genes involved in plant-pathogen interactions, phosphatidylinositol signaling systems, ubiquitin-mediated proteolysis, carotenoid biosynthesis, alpha-linolenic acid metabolism and phagosomes were significantly enriched. (Fig. 5). The plant-pathogen interaction pathway plays vital roles in the immunity of *Arabidopsis* [72]. In addition, the phosphatidylinositol signaling system pathway relates to a multitude of cellular processes with key importance for plant function, development and regulation of energy, carbon metabolism and stress response [73-75]. The ubiquitin-mediated proteolysis pathway is involved in the immune response in *Arabidopsis* [76], and the carotenoid biosynthesis pathway plays important roles in fruit development, petal coloration and photosynthesis [77-79]. The alpha-linolenic acid metabolism pathway relates to salt tolerance and cold stress [80, 81], and the phagosome pathway is involved in environmental stress [82]. These results suggested that most of the variation loci obtained from environmental correlation analysis are indeed closely related to the adaptation of species. It was further confirmed that the combination of population transcriptome resources and environmental datasets is feasible to study the adaptive genetic differentiation of species by landscape genetic approach. These results also suggested that environmental factors related to habitats of species play keys roles in driving adaptive genetic differentiation of species genome. Furthermore, resistance-related genomic loci will provide useful molecular marker resources for forest tree molecular marker-assisted breeding.

Conclusions

In this study, environmental data and SNPs from population transcriptomes were integrated to investigate the spatial distribution of genetic variations to detect relationships between ecological factors and genetic differentiation. The results showed that the population genetic structure of *Liriodendron* was

closely related to the divergence of multiple environmental factors. The high degree of interspecific genetic differentiation and ecological niche divergence suggests that *L. chinense* and *L. tulipifera* may be have experienced selective pressures from different environments in East Asia and North America. Our results suggested that combining population transcriptome and environmental factor datasets is potential feasible to explore ecological adaptation differentiation of species by landscape genetic approach. The combination of population transcriptome resources and environmental datasets also provides new insights into the study of adaptive genetic differentiation of species. This study is an important step towards understanding the genetic mechanism of ecological adaptation differentiation of species. The resistance-related genomic loci will also provide useful molecular marker resources for forest tree molecular marker-assisted breeding. For researchers, using RNA-seq strategy for *Liriodendron* populations is a good potential approach to obtain their genomic resources. The high-quality SNP library and population transcriptome data will provide the basis for future population genomic analyses of *Liriodendron* and understanding the genetic mechanism of plant ecological adaptive differentiation of plants. Furthermore, our research also provided an empirical reference for studying the adaptive differentiation of allopatric species.

Methods

Sampling and RNA extraction

In order to obtain more orthologous genes, we adopted the strategy of mixing shoot apex samples. Shoot apex materials of 80 individuals from 17 populations (12 populations of *L. chinense* and 5 populations of *L. tulipifera*) were collected from a provenance trial plantation of Nanjing Forestry University located in Zhenjiang, Jiangsu Province (119°13'20"E, 32°7'8"N) [48]. Fresh shoot apexes were quick frozen, transported to the laboratory and stored at -80 ℃ pending RNA extraction. All shoot apex materials were collected on the same day in May 2019 (Fig. 6). The origins of the 17 populations are shown in Fig. 7. There is a voucher specimen for each population, and the voucher specimens and depositary of 17 populations are shown in Table 2. All samples were collected following current Chinese regulations and international guidelines.

Fig. 6. Shoot apex materials of *Liriodendron*.

Fig. 7. The original population distributions of 17 provenances of *Liriodendron*. (The map was created by software DIVA-GIS 7.5, the software and free spatial data were downloaded from http://www.diva-gis.org)

RNA was extracted using an RNAprep Pure Kit (Tiangen, China). RNA purity was checked using a NanoPhotometer® spectrophotometer (Implen, CA, USA). A total amount of 1 μg RNA per sample was used as input material for the RNA sample preparations. Sequencing libraries were generated using the NEBNext® UltraTM RNA Library Prep Kit for Illumina® (NEB, USA) following the manufacturer's recommendations, and index codes were added to attribute sequences to each sample.

Data generation

The cDNA libraries were sequenced on an Illumina HiSeq platform with a 150 bp paired-end read length at Novogene-Tianjin. Raw reads were filtered by removing read adaptors and reads featuring unknown nucleotides and those with low quality (where PHRED values were < 20 for more than 50% of the bases). After filtering, ~8 Gb of clean data were generated for each individual (Additional file 1: Table S1). Clean data were mapped to the *L. chinense* reference genome [52] using Hisat2 v2.0.4 [83]. First, Picard-tools v1.96 (https://broadinstitute.github.io/picard/) and samtools v0.1.18 (https://github.com/samtools/) were used to sort, mark duplicated reads, remove duplicated reads and reorder the bam alignment results of each sample. Second, we used GATK2 v3.2 [84] to perform SNP calling. Third, we used VCFtools v4.1 (https://vcftools.github.io/) to filter the SNPs. We applied the following filters: quality ≥ 20, coverage depth (DP) ≥ 2 for each locus of each individual, and minor allele frequency (MAF) ≥ 0.05. In addition, we retained only diallelic sites. Finally, only those SNP loci that were covered by reads in all individuals were retained for downstream analysis.

Environmental variables

We downloaded 55 climatic variables (11 temperature variables, 8 precipitation variables, 12 solar radiation variables, 12 water vapor pressure variables and 12 wind speed variables) from Worldclim v1.4 (https://www.worldclim.org) from 1970 to 2000 at 30 arc second resolution (~1 km^2) and extracted environmental variable values using ArcGIS 10.2 (ESRI, Redlands, CA, USA) for the 17 population locations of *Liriodendron* (Additional file 12: Tables S9 and Additional file 13: Table S10). In addition, occurrence records of *L. chinense* and *L. tulipifera* were collected from the Chinese Virtual Herbarium (CVH), the Global Biodiversity Information Facility (GBIF) and field investigation to identify their niches. Repetitive geographic coordinate information was deleted, and 145 and 478 geographic coordinate locations of *L. chinense* and *L. tulipifera* were obtained, respectively. A total of 56 environmental variables (55 climatic variables and 1 topographical variable) for the 623 point locations were extracted and downloaded from Worldclim v1.4 (Additional file 13: Table S10 and Additional file 14: Table S11).

Population genetic structure analysis

First, Admixture v1.3.0 (http://www.genetics.ucla.edu/software) [85] was used to determine the population genetic structure of *Liriodendron*. Admixture is a program for estimating ancestry in a maximum likelihood model that allows for faster and more efficient when thousands of SNPs are employed. We ran 10 replicates for each genetic cluster K as 2 to 10. We chose the K value that best fit the data by detecting the minimum cross-validation error of each K. The SNP sites used for population genetic structure analysis were all neutral sites (13990 SNPs); outlier loci identified by Arlequin and BayeScan were removed from all SNPs (16592 SNPs). We performed principal component analysis (PCA) using Eigensoft v7.2.1 [86]. Eigenvectors were generated from the covariance matrix with the R program. The significance level of the eigenvectors was determined using the Tracey-Window test, which was implemented in the program twstats of Eigensoft v 7.2.1 [86]. In addition, the neighbor-joining tree was constructed based on the distance matrix by phylip v3.695 [87], with 1000 bootstraps.

To provide insight into the distribution of genetic variation and genetic diversity at different levels in the 17 populations of *Liriodendron*, molecular variance (AMOVA) and genetic diversity analysis were performed in Arlequin v3.5 [88]. *Fst* (*F*-statistics) was calculated using Arlequin for the 17 populations of *Liriodendron* to detect population differentiation.

Redundancy analysis (RDA)

To infer the influence of the environmental variables on population genetic structure, we performed redundancy analysis, a constrained linear ordination analysis. The analysis was implemented in Canoco 5.0 (http://www.canoco5.com/) [89]. Here, allele frequencies per locus of each population were used as response variables (Additional file 15: Table S12), and the 56 environmental variables were used as explanatory variables. Before performing redundancy analysis, an interactive forward-selection analysis was implemented using Canoco 5.0 to select the top 16 environmental variables that had the greatest impact on the population genetic structure. Subsequently, RDA was implemented for 16 environmental variables to detect those that had significant effects on the population genetic structure. To reduce the false positive rate, the *P*-value was corrected. Finally, environmental factors with corrected *P*-values less than 0.01 were considered to have a significant impact on the population genetic structure.

Spatial evolutionary and ecological vicariance analysis (SEEVA)

To detect ecological divergence among groups of *Liriodendron*, SEEVA [90] was performed, which can analyze environmental data using statistical methods to detect ecological vicariance in genetic differentiation and speciation. The divergence index (D: 0-1) and impact index (I: 0-1) were calculated for each of the 56 environmental factors in interspecies of *Liriodendron* and east-west groups of *L. chinense*. Meanwhile, Fisher's exact test [91] was used to obtain an appropriate *P*-value for tests with small sample sizes. D > 0.75 indicates high-level divergence. A larger impact index implies that an ecological factor has a greater influence on the differentiation between sister groups and sister species. A *P*-value less than 0.0016 indicated significant divergence for splits between sister groups or sister species. SEEVA v1.01 can be obtained from http://seeva.heiberg.se.

Acquisition and analysis of leaf phenotypic data

To understand whether the phenotypes of the two species diverged, the characteristics of leaf phenotypes were also investigated and analyzed. We collected at least 10 mature leaves from different parts of each of the 80 individuals; a total of 947 leaves were collected, 336 leaves from *L. tulipifera* and 611 leaves from *L. chinense*. The relative positions of the lateral sinuses on the right side of the leaf (Fig. 8) and the number of leaf cracks were measured and recorded, and their positions were plotted using the ggplot2 package in R (https://www.r-project.org/).

Fig. 8. Common leaf shapes of *L. chinense* and *L. tulipifera* (x1: The vertical distance from the right lateral sinus to the primary vein. x2: The vertical distance from the right lateral lobe to the primary vein.

y1: The vertical distance from the right lateral sinus to the leaf base. y2: The vertical distance from the right apical lobe to the leaf base).

Environment association analysis of individual loci

To reduce the false positive rate, three methods were used to detect sites related to adaptive genetic differentiation. First, Arlequin v3.5 [88] and BayeScan v2.01 [92] were used to identify the outlier loci for subsequent environment association analysis. The hierarchical island model in Arlequin was used to detect outlier loci. The advantage of this method is that it is very sensitive to individuals with common histories and substructures. The parameters adopted by Arlequin are as follows: 20,000 simulations, 100 demes to simulate (per group), and the number of simulated groups suggested by the results of Admixture. The loci with *FstP*-value ≤ 0.05 were identified as outlier loci. The second method was performed in BayeScan using the Bayesian approach. The parameters used were as follows: 50,000 burn-in iterations, thinning interval of 10, sample size of 5000, 20 pilot runs, each pilot run had a length of 5000, and 10,000 prior odds for the neutral model. As a result, a Bayes factor (PO) of 100 corresponded to a posterior probability (*P*) of 0.99. Therefore, loci with log (PO) ≥ 2 were considered outlier loci. As a way of detecting SNPs potentially under natural selection for adaptation, we tested for associations between outlier loci and each of 55 environmental factors using a latent factor mixed model in LFMM 1.2 [93]. The advantage of this approach is that it can effectively avoid false positives and false negatives caused by demographic history and population structure. The parameters used were as follows: 10,000 sweeps, 1000 burn-in sweeps, and the number of latent factors suggested by the results of population structure analysis. SNP loci with |Z|-score ≥ 4 and *P*-value ≤ 1.0 E-5 were considered environment-associated loci (EAL).

GO and KEGG enrichment analysis

To detect whether the genes where these EAL are located are related to environmental adaptation, functional annotation and enrichment analysis were performed on these genes. First, for the potential genes related to environmental adaptation, we searched the NT (NCBI nucleotide sequences) and NR (NCBI non-redundant protein sequences) databases based on homologous gene theory and performed functional annotations. Then, GO (Gene Ontology) enrichment analysis of these genes was implemented by the GOseq R package (www.bioconductor.org/packages/) [94], and GO terms with *P*-values less than 0.05 were considered significantly enriched. KEGG (Kyoto Encyclopedia of Genes and Genomes) enrichment analysis was also performed using KOBAS software (http://kobas.cbi.pku.edu.cn/kobas3/download/) [95] to test the statistical enrichment of these genes in KEGG pathways.

Abbreviations

Single nucleotide polymorphism (SNP), quantitative trait loci (QTL), redundancy analysis (RDA), spatial evolutionary and ecological vicariance analysis (SEEVA), principal component analysis (PCA), principal component (PC), latent factor mixed model (LFMM), environment-associated loci (EAL), analysis of

molecular variance (AMOVA), chloroplast DNA (cpDNA), *L. tulipifera* populations (NA), eastern populations of *L. chinense* (CE), western populations of *L. chinense* (CW)

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article and its additional files. Material samples are available from authors.

Competing interests:

The authors declare that they have no competing interests.

Funding

The research was supported by funds from National Natural Science Foundation of China (No. 31770718 and No. 31470660) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Authors' Contributions:

HGL conceived the study. YFS completed the data analysis with the help of HX and ZHT. YFS completed the Leaf morphology analysis with the assistance of YXZ and LCY. YFS wrote the manuscript. All authors read and approved the manuscript.

Acknowledgements

The authors are grateful to Shenghua Zhu, Mujun Liu, Ziyuan Hao, Weiping Zhong, Huanhuan Liu, Chengge Zhang and Shaoying Wen for collecting plant materials.

References

1. Rundle HD, Nosil P. Ecological speciation. Ecol Lett. 2005;8:336–52.
2. Nosil P, Harmon LJ, Seehausen O. Ecological explanations for (incomplete) speciation. Trends Ecol Evol. 2009;24:145–56.

3. Rundell RJ, Price TD. Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. Trends Ecol Evol. 2009;24:394–9.
4. Schluter D. Evidence for ecological speciation and its alternative. Science. 2009;323:737–41.
5. Di Pierroa EA, Mosca E, González-Martínez SC, Binelli G, Neale DB, La Porta N. Adaptive variation in natural alpine populations of Norway spruce (*Piceaabies* [L.] karst) at regional scale: landscape features and altitudinal gradient effects. Forest Ecol Manag. 2017;405:350–9.
6. Li JX, Zhu XH, Li Y, Liu Y, Qian ZH, Zhang XX, Sun Y, Ji LY. Adaptive genetic differentiation in *Pterocarya stenoptera* (Juglandaceae) driven by multiple environmental variables were revealed by landscape genomics. BMC Plant Biol. 2018;18:306.
7. Zhang X, Sun YX, Landis JB, Zhang JW, Yang LS, Lin N, Zhang HJ, Guo R, Li LJ, Zhang YH, Deng T, Sun H, Wang HC. Genomic insight into adaptation to heterogeneous environments for the ancient relictual *Circaeaster agrestis* (Circaeasteraceae, Ranunculales). New Phytol. 2020. doi:10.1111/nph.16669.
8. Hampe A, Jump AS. Climate Relicts: Past, Present, Future. Annu Rev Ecol Evol S. 2011;42:313–33.
9. Hoffmann AA, Sgro CM. Climate change and evolutionary adaptation. Nature. 2011;470:479–85.
10. Anderson JT, Inouye DW, McKinney AM, Colautti RI, Mitchell-Olds T. Phenotypic plasticity and adaptive evolution contribute to advancing flowering phenology in response to climate change. P Roy Soc B-Biol Sci. 2012;279:3843–52.
11. Davis MB, Shaw RG. Range shifts and adaptive responses to Quaternary climate change. Science. 2001;292:673–9.
12. Jump AS, Penuelas J. Running to stand still: adaptation and the response of plants to rapid climate change. Ecol Lett. 2005;8:1010–20.
13. Savolainen O, Lascoux M, Merila J. Ecological genomics of local adaptation. Nat Rev Genet. 2013;14:807–20.
14. Csillery K, Lalague H, Vendramin GG, Gonzalez-Martinez SC, Fady B, Oddou-Muratorio S. Detecting short spatial scale local adaptation and epistatic selection in climate-related candidate genes in European beech (*Fagus sylvatica*) populations. Mol Ecol. 2014;23:4696–708.
15. Delph LF. The Study of Local Adaptation: A Thriving Field of Research. J Hered. 2018;109:1–2.
16. Ye JW, Zhang Y, Wang XJ. Phylogeographic breaks and the mechanisms of their formation in the Sino-Japanese floristic region. Chin J Plant Ecol. 2017;41:1003–19.
17. Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R. A practical guide to environmental association analysis in landscape genomics. Mol Ecol. 2015;24:4348–70.
18. Li Y, Zhang XX, Mao RL, Yang J, Miao CY, Li Z, Qiu YX. Ten Years of Landscape Genomics: Challenges and Opportunities. Front Plant Sci. 2017;8:2136.
19. Geraldes A, Farzaneh N, Grassa CJ, McKown AD, Guy RD, Mansfield SD, Douglas CJ, Cronk QC. Landscape genomics of *Populus trichocarpa*: the role of hybridization, limited gene flow, and natural selection in shaping patterns of population structure. Evolution. 2014;68:3260–80.

20. Stapley J, Reger J, Feulner PGD, Smadja C, Galindo J, Ekblom R, Bennison C, Ball AD, Beckerman AP, Slate J. Adaptation genomics: the next generation. Trends Ecol Evol. 2010;25:705–12.
21. Messer PW, Petrov DA. Population genomics of rapid adaptation by soft selective sweeps. Trends Ecol Evol. 2013;28:659–69.
22. Nosil P, Funk DJ, Ortiz-Barrientos D. Divergent selection and heterogeneous genomic divergence. Mol Ecol. 2009;18:375–402.
23. Alvarez M, Schrey AW, Richards CL. Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution? Mol Ecol. 2015;24:710–25.
24. Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM. A map of local adaptation in *Arabidopsis thaliana*. Science. 2011;334:86–9.
25. Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F, Bergelson J. Adaptation to Climate Across the *Arabidopsis thaliana* Genome. Science. 2011;334:83–6.
26. Lasky JR, Des Marais DL, Lowry DB, Povolotskaya I, McKay JK, Richards JH, Keitt TH, Juenger TE. Natural variation in abiotic stress responsive gene expression and local adaptation to climate in *Arabidopsis thaliana*. Mol Biol Evol. 2014;31:2283–96.
27. Menon M, Barnes WJ, Olson MS. Population genetics of freeze tolerance among natural populations of *Populus balsamifera* across the growing season. New Phytol. 2015;207:710–22.
28. Derome N, Duchesne P, Bernatchez L. Parallelism in gene transcription among sympatric lake whitefish (*Coregonus clupeaformis* Mitchill) ecotypes. Mol Ecol. 2006;15:1239–49.
29. Whitehead A, Dubansky B, Bodinier C, Garcia TI, Miles S, Pilley C, Raghunathan V, Roach JL, Walker N, Walter RB, Rice CD, Galvez F. Genomic and physiological footprint of the Deepwater Horizon oil spill on resident marsh fishes. Proc Natl Acad Sci USA. 2012;109:20298–302.
30. Guggisberg A, Lai Z, Huang J, Rieseberg LH. Transcriptome divergence between introduced and native populations of Canada thistle, *Cirsium arvense*. New Phytol. 2013;199:595–608.
31. Harris SE, O'Neill RJ, Munshi-South J. Transcriptome resources for the white-footed mouse (*Peromyscus leucopus*): new genomic tools for investigating ecologically divergent urban and rural populations. Mol Ecol Resour. 2015;15:382–94.
32. Pan SK, Zhang TZ, Rong ZQ, Hu L, Gu ZR, Wu Q, Dong SS, Liu Q, Lin ZZ, Deutschova L, Li XH, Dixon A, Bruford MW, Zhan XJ. Population transcriptomes reveal synergistic responses of DNA polymorphism and RNA expression to extreme environments on the Qinghai-Tibetan Plateau in a predatory bird. Mol Ecol. 2017;26:2993–3010.
33. Melillo JM, McGuire AD, Kicklighter DW, Moore B, Vorosmarty CJ, Schloss AL. Global climate change and terrestrial net primary production. Nature. 1993;363:234–40.
34. Trumbore S, Brando P, Hartmann H. Forest health and global change. Science. 2015;349:814–8.
35. Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S. Adaptation, migration or extirpation: climate change outcomes for tree populations. Evol Appl. 2008;1:95–111.

36. Liu J, Moller M, Provan J, Gao LM, Poudel RC, Li DZ. Geological and ecological factors drive cryptic speciation of yews in a biodiversity hotspot. New Phytol. 2013;199:1093–108.
37. Gauthier S, Bernier P, Kuuluvainen T, Shvidenko AZ, Schepaschenko DG. Boreal forest health and global change. Science. 2015;349:819–22.
38. Parks CR, Miller NG, Wendel JF, Mcdougal KM. Genetic divergence within the genus *Liriodendron* (Magnoliaceae). Ann Mo Bot Gard. 1983;70:658–66.
39. Parks CR, Wendel JF. Molecular divergence between Asian and North American species of *Liriodendron* (Magnoliaceae) with implications for interpretation of fossil floras. Am J Bot. 1990;77:1243–56.
40. Hao RM, He SA, Tang SJ, Wu SP. Geographical distribution of *Liriodendron chinense* in China and its significance. J Plant Resour Environ. 1995;4:1–6.
41. Tang CQ, Yang Y, Ohsawa M, Momohara A, Mu J, Robertson K. Survival of a tertiary relict species, *Liriodendron chinense* (Magnoliaceae), in southern China, with special reference to village fengshui forests. Am J Bot. 2013;100:2112–9.
42. He SA, Hao RM, Tang SJ. A study on the ecological factors of endangering mechanism of *Liriodendron chinense* (Hemsl.) Sarg. J Plant Resour Environ. 1996;5:1–8.
43. Wolfe JA. Paleogene floras from the Gulf of Alaska region. Geological Survey Professional Paper. 1977;8:294–6.
44. McCartan L, Tiffney BH, Wolfe JA, Ager TA, Wing SL, Sirkin LA, Ward LW, Brooks J. Late Tertiary floral assemblage from upland gravel deposits of the southern Maryland Coastal Plain. Geology. 1990;18:311–4.
45. Sewell MM, Parks CR, Chase MW. Intraspecific chloroplast DNA variation and biogeography of north American *Liriodendron* L. (Magnoliaceae). Evolution. 1996;50:1147–54.
46. Katherine EK. Assessment of Genetic Variation of *Acer rubrum* L. and *Liriodendron tulipifera* L. Populations in Unmanaged Forests of the Southeast United States. MS thesis, University of Virginia Tech, Blacksburg, VA, USA. 2008.
47. Lee HR, Lee HW, Lee YW, Kazlauskas RJ, Park TH. Improved pretreatment of yellow poplar biomass using hot compressed water and enzymatically-generated peracetic acid. Biomass Bioenerg. 2017;105:190–6.
48. Li HG, Chen L, Liang CY, Huang MR. A Case Study on Provenance Testing of Tulip Tree (*Liriodendron spp*). J Forest Eng. 2005;19:13–6.
49. Zhang XP, Fang YM, Chen YJ. Effect of waterlogging stress on physiological indexes of *Liriodendron* seedlings. J Plant Resour Environ. 2006;15:41–4.
50. Pan XY, Ji KS, Fang Y. Changes in enzyme activities in different clones of *Liriodendron chinense* × *L. tulipifera* under flooding stress. J Northwest Fore Univ. 2007;22:43–6.
51. Cheng YL, Li HG. Interspecies evolutionary divergence in *Liriodendron*, evidence from the nucleotide variations of *LcDHN-like* gene. BMC Evol Biol. 2018;18:195.

52. Chen JH, Hao ZD, Guang XM, Zhao CX, Wang PK, Xue LJ, Zhu QH, Yang LF, Sheng Y, Zhou YW, Xu HB, Xie HQ, Long XF, Zhang J, Wang ZR, Shi MM, Lu Y, Liu SQ, Guan LH, Zhu QH, Yang LM, Ge S, Cheng TL, Laux T, Gao Q, Peng Y, Liu N, Yang SH, Shi JS. *Liriodendron* genome sheds light on angiosperm phylogeny and species-pair differentiation. Nat Plants. 2019;5:18–25.
53. Ellis CM, Nagpal P, Young JC, Hagen G, Guilfoyle TJ, Reed JW. *AUXIN RESPONSE FACTOR1* and *AUXIN RESPONSE FACTOR2* regulate senescence and floral organ abscission in *Arabidopsis thaliana*. Development. 2005;132:4563–74.
54. Schafer M, Meza-Canales ID, Navarro-Quezada A, Brutting C, Vankova R, Baldwin IT, Meldau S. Cytokinin levels and signaling respond to wounding and the perception of herbivore elicitors in *Nicotiana attenuata*. J Integr Plant Biol. 2015;57:198–212.
55. Urban J, Joanne W, Clare L, Scott M, Richard A, Caroline D. Molecular Analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. Science. 2000;209:344–7.
56. Favory JJ, Stec A, Gruber H, Rizzini L, Oravecz A, Funk M, Albert A, Cloix C, Jenkins GI, Oakeley EJ, Seidlitz HK, Nagy F, Ulm R. Interaction of *COP1* and *UVR8* regulates UV-B-induced photomorphogenesis and stress acclimation in *Arabidopsis*. EMBO J. 2009;28:591–601.
57. Rahman AYA, Usharraj AO, Misra BB, Thottathil GP, Jayasekaran K, Feng Y, Hou S, Ong SY, Ng FL, Lee LS, Tan HS, Sakaff MK, Teh BS, Khoo BF, Badai SS, Aziz NA, Yuryev A, Knudsen B, Dionne-Laporte A, Mchunu NP, Yu QY, Langston BJ, Freitas TA, Young AG, Chen R, Wang L, Najimudin N, Saito JA, Alam M. Draft genome sequence of the rubber tree *Hevea brasiliensis*. BMC Genom. 2013;14:1–15.
58. Yang AH, Dick CW, Yao XH, Huang HW. Impacts of biogeographic history and marginal population genetics on species range limits: a case study of *Liriodendron chinense*. Sci Rep. 2016;6:25632.
59. Zhong YD, Yang AH, Liu SJ, Liu LP, Li YQ, Wu ZX, Yu FX. RAD-Seq data point to a distinct split in *Liriodendron* (Magnoliaceae) and obvious east–west genetic divergence in *L. chinense*. Forests. 2018;10:1–13.
60. Yang AH, Zhong YD, Liu SJ, Liu LP, Liu TY, Li YQ, Yu FX. New insight into the phylogeographic pattern of *Liriodendron chinense* (Magnoliaceae) revealed by chloroplast DNA: east-west lineage split and genetic mixture within western subtropical China. PeerJ. 2019;7:e6355.
61. Song ZC, Li MY, Zhong L. Spore-pollen assemblage from Cretaceous to Early Tertiary in Sanshui Basin, Guangdong. Beijing: Science press; 1986. pp. 88–9.
62. Bell WA. Upper Cretaceous Floras of the Dunvegan, Bad Heart, and Milk River Formations of Western Canada. Geological Survey of Canada, Bulletin. 1963;94:1–76.
63. Frumin SI, Friis EM. Liriodendroid seeds from the Late Cretaceous of Kazakhstan and North Carolina, United States. Rev Palaeobot Palyno. 1996;94:39–55.
64. Hebda RJ, Irving E. On the Origin and Distribution of Magnolias: Tectonics, DNA and Climate Change. Timescales of the Paleomagnetic Field Geophysical Monograph Series. 2004;145:43–57.
65. Zhang Q, Chiang TY, George M, Liu JQ, Abbott RJ. Phylogeography of the Qinghai-Tibetan Plateau endemic *Juniperus przewalskii* (Cupressaceae) inferred from chloroplast DNA sequence variation. Mol Ecol. 2005;14:3513–24.

66. Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS. Comparative phylogeography of unglaciated eastern North America. Mol Ecol. 2006;15:4261–93.
67. Shen YF, Cheng YL, Li KQ, Li HG. Integrating Phylogeographic Analysis and Geospatial Methods to Infer Historical Dispersal Routes and Glacial Refugia of *Liriodendron chinense*. Forests. 2019;10:1–26.
68. Li KQ, Chen L, Feng YH, Yao JX, Li B, Xu M, Li HG. High genetic diversity but limited gene flow among remnant and fragmented natural populations of *Liriodendron chinense* Sarg. Biochem Syst Ecol. 2014;54:230–6.
69. Bellard C, Bertelsmeier C, Leadley P, Thuiller W, Courchamp F. Impacts of climate change on the future of biodiversity. Ecol Lett. 2012;15:365–77.
70. Alberto FJ, Aitken SN, Alia R, Gonzalez-Martinez SC, Hanninen H, Kremer A, Lefevre F, Lenormand T, Yeaman S, Whetten R, Savolainen O. Potential for evolutionary responses to climate change - evidence from tree populations. Global Change Biol. 2013;19:1645–61.
71. De La Torre AR, Roberts DR, Aitken SN. Genome-wide admixture and ecological niche modelling reveal the maintenance of species boundaries despite long history of interspecific gene flow. Mol Ecol. 2014;23:2046–59.
72. Maurya R, Srivastava D, Singh M, Sawant SV. Envisioning the immune interactome in *Arabidopsis*. Funct Plant Biol. 2020;47:1–22.
73. Heilmann I. Plant phosphoinositide signaling - dynamics on demand. BBA-Mol Cell Biol L. 2016;1861:1345–51.
74. Ramanan R, Tran QG, Cho DH, Jung JE, Kim BH, Shin SY, Choi SH, Liu KH, Kim DS, Lee SJ, Crespo JL, Lee HG, Oh HM, Kim HS. The Ancient Phosphatidylinositol 3-Kinase Signaling System Is a Master Regulator of Energy and Carbon Metabolism in Algae. Plant Physiol. 2018;177:1050–65.
75. Liu S, Hu ZM, Zhang Q, Yang X, Critchley AT, Duan D. PI signal transduction and ubiquitination respond to dehydration stress in the red seaweed *Gloiopeltis furcata* under successive tidal cycles. BMC Plant Biol. 2019;19:516.
76. Till CJ, Vicente J, Zhang H, Oszvald M, Deery MJ, Pastor V, Lilley KS, Ray RV, Theodoulou FL, Holdsworth MJ. The *Arabidopsis thaliana* N-recognin E3 ligase PROTEOLYSIS1 influences the immune response. Plant Direct. 2019;3:e00194.
77. Lado J, Zacarias L, Rodrigo MJ. Regulation of carotenoid biosynthesis during fruit development. Subcell Biochem. 2016;79:161–98.
78. Llorente B. Regulation of carotenoid biosynthesis in photosynthetic organs. Subcell Biochem. 2016;79:141–60.
79. Hao ZD, Liu SQ, Hu LF, Shi JS, Chen JH. Transcriptome analysis and metabolic profiling reveal the key role of carotenoids in the petal coloration of *Liriodendron tulipifera*. Hortic Res. 2020;7:70.
80. Li Q, Song J. Analysis of widely targeted metabolites of the euhalophyte *Suaeda salsa* under saline conditions provides new insights into salt tolerance and nutritional value in halophytic species. BMC Plant Biol. 2019;19:1–11.

81. Zhang Y, Zhang YT, Lin YX, Luo Y, Wang XR, Chen Q, Sun B, Wang Y, Li MY, Tang HR. A transcriptomic analysis reveals diverse regulatory networks that respond to cold stress in strawberry (*Fragaria* × *ananassa*). Int J Genomics. 2019;2019:1–13.
82. Yoshimoto K, Ohsumi Y. Unveiling the molecular mechanisms of plant autophagy-from autophagosomes to vacuoles in plants. Plant Cell Physiol. 2018;59:1337–44.
83. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg S. TopHat2 accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14:R36.
84. McKenna A, Hanna M, Banks E, Sivachenko A, Clibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.
85. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19:1655–64.
86. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet. 2006;2:e190.
87. Felsenstein J. Phylogeny Inference Package (Version 3.2). Cladistics. 2005;5:164–6.
88. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Res. 2010;10:564–7.
89. Ter Braak CJF, Šmilauer P. Canoco reference manual and user's guide: software for ordination (version 5.0). Ithaca: Microcomputer Power; 2012.
90. Struwe L, Smouse PE, Heiberg E, Haag S, Lathrop RG. Spatial evolutionary and ecological vicariance analysis (SEEVA), a novel approach to biogeography and speciation research, with an example from Brazilian Gentianaceae. Journal Biogeogr. 2011;38:1841–54.
91. Fisher RA. Statistical methods for research workers. Paleobiota of China. New York: Hafner press; 1958.
92. Foll M, Gaggiotti O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics. 2008;180:977–93.
93. Frichot E, Schoville SD, Bouchard G, Francois O. Testing for associations between loci and environmental gradients using latent factor mixed models. Mol Biol Evol. 2013;30:1687–99.
94. Matthew DY, Matthew JW, Gordon KS, Alicia O. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010;11:R14.
95. Xie C, Mao XZ, Huang JJ, Ding Y, Wu JM, Dong S, Kong L, Gao G, Li CY, Wei LP. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res. 2011;39:W316–22.

Figures

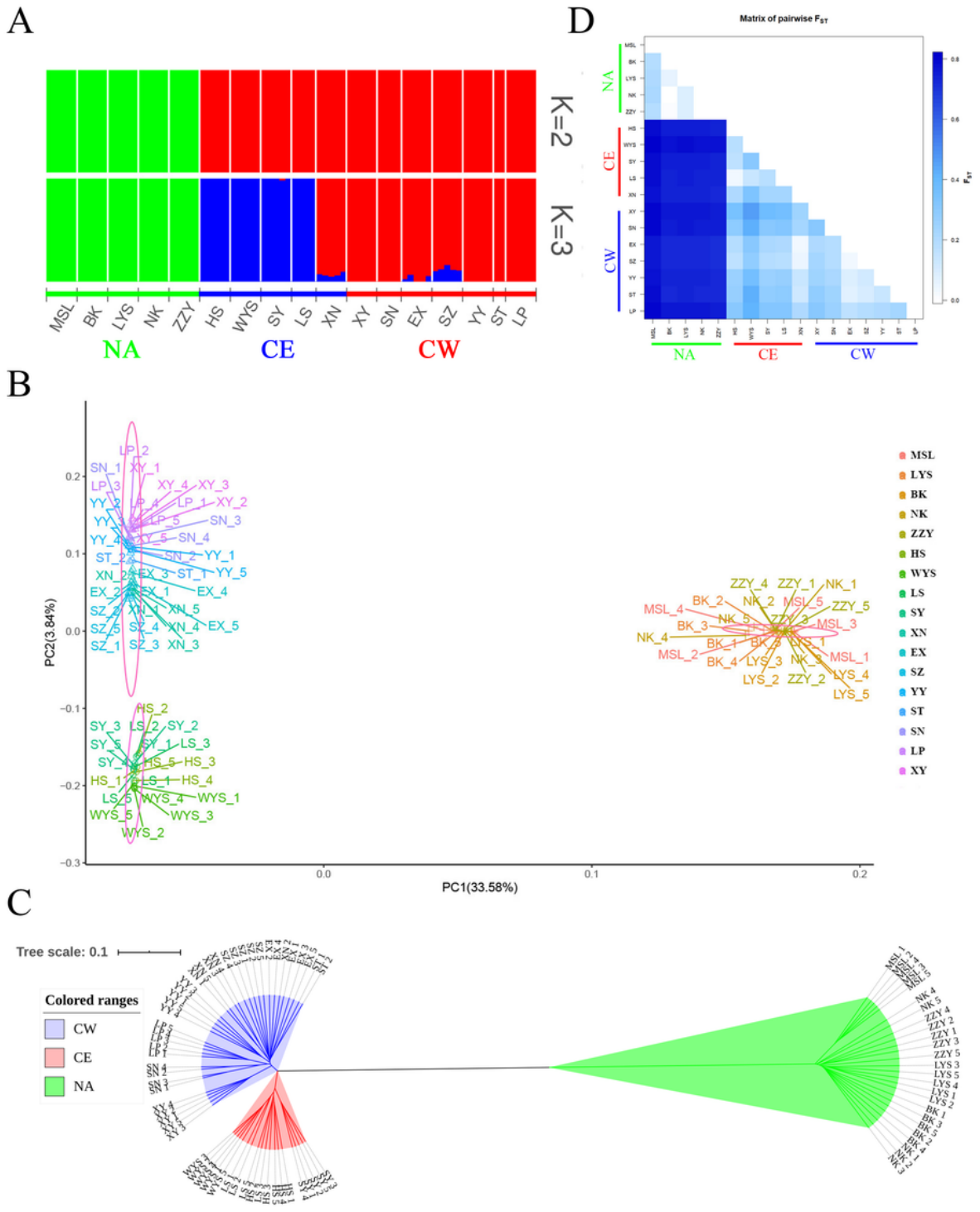

Figure 1

Population genetic structure of Liriodendron. (A) The Admixture plots (K = 2 and 3) based on 13990 neutral loci. (B) The PCA result based on 16592 SNPs identified from Liriodendron. The 17 colors correspond to 17 populations from 3 groups. (C) Neighbor-joining phylogenetic tree of Liriodendron based on 16592 SNPs, with the evolutionary distances measure by p-distances with phylip. The three colors correspond to the 3 groups. (D) Matrix of the pairwise Fst of 17 populations in Liriodendron.

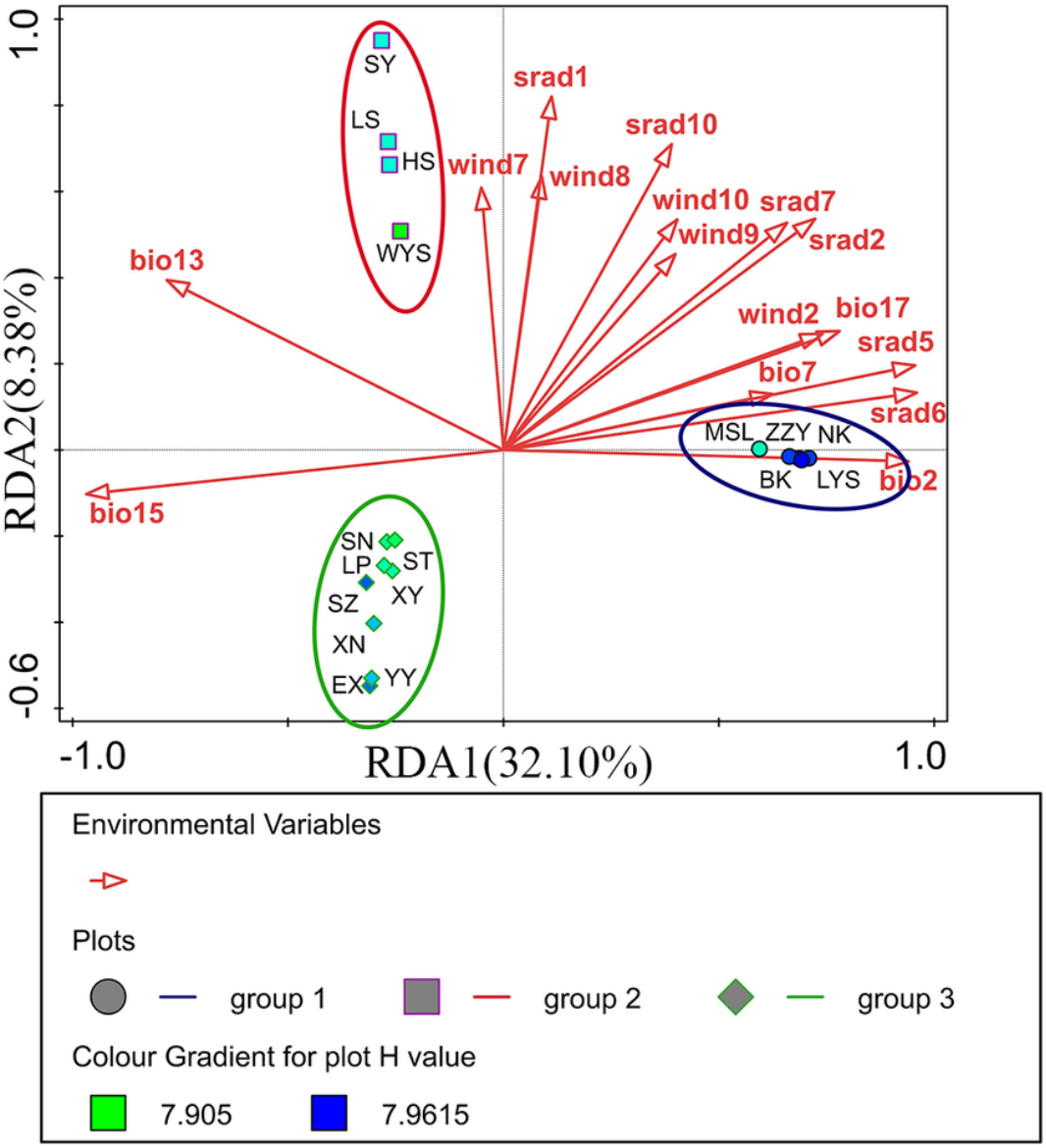

Figure 2

Redundancy analysis of Liriodendron showing the relative contributions of 16 environmental variables to the population genetic structure. The biplot depicts the eigenvalues and the lengths of eigenvectors for the RDA, and the color gradient corresponds to genetic diversity.

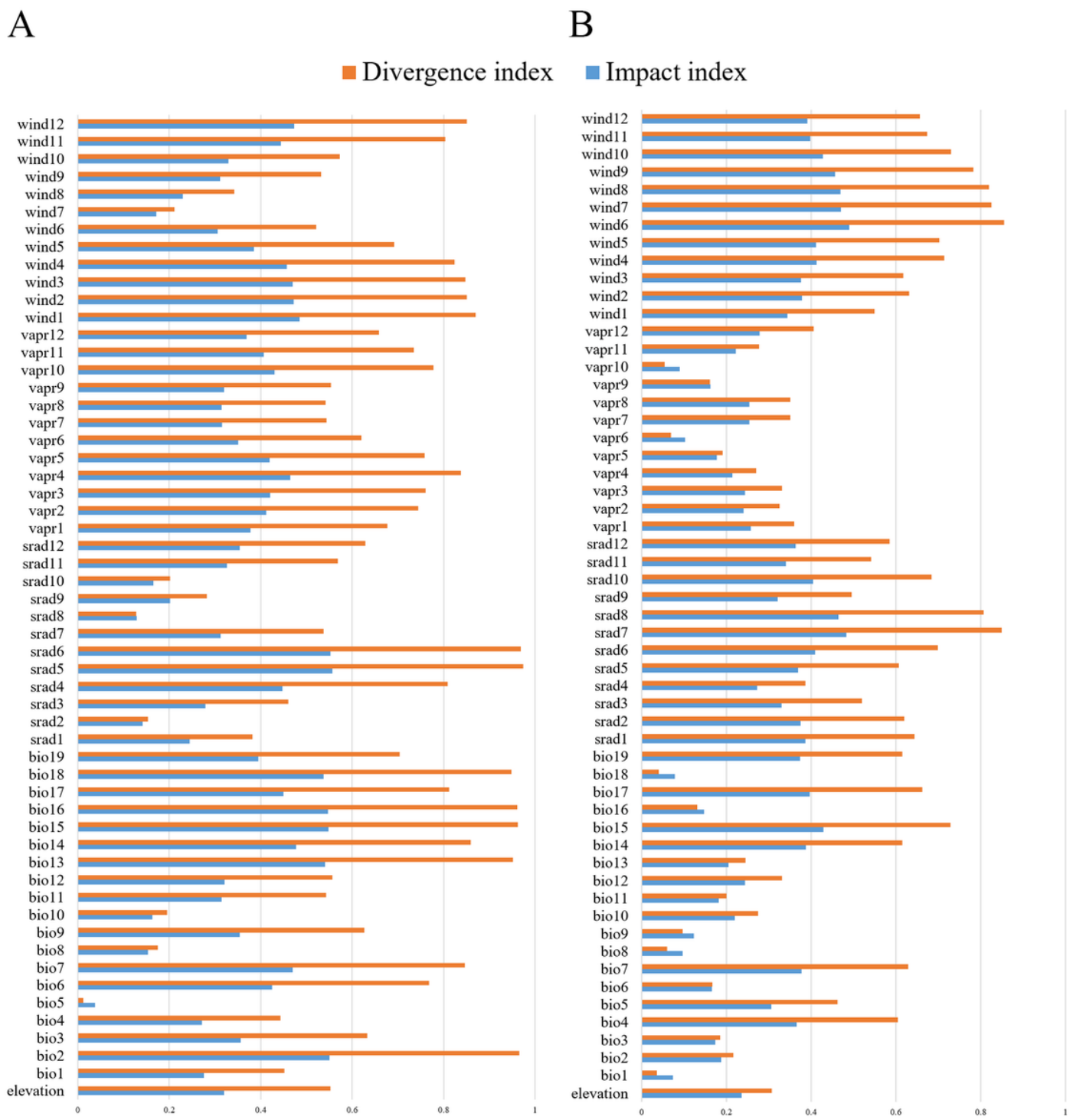

Figure 3

The divergence index and impact index shown as bar diagrams for elevation and 55 climatic variables between sister species and sister groups. (A) The divergence index and impact index shown as bar diagrams for elevation and 55 climatic variables between L. chinense and L. tulipifera. (B) The divergence index and impact index shown as bar diagrams for elevation and 55 climatic variables between east and west groups.

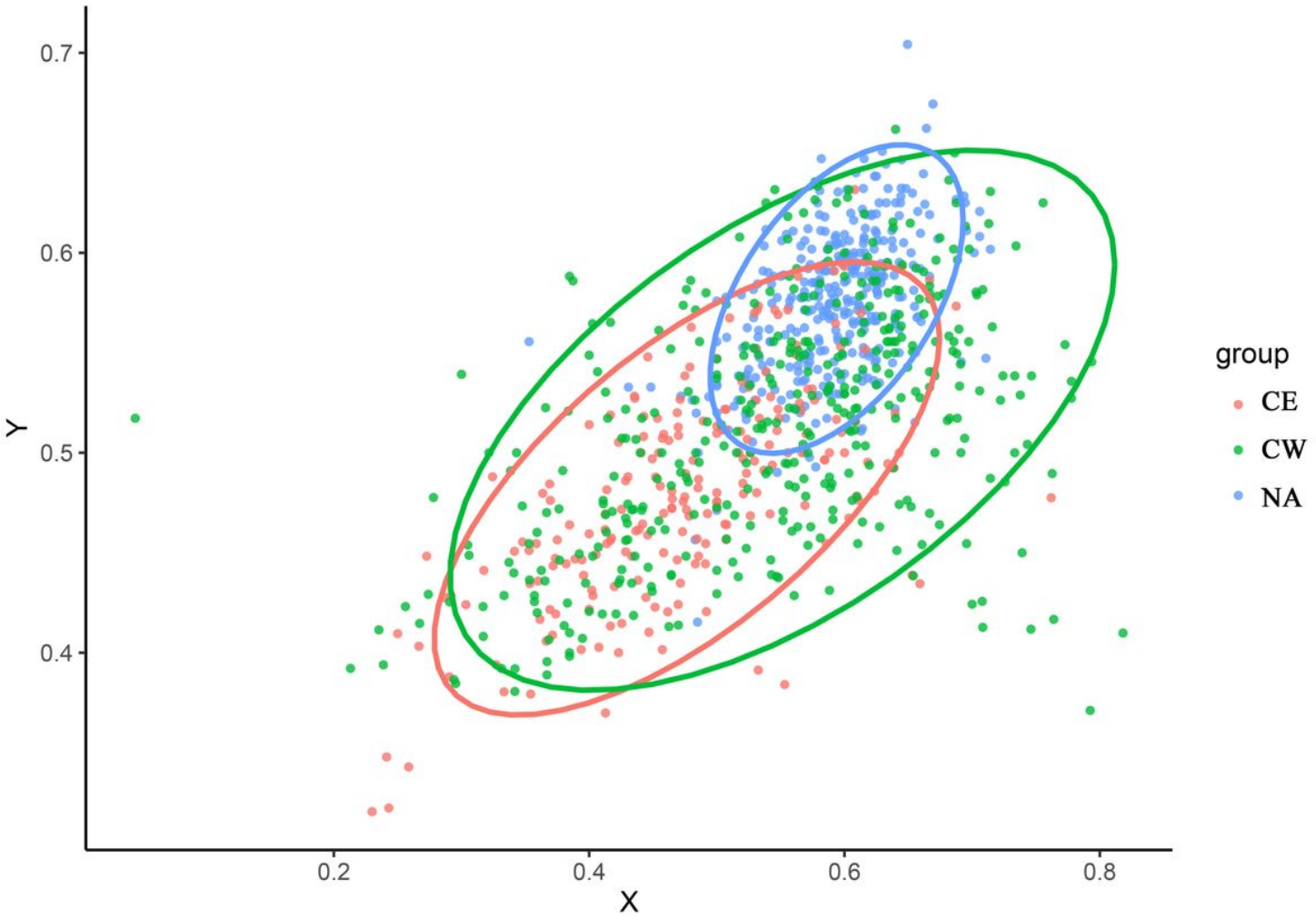

Figure 4

Leaf shape variation distributions in three groups of Liriodendron (X-axis: The ratio of x1 and x2; Y-axis: The ratio of y1 and y2) (Confidence interval: 95%).

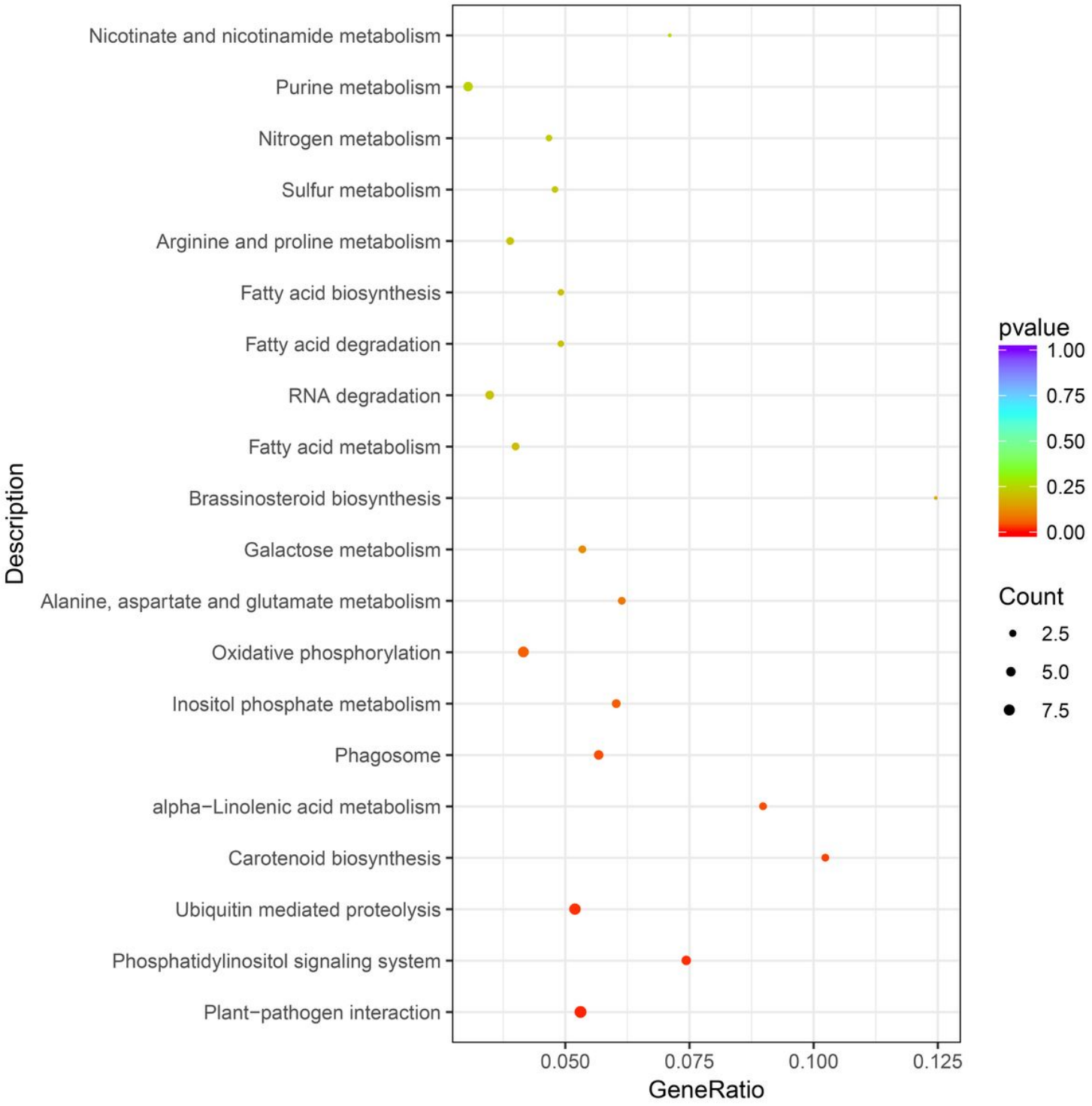

Figure 5

The plots of the KEGG enrichment analysis results.

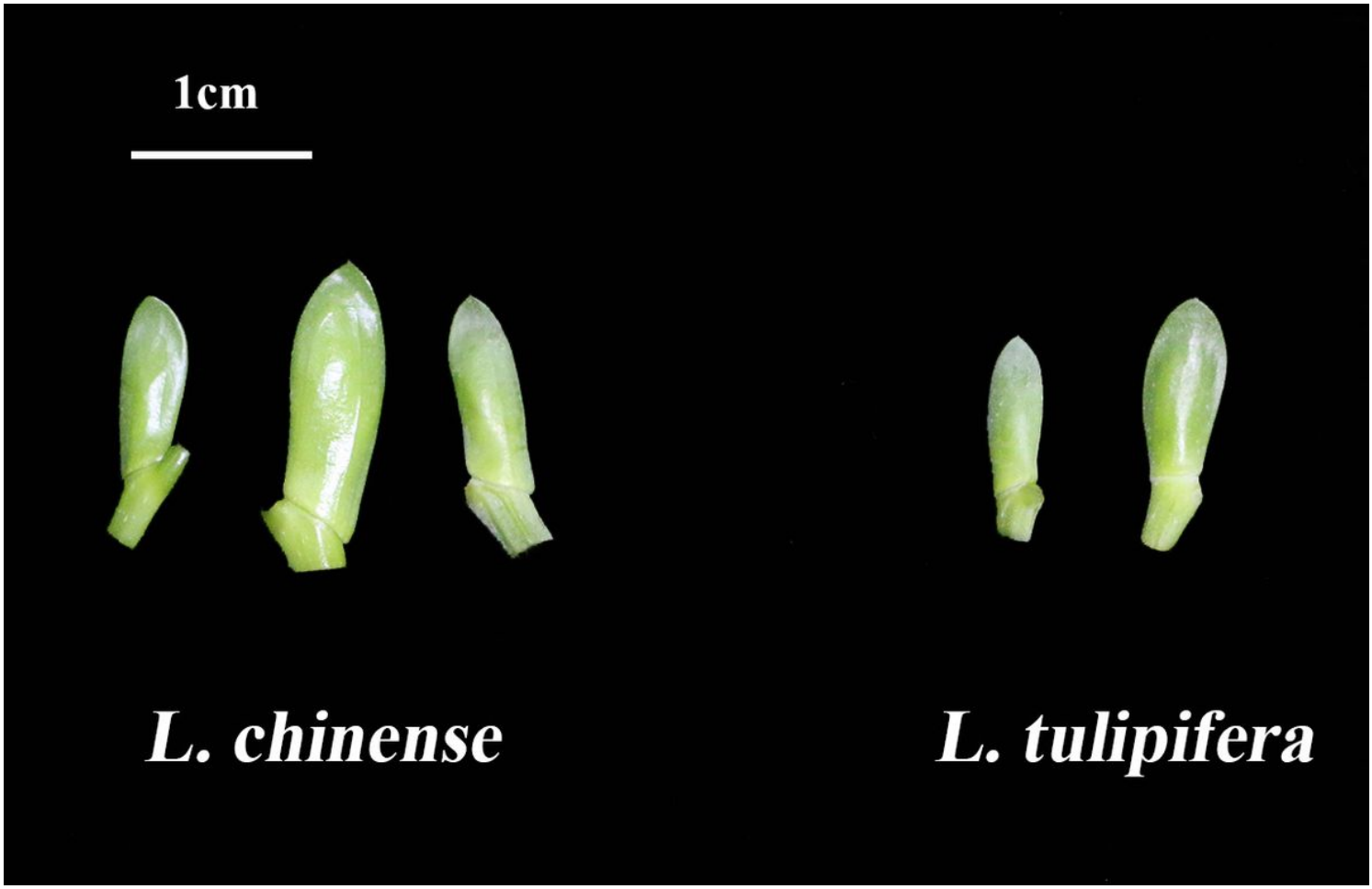

Figure 6

Shoot apex materials of Liriodendron.

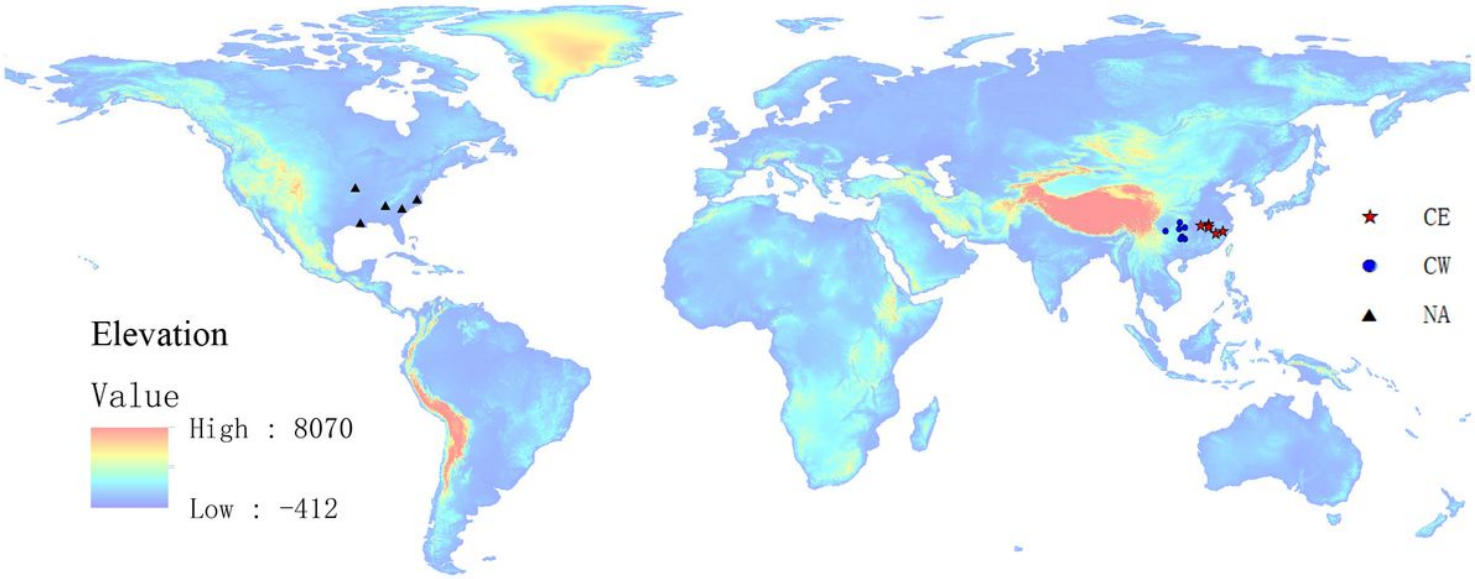

Figure 7

The original population distributions of 17 provenances of Liriodendron. (The map was created by software DIVA-GIS 7.5, the software and free spatial data were downloaded from http://www.diva-gis.org) Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

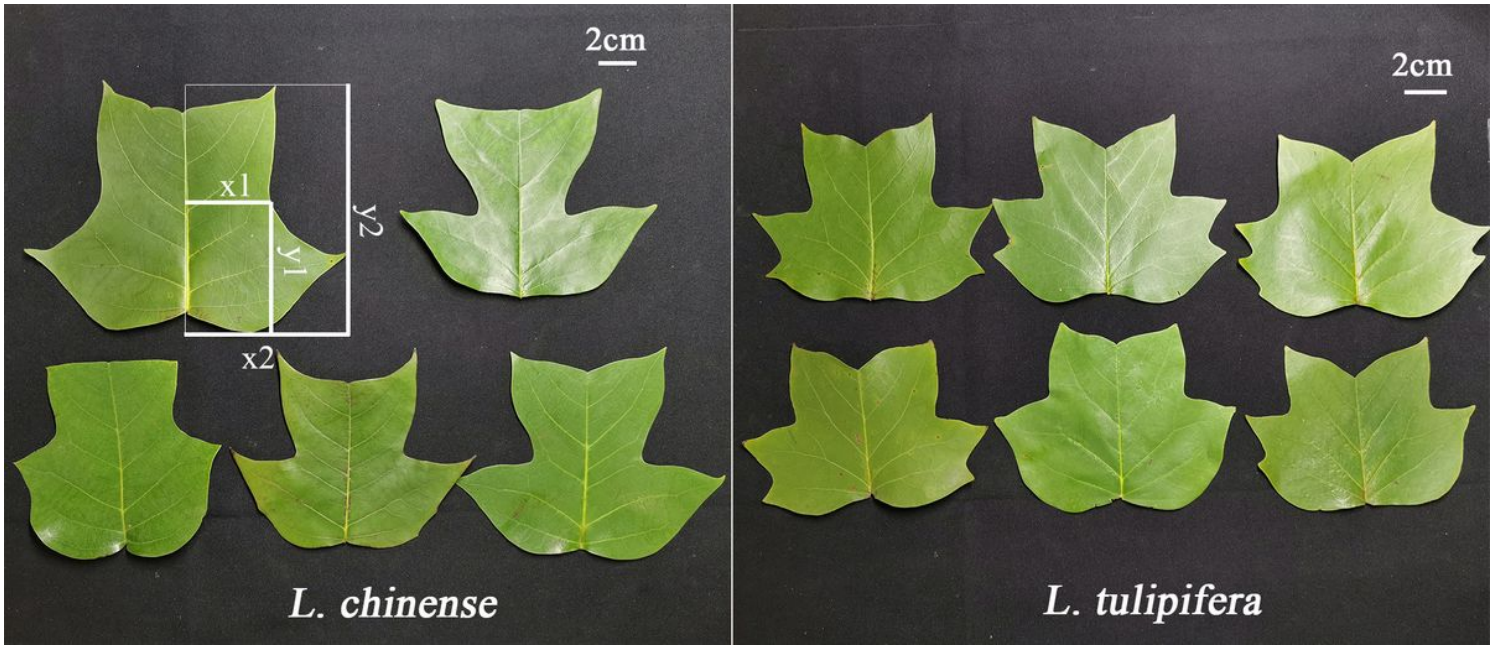

Figure 8

Common leaf shapes of L. chinense and L. tulipifera (x1: The vertical distance from the right lateral sinus to the primary vein. x2: The vertical distance from the right lateral lobe to the primary vein. y1: The vertical distance from the right lateral sinus to the leaf base. y2: The vertical distance from the right apical lobe to the leaf base).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS12.xlsx
- TableS11.xlsx
- TableS10.docx
- TableS9.xlsx
- TableS8.xlsx
- TableS7.xlsx
- TableS6.xlsx
- TableS5.xlsx

- TableS4.xlsx
- TableS3.docx
- TableS2.xlsx
- TableS1.docx
- FigureS3.pdf
- FigureS2.png
- FigureS1.png