

RESEARCH ARTICLE

Open Access



Interspecies evolutionary divergence in *Liriodendron*, evidence from the nucleotide variations of *LcDHN-like* gene

Yanli Cheng and Huogen Li* 

Abstract

Background: *Liriodendron* is a genus of Magnoliaceae, which consists of two relict species, *Liriodendron chinense* and *L. tulipifera*. Although the morphologies are highly similar, the two species exhibit different adaptive capacity. Dehydrins (DHNs) are abiotic stresses resistant proteins in *planta*, which are associated with adaptive evolution. To better understand the evolution divergence between *L. chinense* and *L. tulipifera* and how *DHN* genes are associated with adaptation evolution, we firstly investigated the DNA polymorphisms of the *LcDHN-like* gene in 21 *L. chinense* and 6 *L. tulipifera* populations.

Results: A 707 bp *LcDHN-like* gene was cloned, which included a 477 bp open reading frame (ORF) and coding 158 amino acids. 311 *LcDHN-like* gDNA sequences were obtained from 70 *L. chinense* and 35 *L. tulipifera* individuals. The AMOVA and phylogenetic relationship analysis showed significant differences between the two species. A higher genetic diversity was observed in *L. tulipifera* compared to *L. chinense*, in consistent with the higher adaptive capacity of *L. tulipifera*. Our data also suggested that the *LcDHN-like* genes' polymorphisms were under neutral mutation and purifying selection model in the *L. chinense* and *L. tulipifera* populations, respectively. The distinct expanding range and rate between the two species, haplotypes shared only in *L. chinense's* nearby populations, and wide dispersals in *L. tulipifera* could contribute to the obscure east-west separation in *L. chinense* and entirely unordered phylogeny in *L. tulipifera*. The completely separated nonsynonymous substitution at position 875 and the higher range scope of aliphatic index in *L. tulipifera* populations may be related with its higher adaptive capacity. Taken together, our study suggests *LcDHN-like* gene is a potential mark gene responsible for adaptive evolution divergence in *Liriodendron*.

Conclusions: Significant differences and completely distinct haplogroups between *L. chinense* and *L. tulipifera* showed that the two species have evolved into different directions. The more widely distribution, earlier haplogroups divergence events, and richer SNPs variations in *L. tulipifera* could imply its stronger adaptation in this species. And potential effect of the allelic variations in *LcDHN-like* gene may reflect the difference of water stress and chill tolerance between *L. chinense* and *L. tulipifera*, which could provide some information for further adaption evolution studies of *Liriodendron*.

Keywords: *Liriodendron*, Divergence, Dehydrin, Nucleotide variation, Adaptation

* Correspondence: hgli@njfu.edu.cn

The Southern Modern Forestry Collaborative Innovation Center, College of Forestry, Nanjing Forestry University, Nanjing 210037, China



© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Background

Liriodendron tulipifera and *L. chinense* is an east Asian-east North American sister species pair [1]. They are well-known for rapid growth, valuable wood, attractive leaves and flowers, thus having great ecological and economic potential [1–3]. *L. chinense* distributes in southern China and northern Vietnam as a scattered population pattern [4]. Because of endangering factors including endangering habitat, degraded population structure, low seed germination rates and artificial interference, *L. chinense* was listed in the IUCN Red List of Threatened Species by the International Union for Conservation of Nature and Natural Resources [5–8]. On the contrary, *L. tulipifera* abundantly and widely distributed between 27° and 42° north latitude and predominant in east of the Mississippi River [1, 9, 10].

As an excellent pioneer species, *L. tulipifera* possesses a stronger adaptive capacity than *L. chinense* [11]. Previous studies have shown that *L. chinense* is more sensitive to low-temperature and waterflooding [12–14]. The variations of adaptive genes at DNA level in other plants has been shown to be highly correlated with their adaptation difference, which also could provide a snapshot of evolution divergence. For example, nucleotide variations at several cold candidate genes were surveyed in Scots pine, which showed that significant differentiation in allelic frequency or haplotypes structure between north and south populations was detected at *dhn1*, *dhn3*, and *abaH* loci [15]. The difference of nucleotide diversity, haplotypes, and expression after dehydration of *Dhn1* between ‘African’ dry slope (AS) and ‘European’ humid slope (ES) showed that adaptive natural microclimatic selection is the major evolutionary divergent driving force of *Hordeum spontaneum* in ‘Evolution Canyon’ (ECI) at Lower Nahal Oren, Mount Carmel, Israel [16].

Dehydrin (DHN), also called *LEA-D11* or *LEA II*, belongs to a small gene family of the late embryogenesis abundant (LEA) proteins [17]. DHNs are largely expressed during the later stage of seeds maturation or/and the cell dehydration process caused by environmental stresses, such as drought, cold, and high salinity [18–20]. DHN proteins contain three conserved segments, which are (V/T)D(E/Q)YGNP, (LHRSGS4–10(E/D)3), and EKKGIM(E/D)KIKEKLPG], named Y-, S-, and K-segment, respectively. Based on the appearance and order of these segments, DHN proteins can be classified into five subgroups K_n , SK_n , K_nS , Y_nSK_m , and Y_nK_m [21]. Y-segment, normally found in the N terminus of DHN protein, is similar with partial amino acids of nucleotide binding sites of chaperones in plant and bacterial [3, 21–23]. S-segment, a string of Serine residues, could be phosphorylated and promote DHNs interaction with specific signal peptides for the translocation of DHNs into the nucleus [21, 24, 25]. K-segment, a

Lysine-rich 15-residues, usually exist at the C terminal, could form an amphipathic “ α -helix” like structure, which may play an important role in the effect of hydrophobicity/hydrophilicity [21, 23].

In this study, we firstly investigated the variations of DHN genes’ coding DNA sequence (CDS) among the *L. chinense* and *L. tulipifera* populations. Full-length of *LcDHN-like* gene was cloned by Rapid Amplification of cDNA Ends (RACE) and the corresponding coding protein was predicted. We then explored the evolution dynamics of *LcDHN-like* gene by analyzing the DNA polymorphisms. Second, we conducted population structure, phylogeny and demographic expansion analysis to explore the different evolutionary history in *L. chinense* and *L. tulipifera*, respectively. All together, this study would not only benefit the understanding of the evolutionary divergence interspecies, but also contribute to analyze the evolution of DHN genes and lay the foundation for seeking the difference of water stress, low-temperature adaption in different *Liriodendron* populations.

Materials and methods

Plant materials

The petals for gene cloning were gathered from a 21-year-old tree in April 2015, which located at a provenance trial plantation in Xiashu, Jurong County, Jiangsu Province (119°13′20″E, 32°7′8″N) [26]. Fresh petals were quick frozen and brought back to laboratory storing at -80°C in a freezer prior. Total RNA and DNA were extracted using RNeasy Plant Kit and Plant Genomic DNA Kit, respectively (Tiangen Biotech, China).

27 *Liriodendron* populations, including 21 *L. chinense* and 6 *L. tulipifera*, were sampled from different geographical origins. Each population contains 2 to 5 individuals, altogether, 105 individuals were taken as plant materials for this study (Table 1). Total genomic DNA was isolated from their young leaves or winter buds according to the protocol provided by Plant Genomic DNA Kit (Tiangen Biotech, China).

Full-length cDNA cloning by RACE

The full-length cDNA was obtained by Rapid Amplification of cDNA Ends (RACE). 3′-RACE and 5′-RACE cDNA were synthesized using RACE kits (3′-Full RACE Core Set with PrimeScript RTase, 5′-Full RACE Kit, Takara, Japan). The RACE primers (Additional file 1: Table S1) were designed basing on EST sequence of DHN, which acquired by searching “Dehydrin” annotation in *L. chinense*’s transcriptome database [27]. All the PCR reactions were carried out in a 50 μL reaction mix according to the PCR protocol (Additional file 1: Table S1). The reaction mix consisted of 5 μL 10 \times LA PCR

Table 1 Plant materials and haplotypes of *Liriodendron* populations

Code	Site	Sample size	Sequences number	Haplotype type
ZJ-AJ	Anji, Zhejiang, CHN	4	15	Hap1- Hap6
ZJ-SY	Songyang, Zhejiang, CHN	4	12	Hap7- Hap9
AH-JX	Jixi, Anhui, CHN	3	8	Hap4, Hap5
AH-HS	Huangshan, Anhui, CHN	3	7	Hap5,Hap10,Hap11
JX-LS	Lushan, Jiangxi, CHN	3	9	Hap12, Hap13
FJ-WYS	Wuyishan, Fujian, CHN	3	9	Hap5-Hap7,Hap14- Hap17
HB-XN	Xianning, Hubei, CHN	4	13	Hap18- Hap23
HB-EX	Exi, Hubei, CHN	4	12	Hap20,Hap21,Hap24- Hap26
HN-SN	Suining, Hunan, CHN	3	9	Hap18,Hap27- Hap29
GX-LY	Leye, Guangxi, CHN	3	9	Hap30- Hap33
GX-MES	Maoershan, Guangxi, CHN	3	9	Hap34- Hap36
GX-HP	Huaping, Guangxi, CHN	3	8	Hap33,Hap37, Hap38
GZ-YJ	Yinjiang, Guizhou, CHN	4	11	Hap39- Hap43
GZ-ST	Songtao, Guizhou, CHN	3	9	Hap20,Hap44- Hap46
GZ-XS	Xishui, Guizhou, CHN	2	6	Hap47- Hap49
GZ-LP	Liping, Guizhou, CHN	4	10	Hap50- Hap52
SC-XY	Xuyong, Sichuan, CHN	3	9	Hap48,Hap53- Hap55
SC-YY	Youyang, Sichuan, CHN	3	9	Hap56- Hap58
YN-XC	Xichou, Yunnan, CHN	4	12	Hap34-Hap36, Hap43, Hap58- Hap62
YN-MG	Maguan, Yunnan, CHN	4	8	Hap60
YN-JP	Jinping, Yunnan, CHN	3	8	Hap63, Hap64
Hershey	Pennsylvania, USA	5	14	Hap65- Hap75
BK	North Carolina, USA	6	17	Hap70,Hap72,Hap76- Hap79
MSL	Missouri, USA	6	20	Hap72, Hap80- Hap92
ZZY	Georgia, USA	5	15	Hap68,Hap70, Hap78, Hap93- Hap97
NK	South Carolina, USA	6	21	Hap68,Hap70,Hap76,Hap78,Hap95,Hap96,Hap98-Hap107
LYS	Louisiana, USA	7	22	Hap72,Hap78,Hap108-Hap123
Total	27 populations	105	311	

Buffer II (Mg^{2+} Free), 5 μL MgCl_2 (25 $\text{mmol}\cdot\text{L}^{-1}$), 8 μL dNTP Mixture (2.5 $\text{mmol}\cdot\text{L}^{-1}$ each), 2 μL of each primer (2.5 $\text{mmol}\cdot\text{L}^{-1}$ each), 0.5 μL *TaKaRa LA Taq* (5 $\text{U}\cdot\text{L}^{-1}$), 2 μL each cDNA, 25.5 μL ddH₂O. PCR products were separated on a 1.5% agarose gel and purified using the DNA gel extraction kit (Transgen Biotech, China). The purified PCR products were cloned into pEASY[®]-T1 Cloning Vector and transferred into Trans5 α Chemically Competent Cells for white-blue plaque selection (Transgen Biotech, China). Positive monoclonal was screened for sequencing (Genscript, China). Finally, the gene full-length was assembled according to the 3' and 5' sequencing results.

Open reading frame prediction, verification and analysis

The Open reading frame (ORF) and corresponding protein was forecasted using full-length sequence by ORF

Finder (<https://www.ncbi.nlm.nih.gov/orffinder/>). ORF primers were designed in the untranslated regions (UTR) to ensure getting the whole ORF (Additional file 1: Table S1). cDNA of *L. chinense* was reversed using RevertAid strand cDNA Synthesis Kit (Thermo Scientific, USA). 25 μL 2 \times TransStart FastPfu PCR SuperMix, 2 μL of each ORF primer (2.5 $\text{mmol}\cdot\text{L}^{-1}$ each), 2 μL cDNA, 19 μL ddH₂O were mixed for PCR. The PCR protocol of ORF was same with above, except that inserted PCR products into pEASY[®]-Blunt Cloning Vector (Transgen Biotech, China). ORF and the coding protein were blasted in NCBI to make sure that the cloning result is accurate. The conserved motifs were manually identified according to the conserved sequences (Y, S, K) of DHN. LcDHN-like protein and 13 homologous DHN proteins from other plants (obtaining from the NCBI database) were used to construct an UPGMA tree by MEGA version 6 [28]. Bootstrap values were estimated at 1000 replications.

gDNA cloning of *LcDHN-like* gene among *Liriodendron* populations

The gDNA sequences of *LcDHN-like* gene of all the populations were obtained using the same primers, PCR procedure and cloning methods with ORF. Each individual selected 1–4 positive monoclonal for sequencing and DNA sequences were summarized.

Sequence diversity and selection mode analyses

DNA sequences were aligned using Clustalx1.83 [29]. DnaSP v5 software was employed to achieve statistical estimates of polymorphic sites (S), InDels (Insertion-Deletion) sites, nucleotide diversity (π) and Theta (per site) from S (θ_w) [30].

The departures from the standard neutral model of evolution was evaluated by three models of Tajima's D , Fu and Li's F^* and Fu and Li's D^* statistics with DnaSP v5 [31–33]. Significantly negative value of Tajima's D means existing excess of low frequency polymorphisms, which is consistent with positive directional selection, exhibiting mildly deleterious alleles or a recent population expansion [31, 34, 35]. Significantly positive value of Tajima's D means existing excess of intermediate-frequency polymorphisms, which may be indicative of balancing selection or a population contraction [31, 34, 35]. If purifying or negative selection and advantageous alleles have recently become fixed in the population, the values of Fu and Li's F^* and Fu and Li's D^* are significantly negative [33]. If balancing selection has happened, the values of Fu and Li's F^* and Fu and Li's D^* are significantly positive [33].

The selection pressures were quantified by likelihood ratio test (LRT), which compares dN and dS by paml4.8 package [36]. $dN/dS < 1$ means purifying selection; $dN/dS = 1$ means neutral evolution; $dN/dS > 1$ means positive selection. Z-test of selection with modified Nei–Gojobori algorithm via the Jukes-Cantor was analyzed by MEGA version 6 to test the significance [37]. The population size changes of *Liriodendron* were inferred by DnaSP v5 [30].

Population genetic structure analyses

The AMOVA program within the Arlequinver 3.5.2.2 software package was applied to evaluate the genetic variance between *L. chinense* and *L. tulipifera* populations and within the two species [38]. The number of permutations was set as 1000 and showed significant difference when significance tests at $P < 0.05$ level.

We inferred the population structure of *Liriodendron* by employing the Bayesian clustering algorithm with admixture Model and allele frequencies correlated models in STRUCTURE V2.3 [39, 40]. The K values were ranging from 2 to 27, the upper bound of which was the number of actual sampled populations. Each K value was run 10 times with 100,000 steps after a burn-in period of

10,000 steps. We estimated the most probable K value by StructureHarvester v0.6.93, which was eventually determined by the relationship between ΔK and K [41]. Replicate cluster analyses of the results about optimum K value were performed by CLUster Matching and Permutation Program version 1.1.2 and the final outputs from the Bayesian analyses were visualized clearly by DISTRICT v1.1 [42].

Phylogenetic relationship and geographical distribution of haplotypes

Haplotypes data and diversity of *LcDHN-like* gene in *Liriodendron* were generated by DnaSP version 5.0 [30]. The original UPGMA tree and Time Trees of haplotypes were constructed by MEGA version 6 with 1000 bootstrap replications [37]. The calibration constraints was set according to the separating time of *L. chinense* and *L. tulipifera*, about 10–16 million years ago, by which MEGA could produce absolute divergence times for all branching points in the tree based on the RelTime method [28, 43]. And, representative divided haplogroups were summarized by sketchy phylogenetic tree.

The phylogenetic relationships among haplotypes were constructed by Median-Joining network in NETWORK version 4.2.0.1 [44, 45]. For a high quality final graphics, the Median-Joining network was enhanced by another Reduced-Median network to simplify the outcome (All parameters were designated as the default values) [45].

The geographical distribution of haplotypes from sampled sites was marked on the map of China and American. And different *LcDHN-like* haplotypes in each sample site were drawn as circle pie proportionally by Adobe Illustrator CS6 [46].

Demographic history analysis

Pairwise mismatch distributions were carried out by Arlequinver 3.5.2.2 to infer the historical demography of *L. chinense* and *L. tulipifera*, with the expected frequency based on a constant population size model [30]. Two specific parameters, the sum of squared deviations (SSD) and Harpending's raggedness index (HRag), were used to test the goodness of fit under a spatial expansion model. Their significances were tested using a parametric bootstrap approach with 1000 replications, and showed significant difference when significance tests at $P < 0.05$ level. $\tau = 2ut$ was used to calculate the recent expansion time, where u ($u = \mu k$) is the mutation rate for the whole haplotypes. μ , calculated by eq. 9.62 of Nei, is the mutation rate per nucleotide; k is the number of nucleotides in the sequence [47, 48].

The population history of *L. chinense* and *L. tulipifera* was also investigated using Bayesian Skyline Plot (BSP) in BEAST v.1.4, which employed a coalescent Bayesian Skyline tree prior and uncorrelated relaxed clock fixing

prior substitutions rate at 2.0×10^{-8} per year [49]. The GTR nucleotide substitution model for sequence evolution was estimated by MEGA version 6 [37]. Each MCMC sample was based on a run of 10,000,000 generations, sampled every 1000, with the first 1000,000 generations discarded as burn-in. Bayesian Skyline Plots were summarized by Tracer v1.7 [50].

Protein structure prediction

The exons of *LcDHN-like* gene were selected and translated into proteins by MEGA 6. These proteins were aligned and then classified by pairwise distance in MEGA 6. ExPasy ProtParam (<http://web.expasy.org/protparam/>) was applied to compute the various physical and chemical parameters of *LcDHN-like* proteins. The secondary structure of proteins was predicted using SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html).

Results

Isolation and characterization of *LcDHN-like* gene

A 707 bp full-length cDNA was assembled with the 468 bp 5' sequence and 423 bp 3' sequence (Additional file 2). The prediction and experimental verification proved that *LcDHN-like* gene contained a 477 bp open reading frame (ORF), which encoded a 158 amino acids polypeptide (Fig. 1). The coding region was presented in the genomic sequence to analyze the gDNA structure. *LcDHN-like* gene was composed of two exons (186 bp and 291 bp), which were separated by one 464 bp intron (Fig. 1). Moreover, a 21 bp 5' Untranslated Region (UTR) and a 110 bp 3' UTR were obtained by the ORF primers (Fig. 1). The intron of *LcDHN-like* gene falls to obey the GT-AG rule of exon-intron borders [51]. ORF and amino acids sequences were used to blast against Nucleotide collection (nr/nt) database from Genebank, which indicated it is a homolog of *DHN*. We then named it *LcDHN-like*. The deduced protein product possesses 2 Y-segments, 1 S-segment, and 2 K-segments, suggesting it is Y_nSK_n -type DHN protein (Fig. 1). This result was further confirmed by a phylogenetic analysis that showed *LcDHN-like* protein clustered into the Y_nSK_n subgroup, including homologs of Y_nSK_n -type proteins from *Arabidopsis thaliana*, *Daucus carota*, *Triticum aestivum*, *Jatropha curcas*, and *Coffea canephora* (Fig. 2) [52–55].

DNA diversity in *Liriodendron* populations

311 *LcDHN-like* gDNA sequences were obtained from 70 *L. chinense* and 35 *L. tulipifera* individuals representing 27 populations (Additional file 3). After alignment, in total, 233 SNPs (19.73%) and 152 indels (12.87%) were found in 1181 bp nucleotide positions (Table 2). Of the 233 SNPs, 102, 113, and 18 SNPs were located in exon, intron, and UTR region, respectively (Table 2, Fig. 3). At

a whole population level, the nucleotide diversity, π was 0.02881 and θ_w was 0.03585. π was observed as 0.02890 and 0.03030, as well as θ_w was 0.03542 and 0.04021, in the coding and intron regions, respectively.

Table 2 showed that 92 SNPs (8.50%) and 30 indels (2.77%) were found in the populations of *L. chinense*. Of the 92 SNPs, 30 SNPs exist in exon (12 SNPs in exon1, 18 SNPs in exon2), 55 in intron, and 7 in UTR. The π for the whole sequence, exon and intron are 0.00986, 0.00613, and 0.01410, respectively. The θ_w for the whole sequence, exon and intron are 0.01487, 0.01097, and 0.02046 respectively. In general, the intron does not have clear function and evolving under neutral with higher polymorphism. Higher selection effect in exon region leads to its lower polymorphism and stronger conservation.

161 SNPs (13.67%) and 129 indels (10.95%) were found in *L. tulipifera* populations, including 76 in the exon (32 in exon1, 44 in exon2), 75 in intron (Table 2), and 10 in the 3' UTR. No SNPs were discovered in the 5' UTR. The π for the whole sequence, exon and intron are 0.03348, 0.03891, and 0.03473, respectively. The θ_w for the whole sequence, exon and intron are 0.02916, 0.03085 and 0.03152, respectively. These results indicate that *L. tulipifera* populations possess much higher level of DNA polymorphism compared to the *L. chinense*.

Neutrality tests

We then evaluated the evolutionary selection dynamic of *LcDHN-like* gene using four neutrality tests (dN/dS, Tajima's D, Fu and Li's D*, and Fu and Li's F*). In all studied populations, dN/dS ratio was found to be significantly less than 1. Tajima's D, Fu and Li's D* and Fu and Li's F* were -0.81939, -2.64830, and -2.01595, respectively. Only Fu and Li's D* showed significant (Table 3). Of four neutrality tests, only dN/dS ratio showed significance within the *L. chinense* populations (Table 3). However, in the *L. tulipifera* populations, all four neutrality tests showed no significance, suggesting the *LcDHN-like* gene does not reject the neutral mutation hypothesis (Table 3).

Molecular variance analysis between *L. chinense* and *L. tulipifera* populations

To survive in continuously changes of environment, *Liriodendron* has adapted to a number of distinct environmental conditions. We then performed AMOVA analysis to evaluate more in detail about the differences of *LcDHN-like* gene interspecies (Table 4). The results indicate that most of the variation (61.82%) contributed by the difference between the two species with a highly significant F_{SC} value (0.17) (Table 4). Within species, the variation was contributed mainly by the difference among the individuals (31.78%) and less by populations (6.40%) (Table 4).

```

1  AGCCGTTGGATCTAGTTGAAA
22 ATGTCAGAGACGCGTGATGAGTATGGCAACCAGGTTCCGCAAACC
 * M S E T R D E Y G N Q V R Q T 15
67 GACGAGTATGGGAACCCGATTCAGCATACTGGCACCGGGACAAAG
  D E Y G N P I Q H T G T G T K 30
112 CCCGTTCCGGGCATACATGGTGGGGGCCATGGGATTGGCACAGGT
  P G S G I H G G G H G I G T G 45
157 GGTGGTGGTGGACAAGGCAAGCTCCACCGTCCAGGCTCCGGCTCT
  G G G G Q G K L H R P G S G S 60
202 TCCTCT
   S S 63
208 GTAAGTATATTCGATGTTTCTCAAGGGCCATTTCACTTGGTATGT
253 CGTGTTGGGATGACTAGATTAGTTGATTAACGGTTGTGATTGTAA
298 ATCAGTGATATGGGCTTCTGTGTACCTTATTTCAACAATGATAATC
343 AATTGATACGCTGGATCGTGATGAGCGGACCGGATGAGCCGTCCA
388 TCTTACAGGCTTTTTAAAATAGAATTATAGGGTGGGTGGTATGAG
433 GTCTACCAAGCCTGTAACATGGAATTACAATAATAGCTGTTAATC
478 GGTTCATCTCATCCGCTCATCGTGATCCAGCATGTCAAACAGAGA
523 TGGGTTTATGGAGAAATGGATGTATGTAAGCCAACAAATCACAAA
568 AAAAATCTGACATAATGATGGAGAGAATGCAAGGAGTGAAGTGA
613 ATCTACGCCGTTAAATGGATTTCATTGTACTTAATTCGTTTTGT
658 CAAATATTTTGTAG
672 GACGAGGATGATGGACAAGGTGGGCGTAGGAAGAAGGGT
   D E D D G Q G G R R K K G 75
711 TTGACGGAGAAGATCAAAGAGAAGCTGCCAGGTGGAACAAGACG
   L T E K I K E K L P G G N K T 90
756 ACAGGTGTCTGTCATCCAGGGACTCAGGGAGTACAGGGGGGTCCG
  T G V C H P G T Q G V Q G G R 105
801 GAGCATGAGAAGCCAGGTTGCGGGCAGGGTGACAGGTGGGCCGC
  E H E K P G C G Q G G Q V G R 120
846 GAGCATGAGAAGACAGGTTTCGGTCATCCAGGAACTCAGGGTGG
  E H E K T G F G H P G T Q G G 135
891 CAGGTGGGCAGCGAGCAGGAGAAGAAGGGTATGATTGAGAAGATC
  Q V G S E Q E K K G M I E K I 150
936 AAGGAGAAGCTGCCAGGTCACAAGTAG
   K E K L P G H K ** 158
963 GATGTGTGCCACCATACACGTGGAGCCTACATCTACGTGTCTTA
1008 TGTTACTTATAATACGCATGGAGTGTGTCCGGAGTCTGTAATAAT
1053 CTGCGCGTATCTGTTATGTC
    
```

Fig. 1 Sequences analysis of *LcDHN-like* gene. Note: 1. Nucleotide and amino acid numbers are counted on the left and right separately. 2. The ORF split into two exons (exon 1:22-207 bp; exon 2:672-962 bp, coding 158 amino acids) by one intron (209-672 bp). 3. The Y- segments are highlighted in light grey; The S-segment is highlighted in grey; The K-segments are highlighted in black. 4. Start and stop codons are indicated by asterisks and double asterisks, respectively. 5. Nucleotides are boxed represent 5'and 3' untranslated regions

AMOVA analysis within the species showed that the variation stored at population level in *L.chinense* (39.31%) is bigger than *L. tulipifera* (11.52%) (Fig. 4a, Fig. 4b). However, at individual level, the total variation in *L. chinense* (60.69%) is lower than *L. tulipifera* (88.48%) (Fig. 4a, b). Consistently, the F_{ST} value in *L. chinense* populations (0.39) is bigger than *L. tulipifera* (0.12) (Fig. 4c). These results suggested that *L. tulipifera*

possesses higher genetic diversities at an individual level but less genetic differentiation among the populations compared to that of *L. chinense*.

Two distinct lineages in *Liriodendron*

A total of 123 haplotypes were defined by 233 SNPs, including 64 from *L. chinense* and 59 from *L.tulipifera*. The two species dependent were clustered into two

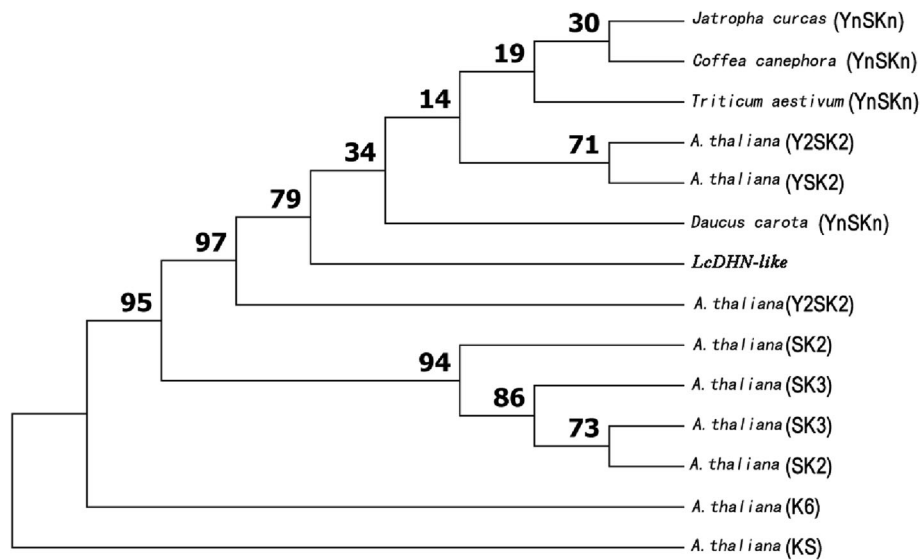


Fig. 2 The UPGMA tree based on LcDHN-like protein and 13 homologous DHN proteins from other plants. Note: 1. The bootstrapping was set 1000 replicates and bootstrap values were labeled beside the branches. 2. *Jatropha curcas*: NP_001295638; *Coffea canephora*: ABC55671; *Daucus carota*: BAD86644; *Triticum aestivum*: AOM63238; *A. thaliana*: NP_201441; *A. thaliana*: NP_190667; *A. thaliana*: NP_179744; *A. thaliana*: NP_195554; *A. thaliana*: NP_177745; *A. thaliana*: NP_564114; *A. thaliana*: NP_173468; *A. thaliana*: NP_190666; *A. thaliana*: NP_175843

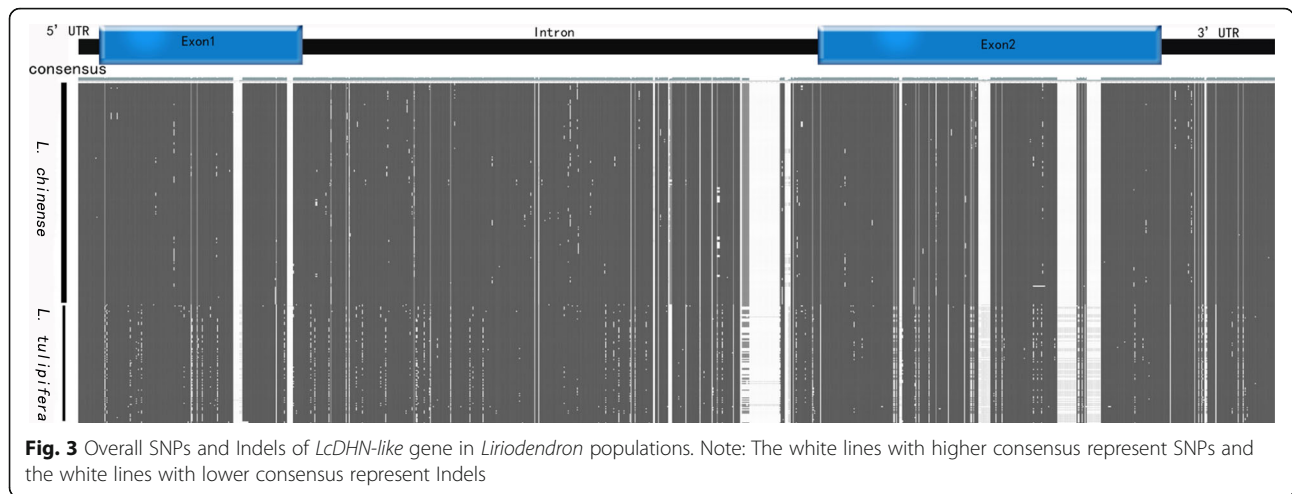
separated clades in the UPGMA phylogenetic tree and network analysis (Fig. 6). Though $K = 4$ was considered as the most adequate number of clusters, all *Liriodendron* populations were divided into 2 major genetic clusters: *L. tulipifera* (green) and *L. chinense* (red) (Fig. 5). Six specific SNPs were identified for distinguished *L. chinense* and *L. tulipifera*, including 3 in exons, 2 in intron, and 1 in 5'UTR (Table 5).

For *L. chinense* populations, the most frequent frequencies of haplotypes were H5, H7, H6 and H18, which were observed 12, 11, 10 and 8 times with frequencies of 5.94, 5.45, 4.95 and 3.96%, respectively. While H (63, 33, 4), H (43, 24, 21, 12), H (58, 56, 53, 52, 48, 34), H (50, 44, 39, 36, 31, 20, 10), H (37, 35, 13, 3, 1) and H (49, 45, 28, 27, 23, 19, 16, H6) were happened 7, 6, 5, 4, 3, 2 times with frequencies of 3.47, 2.97, 2.48,

Table 2 The nucleotide diversity of *LcDHN-like* gene in *Liriodendron* populations

		Exon-1	Exon-2	Exon	5'UTR	Intron	3'UTR	Total
<i>Liriodendron</i>	Size	201	339	540	21	508	112	1181
	SNPs	42	60	102	2	113	16	233
	Indels	21	63	84	0	63	5	152
	π	0.02529	0.03125	0.02890	0.00061	0.03030	0.02776	0.02881
	θ_w	0.03695	0.03442	0.03542	0.01508	0.04021	0.02368	0.03585
<i>L. chinense</i>	Size	186	291	477	21	474	110	1082
	SNPs	12	18	30	2	55	5	92
	Indels	0	12	12	0	17	1	30
	π	0.00573	0.00640	0.00613	0.00094	0.01410	0.00964	0.00986
	θ_w	0.01097	0.01097	0.01097	0.01619	0.02046	0.00780	0.01487
<i>L. tulipifera</i>	Size	201	339	540	21	506	111	1178
	SNPs	32	44	76	0	75	10	161
	Indels	21	51	72	0	54	3	129
	π	0.03237	0.03757	0.03891	0	0.03473	0.02572	0.03348
	θ_w	0.03377	0.02902	0.03085	0	0.03152	0.01759	0.02916

SNPs single nucleotide polymorphism sites, Indels insertion or deletion sites, π nucleotide diversity, θ_w theta (per site) from S



1.98, 1.49 and 0.99%, respectively. The other haplotypes were unique with a frequency of 0.50%. For *L. tulipifera* populations, the most frequent frequencies of haplotypes were H78, H72 and H70, which were observed 14, 8 and 7 times with frequencies of 12.84, 7.34 and 6.42% respectively. While H (76, 68), H (103, 80), H118 and H (117, 115, 105, 96, 95, 93, 85, 77) were happened 5, 4, 3 and 2 times with frequencies of 4.59, 3.67, 2.75 and 1.83%, respectively. The other haplotypes were unique with a frequency of 0.92%. The numerous haplotypes with relative lower frequencies were corresponding with the higher haplotypes diversity of studied *Liriodendron* individuals (0.9870). Both of the two species' lineage networks contained multiple clades, indicating a complex relationship patterns between populations (Fig. 6). In *L. chinense*, the eastern populations, as a big clade, separated from western populations, while in *L. tulipifera*, no obvious intraspecific phylogeographic pattern was found (Fig. 6b).

Haplotypes geographic distribution of *Liriodendron*

For *L. chinense*, haplotypes H (4, 6, 7, 18, 21, 33, 34, 35, 36, 43, 48, 58, 60), H20 and H5, as shared haplotypes, were only found in nearby 2, 3 and 4 populations respectively (Fig. 7). For *L. tulipifera*, H (76, 95, 96), H68 and H (70, 72, 78), as common and dispersed haplotypes, were presented in 2, 3 and 4 populations respectively (Fig. 7). The remaining haplotypes for both *L.*

chinense and *L. tulipifera* were restricted in a single population (Fig. 7). Geographic distribution of haplotypes suggested that all populations contained their own particular haplotypes. Although existing shared haplotypes with higher frequencies, no prominent central and ancestral haplotypes were found in gene genealogies of *L. chinense* and *L. tulipifera* populations.

The analysis of divergences times and demographic expansion in *Liriodendron*

As a Timetree, the UPGMA phylogenetic tree also showed the divergent time of haplotypes in *Liriodendron* populations, in which the earliest haplotypes divergence time were approximately 4.62 and 11.7 Ma ago in *L. chinense* and *L. tulipifera*, respectively (Fig. 6a). The completely separated western haplotypes in *L. chinense* were happened about 3.67 Ma ago.

The values of SSD/H_{Rag} were 0.205067/0.480071 and 0.480071/ 0.436351 with nonsignificant *p*-value in both *L. chinense* and *L. tulipifera* populations, which supported a recently demographic expansion in these two species. In view of the tau value (τ) and the mutation rates of *LcDHN-like* gene (2.0×10^{-8} per year), we deduced that 0.263 Ma and 2.02 Ma ago would be the time of recent population expansion to *L. chinense* and *L. tulipifera* populations respectively.

The demographic history of *Liriodendron*

As showed in the UPGMA phylogenetic tree (Fig. 8b), five main haplogroups, C1, C2, C3; T1, T2, were obviously grouped on account of separated time nodes. Haplogroup C1 was found in most western populations except Jingping and Maguang from Yunnan. Though haplogroup C2 contained populations from eastern and western China, it didn't present at Songtao, Liping from Guizhou; Leye, Maoershan, Huaping from Guangxi, Sichuan-Youyang, Hubei-Exi and Anhui-Huangshan.

Table 3 Neutrality tests values of *LcDHN-like* gene based on nucleotide variation

	dN/dS ratio	Tajima's D	Fu & Li's D*	Fu & Li's F*
<i>Liriodendron</i>	0.18166*	-0.81939	-2.64830*	-2.01595
<i>L. chinense</i>	0.06065*	-1.12616	-0.42516	-0.89998
<i>L. tulipifera</i>	0.32699	0.24844	-1.88595	-1.14004

*significant; *P* < 0.05

Table 4 Analysis of molecular variance of *LcDHN-like* gene of *Liriodendron* populations

Source of variation	Degrees of freedom	Sum of squares	Variance components	Percentage variation	Fixation Indices	P-Value
Among species	1	3862.12	26.87	61.82	$F_{SC} = 0.17$	< 0.001*
Within species among populations	25	1130.85	2.78	6.40	$F_{ST} = 0.68$	< 0.001*
Within populations	284	3921.82	13.81	31.78	$F_{CT} = 0.62$	< 0.001*

*indicates that Fixation Indices showed significant difference when significance tests (1023 permutations) were at $P < 0.001$ level

The haplogroup C3, a widely distributed haplogroup across southern China from east to west, presented in most populations except four populations as Songtao, Liping, Maoershan, Jingping, and Maguan. Haplogroup T1 and T2 were both common in North America, while T2 included more numerous haplotypes. We performed BSPs for the total populations of *L. chinense* and *L. tulipifera* and haplogroup C1, C(2, 3); C(1, 2); C3; T1; T2 to detect their changes in effective population size over time.

The Bayesian skyline plot (Fig. 8a) indicated that both *L. chinense* and *L. tulipifera* populations experienced demographic expansions around 0.5 to 1.5 Ma ago, Pleistocene, which were consistent with the results of pairwise mismatch distributions analysis. Figure 8a also showed that the effective population size of *L. chinense* was smaller than that of *L. tulipifera* in the beginning, followed by a slow and rapid growth in both species, and finally, the effective population size of *L. tulipifera* increased much higher than that of *L. chinense*. Thus, we deduced that the population expanding rate of *L. tulipifera* was rapider than that of *L. chinense*. The haplogroup C1 did not undergo large fluctuations in population size, whereas a marked increase of effective population size in haplogroup C(2, 3) was estimated about 3.3 Ma ago (Fig. 8c). Haplogroup C(1, 2) and C3 were severally experienced population expansion during 2–2.25 Ma and 0–1 Ma respectively, however, no

obvious change of population size was found about 4 Ma ago in haplogroup C3 (Fig. 8d). Figure 8e showed that haplogroup T1 kept a long history of nearly constant population size, before a recent population expansion over the last 0.3–1.0 Ma. A slow population growth during 2.5–3.5 Ma was observed in haplogroup T2 (Fig. 8e).

The variations of LcDHN-like proteins

Fifty-eight LcDHN-like protein isoforms were identified from the studied populations, which derived from non-synonymous SNPs or Indels. There are 17 and 41 specific isoforms for *L. chinense* and *L. tulipifera*, respectively (Additional file 4). In the studied populations, the LcDHN-like proteins ranges from 154 to 175 AA length, resulting molecular weight range between 15.94 and 18.00 kDa (Additional file 1: Table S2). Generally, LcDHN-like proteins in *L. tulipifera* are slightly bigger than that of *L. chinense*. The hydrophilic index ranges from 32.66 to 39.63, in which *L. chinense* contributed more to the lower part of this range (32.66–36.33) and *L. tulipifera* more to the higher part (32.99–39.63).

Discussion

The polymorphism of LcDHN-like gene in *L. chinense* and *L. tulipifera* populations

In this study, we investigated the *LcDHN-like* gene polymorphisms from 21 *L. chinense* populations and 6 *L. tulipifera* populations. Interestingly, we observed

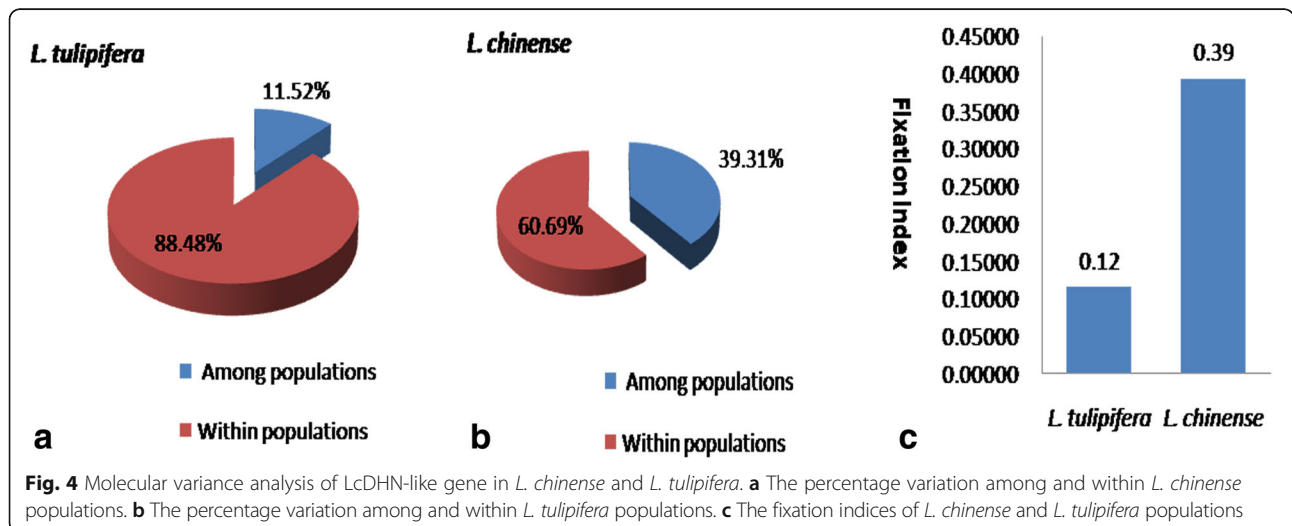


Fig. 4 Molecular variance analysis of LcDHN-like gene in *L. chinense* and *L. tulipifera*. **a** The percentage variation among and within *L. chinense* populations. **b** The percentage variation among and within *L. tulipifera* populations. **c** The fixation indices of *L. chinense* and *L. tulipifera* populations

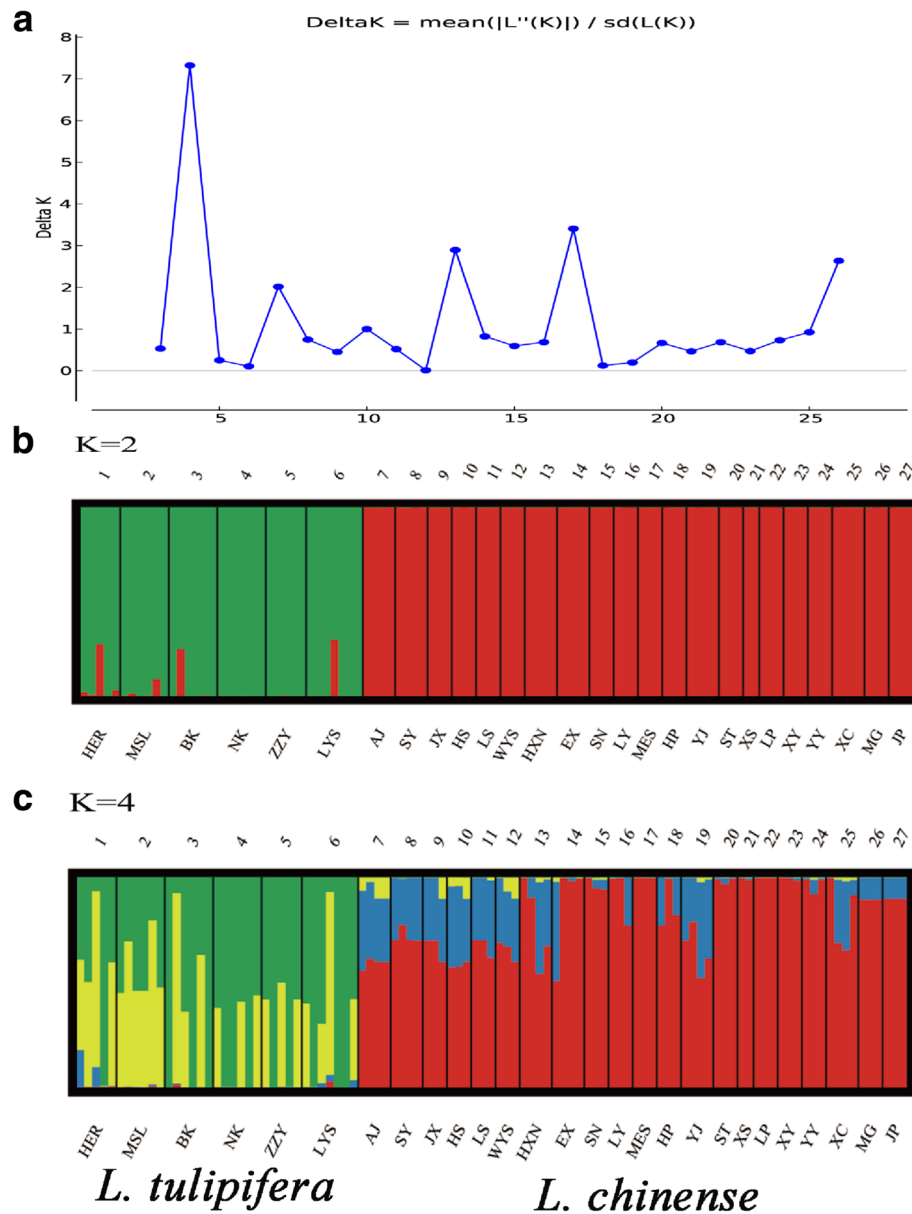
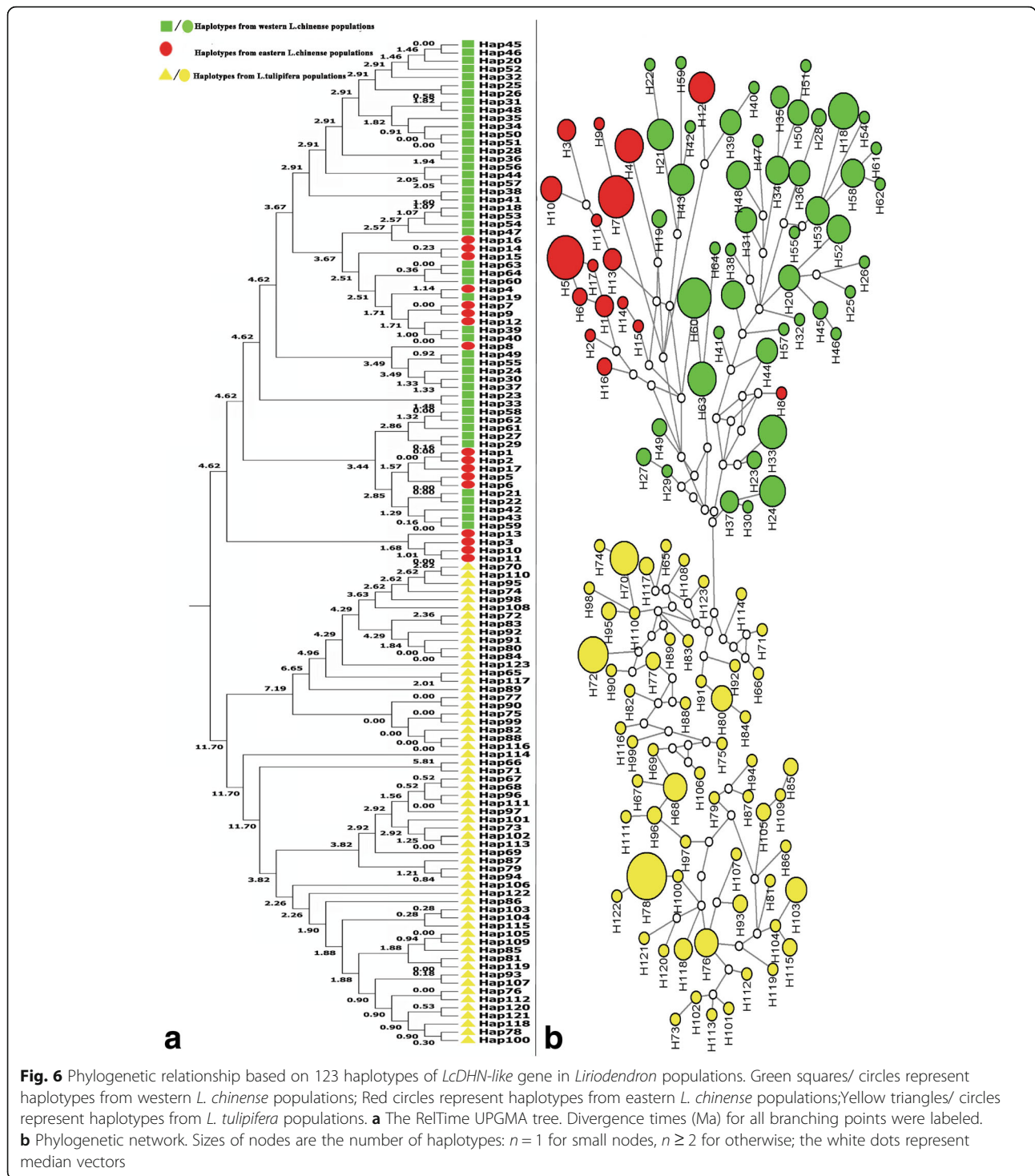


Fig. 5 Population structure of *Liriodendron* populations. **a** The StructureHarvester result of relationships between ΔK and K . **b** the clusters result of $K=2$. **c** the clusters result of $K=4$

Table 5 The completely separated positions between *L. chinense* and *L. tulipifera* populations

region	position	Variation	<i>L. chinense</i>	<i>L. tulipifera</i>	Codon	Amino acid	Effect
Exon-2	733	C → T	C	T	GAC → GAT	Asp → Asp	SC
	859	T → C	T	C	GGC → GGT	Gly → Gly	SC
	875	C → A	C	A	CCA/CAA → ACA	Pro/Gln → Thr	NSC
Intron	647	A → G	A	G			
	703	-C → T/G	-C	T/G			
5'UTR	1078	T → C	T	C			

NSC nonsynonymous change, SC synonymous change



much higher genetic diversity in *L. tulipifera* populations which was sampled from a region with a smaller scale of Appalachian uplands and the southeastern coastal plains, compared to *L. chinense* (Tables 1 and 2). These results are consistent with previous studies based on expressed sequence tag derived SSR (EST-SSRs) markers and RAPD markers [56, 57].

Moreover, this precise survey of diversity at the nucleotide level provided a snapshot of the *LcDHN-like* gene evolution and represented its reservoir of genetic diversity for short-term (ecological) and/or long-term (evolutionary) adaptation [58–61]. Therefore, our data suggest that *L. tulipifera* possesses a higher genetic diversity which might result in a stronger adaptation

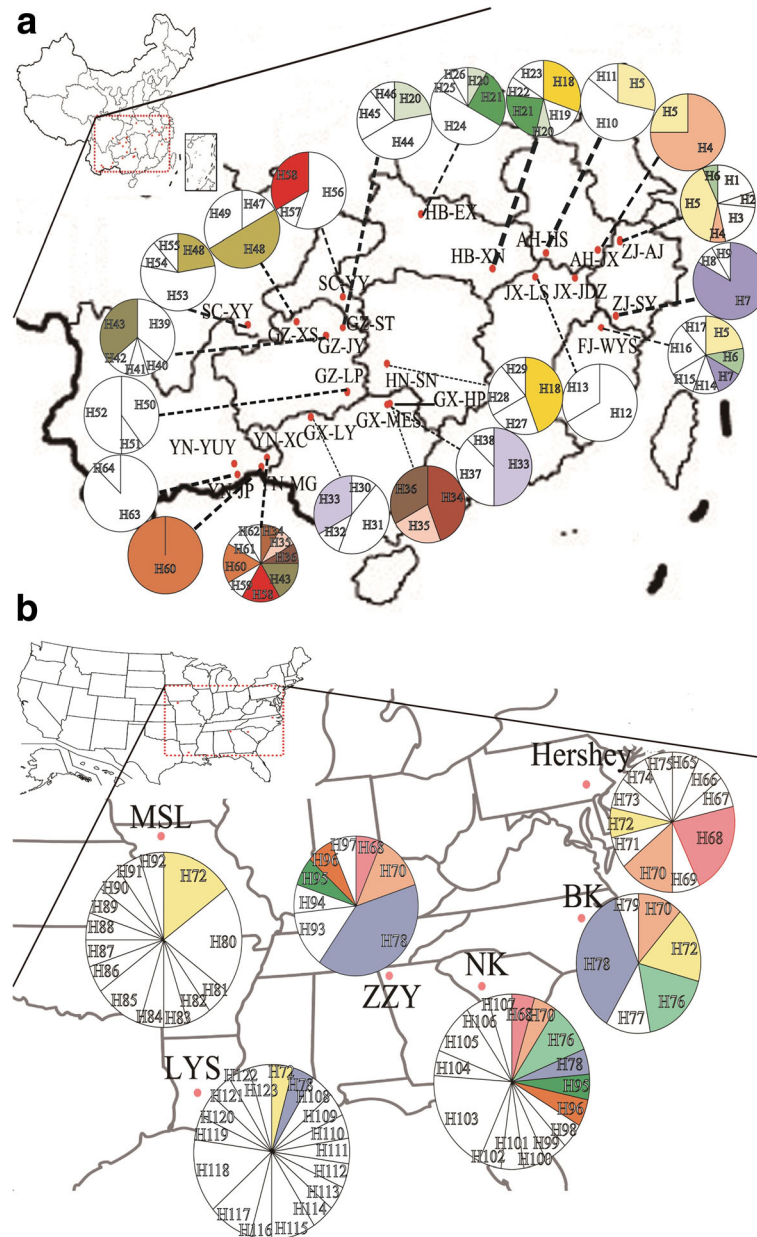


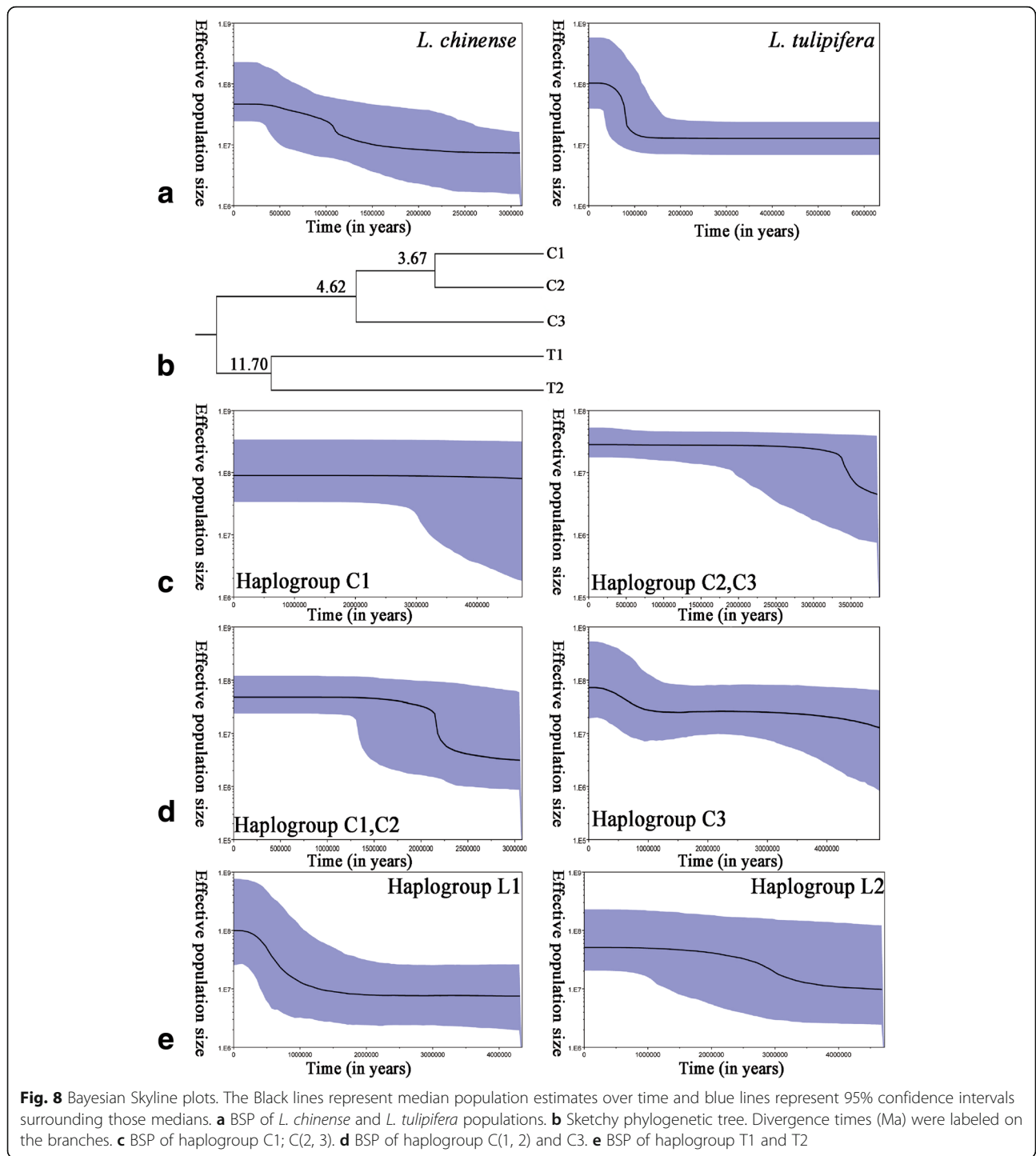
Fig. 7 Geographic distribution of *LcDHN-like* haplotypes about *Liriodendron*. The white parts of circles represent the unshared haplotypes, the circles with same colour represent they share same haplotypes; the divided area of each circle corresponds to the frequency of each haplotype. The maps of CHN and USA were downloaded from <https://commons.wikimedia.org/wiki/File:China-map.xcf> and https://upload.wikimedia.org/wikipedia/commons/c/ca/Blank_US_map_borders.svg, separately. **a** The haplotypes distribution of *L. chinense* in China. **b** The haplotypes distribution of *L. tulipifera* in America

and resilience to the unforeseen environmental changes.

The selection patterns of *LcDHN-like* gene in *L. chinense* and *L. tulipifera* populations

All the studied 27 *Liriodendron* or 21 *L. chinense* populations showed dN/dS values that were significantly less than unity, suggesting that this gene was under purifying selection in these populations [32, 62]. The significant

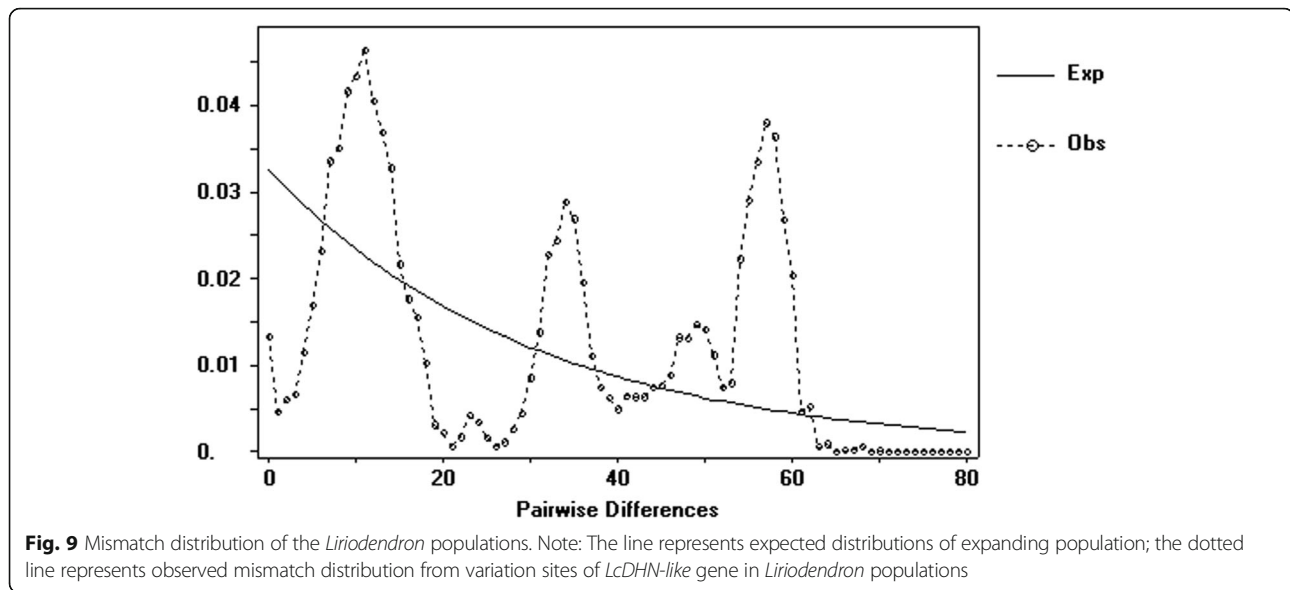
negative value of Fu and Li's D* may indicate events of expansion or selection of the *Liriodendron* populations during the evolution [34]. The mismatches distribution normally exhibits a Poisson distribution in expanding populations, while a variety of geometric distributions were often observed for populations with constant size [63]. The distribution we observed for *LcDHN-like* gene mismatches confirmed that the studied *Liriodendron* populations are under purifying selection (Fig. 9). Taken



together, these results indicated that a natural selection is occurring to suppress LcDHN protein mutations and reduce the nucleotide diversity [64].

Intriguingly, no significance was detected in the neutrality tests for *L. tulipifera* populations, indicating that the gene in this species was under the neutral mutation hypothesis. The neutral theory thought that most

evolutionary changes at the molecular level are caused by natural drift of selectively neutral or nearly neutral mutations rather than by natural selection [65]. Therefore, these results suggested that *LcDHN-like* gene in *L. tulipifera* currently are not under a natural selection. On the contrary, *L. chinense* is under a purifying selection resulting in its lower genetic diversity [66].



The genetic structure difference between *L. chinense* and *L. tulipifera* populations

The AMOVA analysis has shown that the greatest variation of *LcDHN-like* gene sequence in *Liriodendron* populations came from interspecies, which presumably caused by their long time geographic separation. *L. chinense* and *L. tulipifera* were separated about 10–16 Ma ago according to the molecular and paleobotanical evidences, which indicated restricted gene flow with isolation by distance [43]. Fossil floras data indicated that *Liriodendron* may not survived in Beringia after the late Miocene [43]. And the Atlantic Ocean broke the North American-European connection [67]. These two events resulted in a breakage of gene flow between North American and Asian. In summary, the significantly variation between *L. chinense* and *L. tulipifera* populations, which is consistent with completely distinct phylogenetic relationship, were the result of distance and evolution under different environment for a long time.

L. chinense showed greater variation than *L. tulipifera* at population level and lower variation at individual level. *L. tulipifera* populations distribute more widely, causing possibly higher gene flow, which could decrease variation among populations [1, 10]. On the contrary, the small, scattered and isolated *L. chinense* populations are more likely facing barriers of gene flow, which could increase the variation among populations [1].

The distinct expanding range between *L. chinense* and *L. tulipifera* populations

The same lineages of *L. chinense* and *L. tulipifera*'s populations could contribute to the ambiguous phylogeographic patterns detected in the two species. And the shared haplotypes related to demographic expansion in Pleistocene

also facilitated the unordered phylogeographic patterns. Unlike the disordered haplotypes of *L. tulipifera*, the eastern populations of *L. chinense* separated from western population as a big clade in the haplotypes networks and coincided with the biogeographic districts described by Hao and He [68, 69]. The distinct expanding range, haplotypes shared only in nearby *L. chinense* populations while wide dispersals in *L. tulipifera* populations could contribute to the obscure east and west separation in *L. chinense* and entirely unordered phylogeny in *L. tulipifera*.

The haplotypes separation and diverse demography history between *L. chinense* and *L. tulipifera*

The completely separated haplotypes and the distinguished mutation sites between species have been proposed to be the consequence of their apart evolution under different environments for a long time [43, 70]. Thus, climatic oscillations and past geology were important to understand their demography and evolution history [70, 71]. Hereinafter, we took the inferred haplogroups patterns, together with their divergence time in *L. chinense* and *L. tulipifera* and relevant paleo-climatic fluctuations for further discussing.

Our phylogenetic analysis suggested that the earliest divergence time among haplotypes in *L. tulipifera* was occurred about 11.7 Ma ago, during the late Miocene (Fig. 6a). This event might be the consequence of southward migrating Arctic air masses during the late Miocene, which led to the warm temperate mesophytic vegetation arise range restriction [43]. The separated haplogroups T1 and T2 expanded during 0.3–1.0 Ma and 2.5–3.5 Ma with different growth trend, which could contribute to the further haplotypes divergence of Quaternary period (Fig. 6a, Fig. 8e). In *L. chinense*, the earliest divergence time of

haplotypes was approximately 4.62 Ma ago. Considering that no apparent population changes were found 4 Ma ago in haplogroup C3, we deduced that the divergence of haplotypes might be the result of continued cold environment during the late tertiary [72, 73]. The earlier haplotypes divergence event in *L. tulipifera* might reflect the earlier adaptation to cold environment, which contributed to the stronger cold tolerance of *L. tulipifera* when compared to *L. chinense*.

The completely separated western haplotypes in *L. chinense* occurred about 3.67 Ma ago, which coincided with the rapid expansion signal in haplogroup C(2, 3) about 3.3 Ma ago (Fig. 8c). And the warm period in Northern Hemisphere during Middle Pliocene (~ 3 Ma) also could stimulate the expansion of haplogroup C(2, 3) [74]. The nearly identical time of separation and expansion might reflect the founder population of haplogroup C(2, 3) from which the lineages C1 arose [75].

The LcDHN-like protein variation between *L. chinense* and *L. tulipifera* populations

LcDHN-like protein (15.94–18.00 kDa) could be classified as a stable DHN protein ranging from 9kD to 200kD [21]. The DHNs are part of the intrinsically disordered proteins expressed under conditions of water-related stress [23]. Hydrophilic properties reinforce the flexibility and the disorder aspect that can be conferred to LcDHN-like protein [76]. The lower hydrophilic indices observed in *L. chinense* suggested potential stronger abiotic stresses resistance of LcDHN-like proteins in *L. tulipifera* compared to *L. chinense*. The observed nonsynonymous nucleotides substitutions might affect LcDHN-like proteins' function and result in adaptation difference between *L. chinense* and *L. tulipifera*, which would be very interesting to investigate further in the future [77].

Conclusions

In this paper, all of the investigated results from AMOVA, population structure and phylogenetic relationship provided a sharp phylogeographic break between *L. chinense* and *L. tulipifera* populations. We also provided comprehensive and important insights into the degree of interspecies evolutionary divergence by analyzing DNA polymorphism difference, evolution dynamics, molecular variation and haplotypes of *LcDHN-like* gene between the two species. The distinct expanding range and rate between the two species, haplotypes shared only in *L. chinense's* nearby populations while widespread in *L. tulipifera*, could contribute to the unclear east-west separation detected in *L. chinense's* network tree and unordered phylogeographic pattern in *L. tulipifera*. And the more widely distribution, earlier haplotypes divergence events and richer SNPs variations in *L. tulipifera* could imply their stronger adaptation.

Additional files

Additional file 1: Table S1. Primers and PCR protocol for RACE amplification and ORF testing. **Table S2.** Characteristics of properties and structure about LcDHN-like proteins in *Liriodendron*. P1-P17, proteins from *L. chinense*; P18-P58, proteins from *L. tulipifera*. (DOCX 27 kb)

Additional file 2: 5' RACE, 3' RACE and cDNA sequences of *LcDHN-like* gene. (DOCX 15 kb)

Additional file 3: 311 *LcDHN-like* gDNA sequences in *Liriodendron*. (DOCX 152 kb)

Additional file 4: The list of 58 predicted proteins in *Liriodendron*. (DOCX 19 kb)

Abbreviations

dN: Number of non-synonymous substitutions per non-synonymous site; dS: Number of synonymous substitutions per synonymous site; F_{ST} : Fixation index; Indels: Insertion/Deletion sites; SNPs: Single nucleotide polymorphisms; Θ_w : Scaled measure of the number of polymorphic nucleotide sites per nucleotide; π : Nucleotide diversity

Acknowledgements

We are grateful to Xi Wang, Weiping Zhong, Ziyuan Hao, Tengfei Zhu, Weijie Si for volunteering to assist with laboratory work. The authors also thank Bo Zhang for revising the manuscript.

Funding

This study was financially supported by grants from the National Natural Science Foundation of China (No.31770718 and No.31470660), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD). The funding bodies had not any role in the design or implementation of this study or writing the manuscript.

Availability of data and materials

All data generated or analysed during this study are included in this manuscript and its supplementary information files.

Authors' contributions

YLC designed the study, performed the experiment, analyzed the data and wrote the manuscript. HGL conceived and designed the study. HGL also contributed to the data analysis and revised the manuscript. Both authors read and approved the final manuscript.

Ethics approval

We confirm that the leaves and winter buds collection presented here were conducted in accordance with the wild plant care regulations and natural reserves regulations set forth by the Decree of the state council of the People's Republic of China.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 12 October 2017 Accepted: 4 December 2018

Published online: 19 December 2018

References

1. Wang Z. Utilization and species hybridization in *Liriodendron*. 1rd ed. Beijing: Chinese Forestry Press; 2005. (In Chinese).
2. Williams RS, Feist WC. Durability of yellow-poplar and sweetgum and service life of finishes after long-term exposure. For Prod J. 2004;54:96–101.
3. Bonwook K, Hoyong K, Nahyun P, Soomin L, Hwanmyeong Y. Organosolvent pretreatment of *Liriodendron tulipifera* and simultaneous saccharification

- and fermentation for bioethanol production. *Biomass Bioenergy*. 2011;35:1833–40.
4. Li K, Chen L, Feng Y, Yao J, Li B, Xu M, Li H. High genetic diversity but limited gene flow among remnant and fragmented natural populations of *Liriodendron chinense* Sarg. *Biochem Syst Ecol*. 2014;54:230–6.
 5. Hao R, He S. Geographical distribution of *Liriodendron chinense* in China and its significance. *J Plant Resour Environ*. 1995;4:1–6 (In Chinese).
 6. He S, Hao R. Study on the natural population dynamics and the endangering habitat of *Liriodendron chinense* in China. *Acta Phytocologica Sinica*. 1999;23:87–95 (In Chinese).
 7. Li M, Wang K, Wang X, Yang P. Morphological and proteomic analysis reveal the role of pistil under pollination in *Liriodendron chinense* (Hemsl.) Sarg. *PLoS One*. 2014;9:e99970.
 8. Sun X, Lei T, Yuan S, Lin H. Progress in research of Dehydrins. *J Wuhan Botanical Res*. 2005;23:299–304 (In Chinese).
 9. Sewell MM, Parks CR, Chase MW. Intraspecific chloroplast DNA variation and biogeography of north American *Liriodendron L.* (Magnoliaceae). *Evolution*. 1996;50:1147–54.
 10. Parks CR, Miller NG, Wendel JF, McDougal KM. Genetic divergence within the genus *Liriodendron* (Magnoliaceae). *Ann Mo Bot Gard*. 1983;70:658–66.
 11. Liang HY, Carlson JE, Leebens-Mack JH, Wall PK, Mueller LA, Buzgo M, Landherr LL, Hu Y, DiLoreto DS, Ilut DC, Field D, Tanksley SD, Ma H, dePamphilis CW. An EST database for *Liriodendron tulipifera L.* floral buds: the first EST resource for functional and comparative genomics in *Liriodendron*. *Tree Genetics & Genomes*. 2008;4:419–33.
 12. Pan XY, Ji KS, Fang Y. Changes in Enzyme Activities in Different Clones of *Liriodendron chinense* × *L. tulipifera* under Flooding Stress. *J Northwest Forestry Univ*. 2007;22:43–6 (In Chinese).
 13. Lu C, Li B, Zheng Y. Analysis on differential expression of cold resistance related genes of *Liriodendron chinense* under low temperature stress. *J Plant Resour Environ*. 2015;24:25–31 (In Chinese).
 14. Zhang X, Fang Y, Chen Y. Effect of waterlogging stress on physiological indexes of *Liriodendron* seedlings. *J Plant Resour Environ*. 2006;15:41–4 (In Chinese).
 15. Wachowiak W, Balk PA, Savolainen O. Search for nucleotide diversity patterns of local adaptation in dehydrins and other cold-related candidate genes in Scots pine (*Pinus sylvestris L.*). *Tree Genetics Genomes*. 2009;5:117–32.
 16. Yang Z, Zang T, Bolshoy A, Beharav A, Nevo E. Adaptive microclimatic structural and expressional dehydrin 1 evolution in wild barley, *Hordeum spontaneum*, at 'Evolution Canyon', Mount Carmel, Israel. *Mol Ecol*. 2009;18:2063–75.
 17. Close TJ, Fenton RD, Moonan F. A view of plant dehydrins using antibodies specific to the carboxy terminal peptide. *Plant Mol Biol*. 1993;23:279–86.
 18. Choi DW, Zhu B, Close TJ. The barley (*Hordeum vulgare L.*) dehydrin multigene family: sequences, allele types, chromosome assignments, and expression characteristics of 11 Dhn genes of cv Dicktoo. *Theor Appl Genet*. 1999;98:1234–47.
 19. Sterky F, Bhalerao RR, Unneberg P, Segerman B, Nilsson P, Brunner AM, Charbonnel-Campaa L, Lindvall JJ, Tandre K, Strauss SH, et al. A Populus EST resource for plant functional genomics. *PNAS*. 2004;101:13951–6.
 20. Close TJ, Kortt AA, Chandler PM. A cDNA-based comparison of dehydration-induced proteins (dehydrins) in barley and corn. *Plant Mol Biol*. 1989;13:95–108.
 21. Close TJ. Dehydrins: emergence of a biochemical role of a family of plant dehydration proteins. *Physiol Plant*. 1996;97:795–803.
 22. Zhu W, Zhang L, Lv H, Zhang H, Zhang D, Wang X, Chen J. The dehydrin wzy2 promoter from wheat defines its contribution to stress tolerance. *Funct Integr Genomics*. 2014;14:111–25.
 23. Brini F, Yamamoto A, Jlaiel L, Takeda S, Hobo T. Pleiotropic effects of the wheat dehydrin DHN-5 on stress responses in Arabidopsis. *Plant Cell Physiol*. 2011;52:676–88.
 24. Qin Y, Qin F. Dehydrins from wheat × *Thinopyrum ponticum* amphiploid increase salinity and drought tolerance under their own inducible promoters without growth retardation. *Plant Physiol Biochem*. 2016;99:142–9.
 25. Kosova K, Vitamvas P, Prasit IT. Wheat and barley dehydrins under cold, drought, and salinity - what can LEA-II proteins tell us about plant stress response? *Front Plant Sci*. 2014;5:343.
 26. Li H, Cheng L, Liang C, Huang M. A case study on provenance testing of tulip tree (*Liriodendron* spp). *China Forestry Sci Technol*. 2005;19:13–6 (In Chinese).
 27. Yang Y, Xu M, Luo Q, Wang J, Li H. De novo transcriptome analysis of *Liriodendron chinense* petals and leaves by Illumina sequencing. *Gene*. 2014;534:155–62.
 28. Tamura K, Battistuzzi FU, Billings-Ross P, Murillo O, Filipiński A, Kumar S. Estimating divergence times in large molecular phylogenies. *PNAS*. 2012;109:19333–8.
 29. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. 1997;25:4876–82.
 30. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 2009;25:1451–2.
 31. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989;123:585–95.
 32. Hurst LD. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet*. 2002;18:486.
 33. Fu YX, Li WH. Statistical tests of neutrality of mutations. *Genetics*. 1993;133:693–709.
 34. Tajima F. Evolutionary relationship of DNA sequences in finite populations. *Genetics*. 1983;105:437–60.
 35. Nachman M. Detecting selection at the molecular level. In: Fox CW, Wolf J, editors. *Evolutionary Genetics: Concepts and Case Studies*. Oxford: Oxford University Press; 2006. p.103–18.
 36. Nielsen R, Yang Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics*. 1998;148:929–36.
 37. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 2013;30:2725–9.
 38. Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online*. 2007;1:47–50.
 39. Falush D, Stephens M, Pritchard JK. Inference of population structure: extensions to linked loci and correlated allele frequencies. *Genetics*. 2003;164:1567–87.
 40. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155:945–59.
 41. Earl DA, VonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour*. 2012;4:359–61.
 42. Rosenberg NA. Distruct : a program for the graphical display of population structure. *Mol Ecol Notes*. 2010;4:137–8.
 43. Parks CR, Wendel JF. Molecular Divergence Between Asian and North American Species of *Liriodendron* (Magnoliaceae) with Implications for Interpretation of Fossil Floras[J]. *Am J Bot*. 1990;77:1243–56.
 44. Bandelt HJ, Forster P, Sykes BC, Richards MB. Mitochondrial portraits of human populations using median networks. *Genetics*. 1995;141:743–53.
 45. Polzin T, Daneshmand SV. On Steiner trees and minimum spanning trees in hypergraphs. *Oper Res Lett*. 2003;31:12–20.
 46. Adobe Photoshop CS6. *Adobe Illustrator CS6*. Adobe Press. 2013; 34:125.
 47. Li WH. Distribution of nucleotide differences between two randomly chosen cistrons in a finite population. *Genetics*. 1977;85:331–7.
 48. Nei M. *Molecular evolutionary genetics*. Columbia: Columbia University Press, New York; 1987.
 49. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees[J]. *BMC Evol Biol*. 2007;7(1):214.
 50. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarisation in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*. 2018;67:901–04.
 51. Breathnach R, Chambon P. Organization and expression of eucaryotic split genes coding for proteins. *Annu Rev Biochem*. 1981;50:349–83.
 52. Hundertmark M, Hincha DK. LEA (late embryogenesis abundant) proteins and their encoding genes in Arabidopsis thaliana. *BMC Genomics*. 2008;9:118.
 53. Hinniger C. Isolation and characterization of cDNA encoding three Dehydrins expressed during Coffea canephora (Robusta) grain development. *Ann Bot*. 2006;97:755–65.
 54. Omar SA, Elsheery NI, Kalaji HM, Ebrahim M, Pietkiewicz S, Lee C, Allakhverdiev S, Xu Z. Identification and differential expression of two dehydrin cDNAs during maturation of *Jatropha curcas* seeds. *Biochem Mosc*. 2013;78:485–95.
 55. Shiota H, Yang G, Shen S, Eun C, Watabe K, Tanaka I, Kamada H. Isolation and characterization of six abscisic acid-inducible genes from carrot somatic embryos. *Plant Biotechnol*. 2004;21:309–14.
 56. Li K. Studies on population genetics and molecular phylogeography of *Liriodendron*. Nanjing: Nanjing Forestry University; 2013. (In Chinese)

57. Luo G, Shi J, Yin T, Huang M, Wang M. Comparison of genetic diversity between *Liriodendron tulipifera* Linn. And *Liriodendron chinense* (Hemsl.) Sarg. By means of RAPD markers. *J Plant Resour Environ*. 2000;9:9–13 (In Chinese).
58. Riahi L, Zoghalmi N, Dereeper A, Laucou V, Mliki A, This P. Molecular characterization and evolutionary pattern of the 9- cis -epoxycarotenoid dioxygenase NCED1 gene in grapevine. *Mol Breed*. 2013;32:253–66.
59. Cheng CA, O'Brien MJ, KKS N, Lee P, Hector A, Schmid B, Shimizu KK. Genetic diversity of two tropical tree species of the Dipterocarpaceae following logging and restoration in Borneo: high genetic diversity in plots with high species diversity. *Plant Ecol Divers*. 2017;9:1–11.
60. Templeton AR. Biodiversity at the molecular genetic level: experiences from disparate macroorganisms. *Philos Trans R Soc Lond*. 1994;345:59–64.
61. Upendra JM, Rao SR, Dagla HR. Genetic diversity analysis of *Salvadora persica* : an evergreen halo-xeric species of semi-arid and sub-humid regions of Rajasthan. *India Ecol Genet Genomics*. 2017;2:35–41.
62. Skibinski DO, Ward RD. Average allozyme heterozygosity in vertebrates correlates with Ka/Ks measured in the human-mouse lineage. *Mol Biol Evol*. 2004;21:1753–9.
63. Rogers AR, Harpending H. Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol*. 1992;9:552–69.
64. Kryazhimskiy S, Plotkin JB. The population genetics of dN/dS. *PLoS Genet*. 2008;4:e1000304.
65. Kimura M. The neutral theory of molecular evolution. *Sci Am*. 1979;241:98–100 102, 108.
66. Wright SI, Gaut BS. Molecular population genetics and the search for adaptive evolution in plants. *Mol Biol Evolution*. 2005;22:506–19.
67. Hamilton W. Cretaceous and Cenozoic history of the northern continents. *Ann Mo Bot Gard*. 1983;70:440–58.
68. Parks CR, Wendel JF, Sewell MM, Qiu YL. The Significance of Allozyme Variation and Introgression in the *Liriodendron tulipifera* complex (Magnoliaceae). *Am J Bot*. 1994;81: 878–89.
69. Yang A, Dick CW, Yao X, Huang H. Impacts of biogeographic history and marginal population genetics on species range limits: a case study of *Liriodendron chinense*. *Sci Rep*. 2016;6:25632.
70. Li L, Abbott RJ, Liu B, Sun Y, Li L, Zou J, Wang X, Miehle G, Liu J. Pliocene intraspecific divergence and Plio-Pleistocene range expansions within *Picea likiangensis* (Lijiang spruce), a dominant forest tree of the Qinghai-Tibet plateau. *Mol Ecol*. 2013;22:5237–55.
71. Collevatti RG, Limaribeiro MS, Terribile LC, Guedes LBS, Rosa FF, Telles MPC. Recovering species demographic history from multi-model inference: the case of a neotropical savanna tree species. *BMC Evol Biol*. 2014;14:1–13.
72. Lamb HH. *Climatic history and the future*. Princeton: Princeton University; 1977.
73. Fedorov A, Brierley C, Lawrence K, Liu Z, Dekens P, Ravelo A. Patterns and mechanisms of early Pliocene warmth. *Nature*. 2013;496:43.
74. Dowsett H, Thompson R, Barron J, Cronin T, Fleming F, Ishman S, Richard P, Debra W, Thomas HJ. Joint investigations of the middle pliocene climate I: prism paleoenvironmental reconstructions. *Global Planetary Change*. 1994; 9(9):169–95.
75. Atkinson Q, Gray R, Drummond A. Bayesian coalescent inference of major human mitochondrial DNA haplogroup expansions in Africa. *Proc Roy Soc Lon B*. 2009;276:367–73.
76. Saibi W, Drira M, Yacoubi I, Feki K, Brini FA. Empiric, structural and in silico findings give birth to plausible explanations for the multifunctionality of the wheat dehydrin (DHN-5). *Acta Physiol Plant*. 2015;37:1–8.
77. Wu X. *Basal biochemistry*. 3rd ed. Beijing: China Agriculture Press; 1997. (In Chinese).

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

