# Phylogenomics and biogeography of Wisteria (Fabaceae), with implication on plastome evolution of inverted repeat-lacking clade

**Mao-Qin Xia**
Zhejiang University

**Ren-Yu Liao**
Zhejiang University

**Jin-Ting Zhou**
Zhejiang University

**Han-Yang Lin**
Zhejiang University

**Jian-Hua Li**
Hope College

**Pan Li** ( ✉ panli_zju@126.com )
Zhejiang University    https://orcid.org/0000-0002-9407-7740

**Cheng-Xin Fu**
Zhejiang University

**Ying-Xiong Qiu**
Zhejiang University

---

---

# Abstract

Background: The inverted repeat-lacking clade (IRLC) of Fabaceae is characterized by loss of an IR region in plastomes. Both the loss of an IR region and the life history may have affected the evolution of the plastomes in the clade. Nevertheless, few studies have been done to test the impact explicitly. Wisteria , an important member of IRLC and has a disjunct distribution between eastern Asia and eastern North America, has confused interspecific relationships and biogeography, which need to elucidate in depth. Results: The plastome of six newly sequenced Wisteria species and a Millettia japonica ranged from 130,116 to 132,547 bp. Phylogenetic analyses recognized two major clades in IRLC: Glycyrrhiza - Millettia - Wisteria clade and a clade containing the remaining genera. North American Wisteria species and Asian species formed reciprocal clades. Within Asian clade, each of the two Japanese species was sister to a species in the Asian continent. A ~16kb inversion from ndh F to ycf 1 in all IRLC species. Wisteria and Millettia japonica have two intron of rps 12 gene but all other IRLC species just have one. Synonymous substitution rates ( d S ) of protein coding genes were higher in the IRLC species than non-IRLC species. Woody species have lower substitutions rates than herbs. Wisteria may have originated in East Asia by the boundary of Oligocene and Miocene and the eastern Asian-eastern North American disjunction formed in the Late Miocene, while two vicariance events formed the disjunct distributions between the Asian continent and the Japanese islands in the Quaternary. Conclusions: In the IRLC clade, Wisteria , Milletia japonica and Glycyrrhiza form a clade to the remaining genera, most of which are herbaceous. Both the loss of one IR region and the herbaceous habit elevated mutation rates of the plastomes. Multiple vicariance events between eastern Asia and eastern North America, and between the Asian continent and the Japanese Islands may have promoted speciation of Wisteria since the Late Miocene. Plastomes contain rich genetic diversity for studying genetic structure and migration of populations in response to climatic changes, which benefits conservation of rare and endangered species.

# Background

The plastomes of most vascular plants usually have a conservative quartile structure with a large single-copy (LSC) region and a small single-copy (SSC) region separated by two inverted repeat (IR) regions [1-5]. They are about 150 kb in size and contain around 114 unique genes, including four rRNA genes, 30 tRNA genes, and 80 protein genes [6]. Comparative analyses of plastomes have identified some unique features, such as large segment inversion [7], high frequency of repeats [8], gene gain and loss events (including pseudogenization) [9], and expansion or contraction of IRs [10]. The IR region varies greatly among plant lineages, which may have affected the stability of chloroplast structure resulting in an elevated variation of single copy regions [11, 12]. One of the most notable changes of the IR region is the loss of a repeat in a lineage of Fabaceae, namely the inverted-repeat-lacking clade (IRLC), which has been supported by various molecular systematic studies [15-21].

The IRLC clade is composed of seven tribes, (Carmichaelieae, Cicereae, Galegeae, Hedysareae, Trifolieae, Vicieae and Wisterieae), about 37 genera and 2000 species [13-19]. Some plastomes in IRLC have recently been sequenced and analyzed [15, 20-23] revealing important evolutionary patterns: the loss of rps16 gene, accD gene and one intron of clpP [24], multiple sequence inversions [21, 25, 26], gene transfers to the nucleus [15, 27, 28], and elevated mutation rates in the IRLC clade [12]. Most of the genera in the IRLC clade are herbaceous, while a few are woody shrubs or lianas including Caragana, Carmichaelia R. Br., and Wisteria Nutt.. Such life history characteristics have been recognized as an important factor affecting mutation rates with significantly lower rates in woody plants than in herbaceous ones [29, 30]. Therefore, the IRLC clade is a good model system for understanding the impact of both the loss of one IR region and the life history traits on the evolution of plastomes.

Wisteria, one of the most romanticized and spectacular garden plants, has been widely cultivated around the world. Species of Wisteria are woody, deciduous lianas, and bear odd pinnate compound leaves, pendulous racemes, typical papilionaceous corolla in purple, violet or white [31-33]. Wisteria composed of 6-8 species with a disjunct distribution in Eastern Asian and Eastern North American temperate deciduous forest [16, 21, 34-38]. Wisteria frutescens (L.) Poir. occurs in eastern USA; W. sinensis (Sims) Sweet, W. villosa Rehder, W. venusta Rehder & E. H. Wilson., and W. brevidentata Rehder (sometimes considered to be a variant of W. sinensis) are native to China; whereas W. floribunda (Willd.) DC. and W. brachybotrys Siebold & Zucc. are endemic to Japan [39]. Millettia japonica, a close relative of Wisteria, has been recently treated as Wisteriopsis japonica (Siebold & Zucc.) J. Compton & Schrire [19, 40]. Previous studies support the reciprocal monophyly of eastern Asian and eastern North American species, but the interspecific relationships of Asian species have not been resolved [37], largely due to

the small amount of data. Plastomes have been widely used in phylogenetic studies of closely related species [1, 4, 9, 41]. Within *Wisteria*, plastomes of *W. sinensis* and *W. floribunda* have been reported before [42]. Thus, we obtained plastomes from other species of *Wisteria* for generating a robust phylogeny of the genus.

The objectives of the study were 1) exploring the impact of the loss of an IR and life history traits on the evolution of plastomes in the IRLC clade, 2) resolving phylogenetic relationships of the IRLC clade and *Wisteria*, and 3) examining the implications of the phylogeny for the systematics and biogeography of *Wisteria*.

# Results

## Size, content, and arrangement of plastomes

After the low-quality reads and adaptor sequences were filtered out, 20,453,560−28,744,390 clean reads were obtained for six *Wisteria* and *Millettia japonica*. The NOVOPlasty assembly produced a long contig representing the whole plastome. These seven newly sequenced plastomes ranged from 130,116 to 132,547 bp in length, and the GC contents were from 34.20% to 34.50% (Fig. S1; Table 1). All plastomes contained 110 unique genes arranged in same order, with 76 protein-coding genes, 30 tRNA genes, and four rRNA genes (Table 1), and there was a single inverted repeat. In all seven sequences, ten protein-coding genes (*atp*F, *clp*P, *ndh*A, *ndh*B, *pet*B, *pet*D, *rps*12, *rpl*2, *rpl*16, *rpo*C1) and six tRNA genes (*trn*A-UGC, *trn*G-UCC, *trn*I-GAU, *trn*K-UUU, *trn*L-UAA, *trn*V-UAC) possessed a single intron, whereas two genes (*rps*12, and *ycf*3) contained two introns (Table S3). All seven sequenced species had one intron in *clp*P gene, which was not absent in *Glycyrrhiza glabra*. For *rps*12 gene, *Wisteria* had 2 introns, whereas *G. glabra* just had one. All seven plastome sequences have been submitted to the GenBank (accession numbers: MN311163, MN311167-MN311171; Table 1).

Gene rearrangements were not found in *Glycyrrhiza glabra*, *Millettia japonica*, *Wisteria sinensis*, *Caragana kozlowii* Kom., *Carmichaelia australis* R. Br., *Astragalus mongholicus* Bunge, *Cicer arietinum* L., or *Medicago truncatula* Gaertn.. On the contrary, large segments of the plastomes were either reversed and lost in *Trifolium subterraneum* L., *Pisum sativum* L., *Lens culinaris* Medik., *Lathyrus sativus* L. and *Vicia faba* L., (Fig. 1). In addition, plastome structure did not change among species of *Wisteria* (Fig. S2). The MISA analysis identified a total of 540 SSRs across the six *Wisteria* plastomes. The number of SSRs ranged from 81 (*W. villosa*) to 99 (*W. floribunda*). Mono-nucleotide SSRs (A/T/G) took the largest percentage, followed by di-nucleotide SSRs (AG/AT/TA/TC) (Fig. 2A). The combined total of tri-, tetra-, penta- and hexa-nucleotide SSRs was no more than 15%. Besides, only *W. floribunda* and *W. venusta* contained hexa-nucleotide SSRs, and *W. frutescens* did not have penta-nucleotide SSRs (Table S4).

The forward, palindromic and reverse repeats were detected in *W. floribunda*, *W. venusta* and *W. villosa*, while only forward and palindromic repeats were found in *W. brachybotrys*, *W. frutescens* and *W. sinensis* (Fig. 2B, Table S4). The total number of repeats ranged from 43 (*W. villosa*) to 70 (*W. venusta*) and most repeats were in no-coding regions (82.09%-87.14%, except 65.85% for *W. villosa*). Among the six plastomes of *Wisteria*, the proportion of forward repetition was the highest, ranging from 31 in *W. villosa* to 65 found in *W. venusta* (Fig. 2C, Table S5).

In *Wisteria*, nucleotide diversities (*Pi*) of all genes, intergenic spacers, and introns ranged from 0.0008 (*ndh*D) to 0.04648 (*trn*N-*ycf*1) (Table S6). There were seven hypervariable regions with *Pi* greater than 0.025 including *trn*N-*ycf*1, *rpl*33-*rps*18, *trn*S-*trn*G, *ndh*F-*trn*H, *rps*18, *clp*P-*psb*B, and *rps*8-*rpl*14 (Fig. 3).

## Phylogenetic relationships

The matrix of 28 plastome CDS sequences was 71,545 bp in length. In both the ML and BI trees *Glycyrrhiza glabra*, *Millettia japonica* and *Wisteria* formed a clade (BS = 90, PP = 0.5), which was sister to the remaining genera. *M. japonica* was the closest taxon to *Wisteria* (BS = 1, PP = 1). *Caragana* was sister to (*Astragalus* + *Carmichaelia*), they then were sister to the clade containing *Cicer* L., *Medicago*, *Trifolium*, *Pisum* L., *Lathyrus*, *Lens* Mill. and *Vicia* (BS = 1, PP = 1).

The interspecific relationships of *Wisteria* were well resolved (BS = 100, PP = 1 for all nodes, Fig. 4). *Wisteria frutescens* was sister to the clade of Asian species (BS = 1, PP = 1). There were two subclades within the Asian clade (BS = 1, PP = 1): *W. brachybotrys* and *W. venusta*, and the other three species. Two individuals of *W. floribunda* formed a clade, which was sister of *W. sinensis and W. villosa* (BS = 1, PP = 1). However, relationship between *W. villosa* and *W. sinensis* was still unresolved.

## Substitution rates

The mean $d_S$ of SC genes was 0.022 ± 0.001 in the IRLC pairs and 0.009 ± 0.000 in the non-IRLC pairs. For IR genes, they were 0.014 ± 0.003 and 0.002 ± 0.000 respectively. All herbaceous pairs had significant higher $d_S$ than the woody pairs ($P < 0.01$). The herbaceous pairs in the IRLC had 2.07-fold $d_S$ in SC region than woody pairs, whose average $d_S$ of the SC genes was 0.010 ± 0.000 (Table S2).

## Divergence time estimation

The estimated divergence time of IRLC was 40.11 (35.73−44.36) Ma (node c, Fig. 5, Table 2). The stem age and crown age of *Wisteria* were 23.85 (23.26−24.86) Ma and 8.66 (2.40−17.38) Ma, respectively (node d, node 2). Then, the clade of *W. brachybotrys* and *W. venusta* was separated from other Asian wisteria at 2.22 (0.75−5.67) Ma (node 3), and the split of these two species at 0.06 (0.00−0.23) Ma (node 6). The crown age of *W. floribunda*, *W. sinensis* and *W. villosa* was estimated to be 0.87 (0.25−2.15) Ma (node 4). The crown ages of other genera in IRLC were listed in Table 2.

## Ancestral area reconstruction

The result of BMM analysis indicated that the most recent common ancestor (MRCA) of *Wisteria* was distributed in China with 33.57% probability, in Japan with 33.22% probability, meanwhile, in eastern North America with 12.22% probability (Fig. 5). There may be a dispersal event from China to eastern North America, then followed by a vicariance event, with the probability of 12.78%. For Asian wisterias, the ancestral range was indistinct, with 37.09% chance in China, a 31.18% chance in Japan and 31.43% probability in China and Japan. Vicariance events were also identified between *W. brachybotrys* and *W. venusta* with 37.83% probability, between *W. floribunda* and *W. sinensis*-*W. villosa* with 34.26% probability.

# Discussion

## Phylogenetic relationships among IRLC and within *Wisteria*

Previous studies have employed some plastid and nuclear sequences to determine the phylogeny of legume family, however, the phylogenetic relationships among IRLC taxa have not been resolved satisfactorily [16, 34-36]. In this study, The IRLC contained two strongly supported subclades (BS = 100, PP = 1; Fig. 4). *Glycyrrhiza* was the basal group of the IRLC in the *mat*K gene tree [16], or formed a clade with Vicioid species (such as *Cicer*, *Medicago*, *Trifolium* and *Pisum*) in the nuclear ribosomal ITS/5.8S trees [35]. Our result is consistent with Wang et al. in that *Wisteria*, *M. japonica* and *G. glabra* form a clade (BS = 90, PP = 0.5; Fig. 4), which is sister to the remaining genera [42, 62]. *Millettia japonica* has a similar phenotype with *Wisteria*, and was initially described as *Wisteria japonica* Siebold & Zucc. Recently, it was recognized as a separate genus *Wisteriopsis* J. Compton & Schrire [19]. In addition, *Caragana* and (*Astragalus* + *Carmichaelia*) form a clade, which is sister to the clade of the remaining seven genera (Fig. 4). This is consistent with previous research based on *mat*K gene [38].

Our plastome sequence data provide strong support for the monophyly of *Wisteria* and for the sister relationship of the American *W. frutescens* and the Asian species (Fig. 4). Furthermore, The Japanese *W. floribunda* and *W. brachybotrys* are not sister species, instead they form sister relationships with *W. sinensis* and *W. venusta*, respectively. *Wisteria brachybotrys* and *W. venusta* have been considered as conspecific by Valder [39], or different species because of different ovule numbers and floral structures [31, 40]. However, the divergence time estimation show that they diverged very recently (0.06 Ma). Additional sampling of multiple populations of the two species is needed to test the species boundary. In previous research, the phylogenetic relationships among *W. sinensis*, *W. villosa* and *W. floribunda* were not resolved [37], most likely due to the small amount of data. Here the two *W. floribunda* from Japan and Korea form a clade sister to the unresolved clade of *W. sinensis*

and *W. villosa*. Leaf pubescence, pedicel length, and flower size of *W. sinensis* and *W. villosa* are highly variable and even overlap, thus it is unreliable to distinguish them using the so-called diagnostic characteristics [39, 63]. This suggests that future studies on population genetics should be used to test whether they represent two separately evolving lineages or species.

## Plastome evolution in the IRLC clade

Plastomes have experienced various evolutionary changes in plants, one of which is the loss of one copy of the IR in the IRLC clade of Fabaceae, which may disrupt the gene stability and increase the rates of substitution in single copy regions [8, 12, 15, 18, 64-66]. The plastomes of six species of *Wisteria* and *Millettia japonica* have a single IR region (Fig. S1), and share similar size (130,116 to 132,547 bp), overall structure, gene order and content (Table 1). In addition, compared to the non-IRLC lineages of Fabaceae, the IRLC plastomes contain the traditional LSC and IR region plus a reversed SSC (a ~16kb inversion from *ndh*F to *ycf1*). However, the loss of one IR region may not be the major driving force for the genomic changes in the IRLC [15], because gene orders from *Glycyrrhiza glabra* to *Medicago truncatula* are consistent and have not experienced much rearrangement compared with non-IRLC members (Fig. 1). The gene order rearrangements appear to be associated with the herbaceous life history since the woody genera in the IRLC clade did not show the rearrangements (Fig. 1). Short generation time with great opportunities for recombination may have led to the structural changes in the herbaceous genera [29]. The localized hypermutation in *Vicia*, *Lens*, *Lathyrus*, *Pisum* and *Trifolium* (Fig. 1) may be due to the existence of a large number of repeat sequences [15, 28]. *Lens*, *Trifoium* and *Vicia* have 9% repetitive DNA on average, however, *G. glabra* has 4.1%, *Wisteria* has around 3% and *Lotus japonica* just has 2.4% [15].

We estimated the synonymous rates of SC for the IRLC species pairs and relative taxa in Fabaceae and the results show that the difference was about 7 times for IR regions, whereas 2.4 times for SC regions. And herbaceous pairs have significant higher $d$s than wood pairs. The loss of an IR region may have disrupted the stability of the plastome leading to the higher substitution rates in IR genes [1, 12]. The impact may be further enhanced by life history in the herbaceous genera.

Structural rearrangement and gene/intron loss (including pseudogenization) may be correlated with increases in the rate of molecular evolution [30]. In the IRLC, the *rps*16-*acc*D-*psa*I-*ycf*4 region shows a higher rate of transfers, substitutions, or losses of genes [15]. For example, *ycf4* gene shows an acceleration in Millettioids, Robinioids, and the IRLC than in other legumes, especially 20-fold in *Lathyrus*. Plastome gene losses are rare in photosynthetic species, the lost gene may have been transferred to the nuclear genome or substituted by a nuclear gene [28]. The loss of *rps16* in all IRLC species may be compensated by dual targeting of mitochondrial ribosomal protein S16, which is encoded by a nuclear gene [28]. The *accD* gene functioning in plant development and possessing recombinationally active repeats has been transferred to the nucleus in two *Trifolium* species [20, 25, 67, 68]. *Pisum* and *Lathyrus* have lost *ycf4* and *psaI* gene, respectively, which can attribute directly to hypermutation of *ycf4* and its neighboring genes [28]. Some intron losses may also have occurred with the origin of the IRLC clade [8]. For example, most angiosperms have two introns in the *rps12* gene, whereas the IRLC species (except *Wisteria* and *Millettia japonica*) just have one, which indicates that the loss of this intron occurred subsequent to the loss of one copy of the IR [21]. There is a loss of intron 1 of *clpP* in all IRLC species, while the intron 2 of *clpP* is lost only in *Glycyrrhiza glabra* (Fig. 4).

## Biogeography of *Wisteria*

Our results show that *Wisteria* has originated at 23.85 (23.26−24.86) Ma (node d, Fig. 5), which is much earlier than a previous estimate of 13.4 (9.7−17.5) Ma [37]. Lacking samples of *Callerya* may result in the older estimated age of *Wisteria*. Wang et al. believed that *Wisteria* may have a more northern distribution in Neogene and the diversification in this genus had taken placed in the middle Miocene [69]. We inferred the ancestral area of *Wisteria* as Eastern Asia with a high probability (78.10%). However, it is equally likely that ancestral area of *Wisteria* may be in China (33.57%) or Japan (33.22%). A subsequent dispersal to eastern North America and a vicariance event has produced the disjunct distribution at 8.66 (2.40−17.38) Ma (Fig. 5). The result is consistent with Li et al., and supports the "out of Asia" hypothesis for the majority of eastern Asian and eastern North American disjunct plant genera [37, 70]. Fossils of *Wisteria* have been found in Asia and Europe indicating a wide distribution until middle Miocene [69]. The Bering land bridge is likely the migratory pathway for *Wisteria* because it accommodated many

temperate components in the middle Tertiary and later times, whereas North Atlantic land bridge disconnected at that time [71-73].

Eastern Asia is the center of species diversity for *Wisteria*. The diversification time of Asian wisterias might be 2.22 (0.75-5.67) Ma in Pleistocene. Climatic oscillations of the Quaternary at around 2 million years ago have deeply affected plant distribution and genetic structure [74, 75]. Asian wisteria is native in 'Sino-Japanese Floristic Region', which has never been directly impacted by extensive ice sheets [76-78]. But drastic climate and environmental changes, such as sea-level fluctuations, promoted fragmentation of species ranges, population isolation and variation, thereby providing opportunities for allopatric speciation through selection or genetic drift [79]. Albeit with a limited sampling of population of the Asian species of *Wisteria*, the sister relationships of the Japanese and Asian continental species support the significant role the allopatry has played in the generation of biological diversity in Asia [80].

# Conclusions

In this study, we compared the plastome variation inside and outside the IRLC clade at both the generic and species levels. Our results suggest that several genomic and substitution rate changes might happened after the loss of an inverted repeat disrupting the genomic stability and may be further affected by life history traits of the taxa. The plastid data provide a satisfactory resolution for relationships of the IRLC clade and support the sister relationship of eastern Asian and eastern North American species. A dispersal of Asian populations to North America and a subsequent vicariance event might have formed the disjunction in the late Miocene. Allopatric speciation of *Wisteria* between the Japanese Islands and the Asian continent in the Quaternary increased the species richness of eastern Asian in comparison with eastern North America. Abundant molecular markers can be developed based on plastome sequence data and useful for future studies on species identification, molecular phylogenetics and population genetics of *Wisteria*.

# Methods

### Plant Samples, DNA extraction and Sequencing

Twenty-eight species were included in the study representing all tribes and 13 genera of the IRLC clade and outgroups *Robinia* L. and *Lotus* L.. Within *Wisteria*, six species from China, Japan and USA were included. Collection information was provided in Table 1. Pan Li did the final identification of the samples and representative voucher specimens are currently deposited at the Herbarium of Zhejiang University (HZU) and Herbarium Institute of Botany, CAS (PE). This study did not involve endangered species. No specific collecting permits required for the collection of plant materials. Research on these plant comply with institutional, national, or international guidelines.

Total genomic DNAs were extracted from 3 mg of dried leaf samples using DNA Plantzol Reagent (Invitrogen) according to the manufacturer's instructions. The quality of the DNA products was assessed using 1.2% agarose gel electrophoresis. Short-insert (800 bp) paired-end libraries were generated by using a Genomic DNA Sample Prep Kit (Illumina) according to the manufacturer's protocol. Genomic DNA from each species was indexed by tags and pooled together in one lane of a HiSeq[TM] X-Ten (Illumina, San Diego, California, USA) for sequencing at Beijing Genomics Institute (BGI, Shenzhen, China).

### Genome Assembly and Annotation

All seven plastome sequences were assembled in NOVOPlasty 2.6.3 with the *mat*K sequence of *Wisteria sinensis* (AF142732) as seed and the complete plastome sequence of *W. sinensis* (KT200359) as reference [43]. Whole sequences were assembled using MAFFT v7 in Geneious 10.2.3 (http://www.geneious.com) [44], then annotated with *Glycyrrhiza glabra* L. (KF201590) as reference, followed by verifying the start and stop codons manually. In addition, tRNAscan-SE was used to confirm the tRNA genes with default parameters [45]. Finally, the circular maps of the plastomes were drawn using the OrganellarGenome DRAW (ORDRAW) [46], followed by manual editing.

### Plastome comparison and hotspot regions identification

In order to examine plastome structure rearrangements in the IRLC clade, whole-genome alignment of 13 IRLC genera (Table S1) was performed in Mauve alignment in Geneious 10.2.3 with *Lotus japonicus* (Regel) K. Larsen (NC_002694) as the reference [47]. Shuffle-LAGAN mode on mVISTA (http://genome.lbl.gov/vista/mvista/submit.shtml) was used to compare six *Wisteria* species [48].

To determine the position and size of repeat sequences, REPUTER [49] was used to find the whole direct (forward), inverted (palindromic), complement and reverse repeats in six *Wisteria* plastomes. The minimal repeat size was set to 30 bp and the sequence identity was > 90%. The SSRs were detected using MISA [50] with thresholds of 10, 5, 4, 3, 3, and 3 for mono-, di-, tri-, tetra-, penta-, and hexa-nucleotides, respectively. Nucleotide variability (*Pi*) of the regions was evaluated in DNASP v5.10 [51].

## Phylogenetic analyses

Akaike Information Criterion (AIC) in jModelTest v2.1.4 was used to determine the optimal nucleotide substitution model for the protein coding genes [52]. Maximum likelihood (ML) analysis was performed in RAxML-HPC v8.2.8 on CIPRES Science Gateway website with 1000 bootstrap replicates [53, 54]. Bayesian analysis was also constructed using MrBayes as implemented on XSEDE 3.2.6 with two independent Markov Chain Monte Carlo chains for 10,000,000 generations and sampling every 1000 generations [55]. The first 25% of calculated trees were discard as burn-in and the remaining trees were used to construct a consensus tree to estimate the posterior probability (PP). Tracer 1.7 was used to make sure that the likelihood scores have reached the plateau [56].

## Estimation of nucleotide substitution rates

In order to investigate the difference of mutation in and outside of the IRLC clade, 10 pairs of plastomes from the same genus (*Vicia* L., *Lathyrus* L., *Trifolium* L., *Astragalus* L., *Medicago* L., *Caragana* Fabr., *Wisteria*, *Glycine* Willd., *Vigna* Savi, *Dalbergia* L. f.) of Fabaceae were chosen to estimate the substitution rates (Table S1). Synonymous substitution rates ($d_S$) were estimated for each CDS via MEGA 7.0 [57]. We also compared the differences between herbs and woody species pairs. Since the dataset did not conform to the normal distribution, a wilcoxon.test in R package was used to test the significance.

## Estimation of divergence time

Divergence times were estimated in BEAST v1.10.4 with the optimal model selected by jModelTest v2.1.4 [58]. We used three secondary calibration points from previous study and one *Wisteria* fossil record to calibrate four nodes: (a) 56 ±0.9 Ma of the Hologalegina crown clade; (b) 48.3±1 Ma for *Lotus* and *Robinia*; (c) 39 ±1 Ma for the crown age of IRLC clade from and (d) the early Miocene for the *Wisteria* crown age (Offset = 23, M = 1, S = 0.5) based on fossil leaflets of *Wisteria fallax* (Nathorst) Tanai et Onoe found in in Abkhazia in the Caucasus (Republic of Georgia) [59, 60]. The monophyly of each group was constrained by the topology of the ML phylogenetic tree (see before). An uncorrelated lognormal relaxed clock and a Yule speciation prior were chosen. Then, MCMC was run for 300 million generations, with sampling every 1000 generations. The log file was analyzed in Tracer v1.7.1 to check convergence; ESS values for all parameters were at least 200 [56]. Finally, the maximum clade credibility tree was calculated using TreeAnnotator 2.4.4 with 1% trees as burn-in. FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/) was used to visualize tree with node ages.

## Reconstruction of ancestral regions

To infer the ancestral area of *Wisteria*, Bayesian Binary MCMC (BBM) analysis was performed in RASP v3.2 [61]. A ML tree was used as the input tree file. Three areas of endemism were defined following the distribution of each species: (A) China; (B) Japan; and (C) eastern North America. Maximum areas were set to 3. Because close taxa of *Wisteria* (such as *Millettia japonica*, *Callerya* Endl., *Afgekia* Craib) were not found in North America, we set the distribution of outgroup in China and Japan. The among-site rate variation was GAMMA + G, the run was done for 10000000 generations, and sampled every 100 generations.

# Abbreviations

LSC: Large single-copy region; SSC: Small-copy region; IR: Invert repeat; ITS: Internal transcribed spacer of ribosomal DNA; IRLC: Inverted repeat-lacking clade; SSR: simple sequence repeats; ML: Maximum Likelihood; BI: Bayesian Inference; BS: Bootstap; PP: posterior probability; MCMC: Markov Chain Monte Carlo; ESS: Effective Sample Size; MRCA: Most recent common ancestor

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Availability of data and materials

The complete chloroplast genomes of *Wisteria brachybotrys*, *W. floribunda*, *W. frutescens*, *W. sinensis*, *W. villosa*, *W. venusta* and *Millettia japonica* were submitted to NCBI database with GenBank accession numbers: MN311163, MN311167-MN311171.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

MQX, PL, JTZ and HYL conducted the field sampling. MQX, RYL analyzed the data. MQX wrote the manuscript. PL, JHL, CXF and YXQ designed the study. PL and JHL revised the manuscript. All authors approved the final manuscript.

## References

1. Zhu A, Guo W, Gupta S, Fan W, Mower JP. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. New Phytol. 2016;209(4):1747-1756.
2. Wicke S, Schneeweiss GM, dePamphilis CW, Muller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 2011;76(3-5):273-297.
3. Shetty SM, Shah MUM, Makale K, Mohd-Yusuf Y, Khalid N, Othman RY. Complete chloroplast genome sequence of *Musa balbisiana* corroborates structural heterogeneity of inverted repeats in wild progenitors of cultivated bananas and plantains. Plant Genome-US. 2016;9(2).
4. Li ZH, Ma X, Wang DY, Li YX, Wang CW, Jin XH. Evolution of plastid genomes of *Holcoglossum* (Orchidaceae) with recent radiation. BMC Evol Biol. 2019;19.
5. Sugiura M. The Chloroplast Genome. Plant Mol Biol. 1992;19(1):149-168.
6. Kim KJ, Lee HL. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. DNA Res. 2004;11(4):247-261.

7. Sun YX, Moore MJ, Lin N, Adelalu KF, Meng AP, Jian SG, Yang LS, Li JQ, Wang HC. Complete plastome sequencing of both living species of Circaeasteraceae (Ranunculales) reveals unusual rearrangements and the loss of the *ndh* gene family. BMC Genomics. 2017;18.

8. Guisinger MM, Kuehl JV, Boore JL, Jansen RK. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Mol Biol Evol. 2011;28(4):1543-1543.

9. Ye WQ, Yap ZY, Li P, Comes HP, Qiu YX. Plastome organization, genome-based phylogeny and evolution of plastid genes in Podophylloideae (Berberidaceae). Mol Phylogenet Evol. 2018;127:978-987.

10. Sun YX, Moore MJ, Zhang SJ, Soltis PS, Soltis DE, Zhao TT, Meng AP, Li XD, Li JQ, Wang HC. Phylogenomic and structural analyses of 18 complete plastomes across nearly all families of early-diverging eudicots, including an angiosperm-wide analysis of IR gene content evolution. Mol Phylogenet and Evol. 2016;96:93-101.

11. Sudianto E, Wu CS, Leonhard L, Martin WF, Chaw SM. Enlarged and highly repetitive plastome of Lagarostrobos and plastid phylogenomics of Podocarpaceae. Mol Phylogenet and Evol. 2019;133:24-32.

12. Perry AS, Wolfe KH. Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. J Mol Evol. 2002;55(5):501-508.

13. Kolodner R, Tewari KK. Inverted repeats in chloroplast DNA from higher-plants. P Natl Acad Sci USA. 1979;76(1):41-45.

14. Zhang YZ, Ma J, Yang BX, Li RY, Zhu W, Sun LL, Tian JK, Zhang L. The complete chloroplast genome sequence of *Taxus chinensis* var. *mairei* (Taxaceae): loss of an inverted repeat region and comparative analysis with related species. Gene. 2014;540(2):201-209.

15. Sabir J, Schwarz E, Ellison N, Zhang J, Baeshen NA, Mutwakil M, Jansen R, Ruhlman T. Evolutionary and biotechnology implications of plastid genome variation in the inverted-repeat-lacking clade of legumes. Plant Biotechnology Journal. 2014;12(6):743-754.

16. Wojciechowski MF, Lavin M, Sanderson MJ: A phylogeny of legumes (Leguminosae) based on analyses of the plastid *mat*K gene resolves many well-supported subclades within the family. AM J BOT. 2004;91(11):1846-1862.

17. Palmer JD, Thompson WF. Rearrangements in the chloroplast genomes of mung bean and pea. P Natl Acad Sci-Biol. 1981;78(9):5533-5537.

18. Lavin M, Doyle JJ, Palmer JD. Evolutionary significance of the loss of the chloroplast-DNA inverted repeat in the Leguminosae subfamily Papilionoideae. Evolution. 1990;44(2):390-402.

19. James A. Compton BDS, Kálmán Könyves, Félix Forest, Panagiota Malakasi, Sawai Mattapha, Yotsawate Sirichamorn. The *Callerya* group redefined and tribe Wisterieae (Fabaceae) emended based on morphology and data from nuclear and chloroplast DNA sequences. PhytoKeys. 2019;125:1-112.

20. Sveinsson S, Cronk Q. Evolutionary origin of highly repetitive plastid genomes within the clover genus (*Trifolium*). Bmc Evol Biol. 2014;14.

21. Jansen RK, Wojciechowski MF, Sanniyasi E, Lee SB, Daniell H. Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of *rps*12 and *clp*P intron losses among legumes (Leguminosae). Mol Phylogenet Evol. 2008;48(3):1204-1217.

22. Samigullin TH, Logacheva MD, Degtjareva GV, Terentieva EI, Sokoloff DD, Vallejo-Roman CM. The complete plastid genome of *Carmichaelia australis* R. Br. (Leguminosae: Papilionoideae). Conserv Genet Resour. 2018;10(3):519-521.

23. Li CY, Zhao YL, Huang HM, Ding Y, Hu YH, Xu ZG. The complete chloroplast genome of an inverted-repeat-lacking species, *Vicia sepium*, and its phylogeny. Mitochondrial DNA B. 2018;3(1):137-138.

24. Doyle JJ, Doyle JL, Palmer JD. Multiple independent losses of 2 genes and one intron from legume chloroplast genomes. Syst Bot 1995;20(3):272-294.

25. Cai ZQ, Guisinger M, Kim HG, Ruck E, Blazier JC, McMurtry V, Kuehl JV, Boore J, Jansen RK. Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) Is associated with numerous repeated sequences and novel DNA insertions. J Mol Evol 2008, 67(6):696-704.

26. Doyle JJ, Doyle JL, Ballenger JA, Palmer JD. The distribution and phylogenetic significance of a 50-kb chloroplast DNA inversion in the flowering plant family leguminosae. Molecular Phylogenet and Evol. 1996;5(2):429-438.

27. Gantt JS, Baldauf SL, Calie PJ, Weeden NF, Palmer JD. Transfer of *rpl*22 to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. Embo J. 1991;10(10):3073-3078.

28. Magee AM, Aspinall S, Rice DW, Cusack BP, Semon M, Perry AS, Stefanovic S, Milbourne D, Barth S, Palmer JD, Gray JC, Kavanagh TA, Wolfe KH. Localized hypermutation and associated gene losses in legume chloroplast genomes. Genome Res. 2010;20(12):1700-1710.

29. Smith SA, Donoghue MJ. Rates of molecular evolution are linked to life history in flowering plants. Science. 2008;322(5898):86-89.

30. Yao G, Jin JJ, Li HT, Yang JB, Mandala VS, Croley M, Mostow R, Douglas NA, Chase MW, Christenhusz MJM, Soltis DE, Soltis PS, Brockington SSF, Moore MJ, Yi TS, Li DZ. Plastid phylogenomic insights into the evolution of Caryophyllales. Molecular Phylogenet and Evol. 2019;134:74-86.

31. Wei Z, Pedley L. *Wisteria*. In: Flora of China. Volume 10. Edited by Wu ZY, Raren PH. St. Louis: Missouri Botanical Garden Press. 2010;188-189.

32. XiaoHong L, XiaoLing Z, JianJun Z. Floral flowering phenology and reproductive features of *Wisteria sinensis* Sweet. Journal of Jinggangshan University (Natural Science). 2010;31.

33. The Plant List [http://www.theplantlist.org/] Accessed on 1st Augest 2019.

34. Doyle JJ, Doyle JL, Ballenger JA, Dickson EE, Kajita T, Ohashi H. A phylogeny of the chloroplast gene *rbc*L in the Leguminosae: taxonomic correlations and insights into the evolution of nodulation. Am J Bot. 1997;84(4):541-554.

35. Hu JM, Lavin M, Wojciechowski MF, Sanderson MJ. Phylogenetic systematics of the tribe Millettieae (Leguminosae) based on chloroplast *trn*K/*mat*K sequences and its implications for evolutionary patterns in Papilionoideae. Am J Bot. 2000;87(3):418-430.

36. Pennington RT, Lavin M, Ireland H, Klitgaard B, Preston J, Hu JM. Phylogenetic relationships of basal papilionoid legumes based upon sequences of the chloroplast *trn*L intron. Syst Bot. 2001;26(3):537-556.

37. Li JH, Jiang JH, Fu CX, Tang SQ. Molecular systematics and biogeography of *Wisteria* inferred from nucleotide sequences of nuclear and plastid genes. J Syst Evol. 2014;52(1):40-50.

38. Wojciechowski MF, Sanderson MJ, Steele K.P. and Liston A. Molecular phylogeny of the "temperate herbaceous tribes" of Papilionoid legumes: a supertree approach. 2000.

39. Valder P. Wisterias, a comprehensive guide. Portland. Timber Press. 1995.

40. Ohwi J. *Wisteria*. In Flora of Japan. Edited by Meyer FG, Walker EH. Washington: Smithsonian Institution. 1965;572.

41. Liu LX, Wang YW, He PZ, Li P, Lee J, Soltis DE, Fu CX. Chloroplast genome analyses and genomic resource development for epilithic sister genera *Oresitrophe* and *Mukdenia* (Saxifragaceae), using genome skimming data. BMC Genomics. 2018;19.

42. Kim NR, Kim K, Lee SC, Lee JH, Cho SH, Yu Y, Kim YD, Yang TJ. The complete chloroplast genomes of two *Wisteria* species, *W. floribunda* and *W. sinensis* (Fabaceae). Mitochondrial DNA A. 2016;27(6):4353-4354.

43. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 2017;45(4).

44. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772-780.

45. Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. 2005;33:W686-W689.

46. Lohse M, Drechsel O, Bock R. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Curr Genet. 2007;52(5-6):267-274.

47. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve. Multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 2004;14(7):1394-1403.

48. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 2004;32:W273-W279.

49. Kurtz S, Schleiermacher C. REPuter: fast computation of maximal repeats in complete genomes. Bioinformatics. 1999;15(5): 426-427.

50. Thiel T, Michalek W, Varshney RK, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet. 2003;106(3):411-422.

51. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25(11): 1451-1452.

52. Posada D. jModelTest: Phylogenetic model averaging. Mol Biol Evol. 2008;25(7): 1253-1256.

53. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312-1313.

54. Miller AM, Pfeiffer P, Schwartz T. Creating the CIPRES science gateway for inference of large phylogenetic trees. 2010.

55. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Systematic Biol. 2012;61(3):539-542.

56. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in bayesian phylogenetics using Tracer 1.7. Systematic Biol. 2018;67(5):901-904.

57. Kumar S, Stecher G, Tamura K: MEGA7. Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33(7):1870-1874.

58. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 2018;4(1).

59. Lavin M, Herendeen PS, Wojciechowski MF. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. Systematic Biol;2005;54(4):575-594.

60. A. K. Shakryl. Leguminosae species from the Tertiary of Abkhazia. In: Advances in Legume Systematics. Part 4: The Fossil Record. Edited by P. S. Herendeen, D. L. Dilcher. Kew: Royal Botanic Gardens. 1992;189-206.

61. Bouckaert R, Heled J, Kuhnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ. BEAST 2: a software platform for bayesian evolutionary analysis. Plos Comput Biol. 2014;10(4).

62. Wang YH, Qu XJ, Chen SY, Li DZ, Yi TS: Plastomes of Mimosoideae: structural and size variation, sequence divergence, and phylogenetic implication. Tree Genet Genomes. 2017;13(2).

63. Wei Z. *Wisteria* Nutt. In: Flora reipublicae popularis Sinicae. Volume 40. Edited by Delectis Florae Reipublicae Popularis Sinicae Agendae Academiae Sinicae. 1994;183-188.

64. Palmer JD, Thompson WF. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. Cell. 1982;29(2):537-550.

65. Hirao T, Watanabe A, Kurita M, Kondo T, Takata K. Complete nucleotide sequence of the *Cryptomeria japonicia* D. Don. chloroplast genome and comparative chloroplast genomics: diversified genomic structure of coniferous species. BMC Plant Biol. 2008;8.

66. Hao DC, Xiao PG, Huang BL, Ge GB, Yang L. Interspecific relationships and origins of Taxaceae and Cephalotaxaceae revealed by partitioned bayesian analyses of chloroplast and nuclear DNA sequences. Plant Syst Evol. 2008;276(1-2):89-104.

67. Kode V, Mudd EA, Iamtham S, Day A. The tobacco plastid *acc*D gene is essential and is required for leaf development. Plant J. 2005;44(2):237-244.

68. Gurdon C, Maliga P. Two distinct plastid genome configurations and unprecedented intraspecies length variation in the *acc*D coding region in *Medicago truncatula*. DNA Res. 2014;21(4):417-427.

69. Wang Q, Dilcher DL, Zhu XY, Zhou YL, Lott TA. Fruits and Leaflets of *Wisteria* (Leguminosae, Papilionoideae) from the Miocene of Shandong Province, Eastern China. Int Journal Plant Sci. 2006;167(5):1061-1074.

70. Donoghue MJ, Smith SA. Patterns in the assembly of temperate forests around the Northern Hemisphere. Philos T R Soc B. 2004;359(1450):1633-1644.

71. Tiffney BH. Perspectives on the Origin of the Floristic Similarity between Eastern Asia and Eastern North-America. J Arnold Arboretum. 1985;66(1):73-94.

72. Xiang QY, Soltis DE, Soltis PS, Manchester SR, Crawford DJ. Timing the eastern Asian-eastern North American floristic disjunction: molecular clock corroborates paleontological estimates. Mol Phylogenet Evol. 2000;15(3):462-472.

73. Graham A. The role of land bridges, ancient environments, and migrations in the assembly of the North American flora. J Sys Evol.2018;56(5):405-429.

74. Hewitt G. The genetic legacy of the Quaternary ice ages. Nature. 2000;405(6789):907-913.

75. Hewitt GM. Genetic consequences of climatic oscillations in the Quaternary. Philos Trans R Soc Lond B Biol Sci. 2004;359(1442):183-195.

76. Wu ZY, Wu SG: A proposal for a new floristic kingdom (realm) – the E. Asiatic kingdom, its delimitation and characteristics. Floristic characteristics and diversity of east Asian plants. 1996.

77. Shi YF, Ren BH, Wang JT, Derbyshire E. Quaternary glaciation in China. Quaternary Sci Rev. 1986;5:503-507.

78. Liu KB. Quaternary History of the Temperate Forests of China. Quaternary Sci Rev. 1988;7(1):1-20.

79. Qiu YX, Fu CX, Comes HP. Plant molecular phylogeography in China and adjacent regions: tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora. Mol Phylogenet Evol. 2011;59(1):225-244.

80. Li JH, Tredici PD, Lemmon A, Moriatrty-Lemmon E, Zhao Y, Fu CX. Allopatric speciation in Asia contributed to the diversity anomaly between Eastern Asia and Eastern North America: evidence from anchored phylogenomics of *Stewartia* (Theaceae). Int J Plant Sci. 2019. (In press)

# Table

| Characteristics | *W. brachybotrys* | *W. floribunda* | *W. frutescens* | *W. sinensis* | *W. venusta* | *W. villosa* | *M. japonica* |
|---|---|---|---|---|---|---|---|
| Accession | PE01162281 (PE) | PE02002773 (PE) | PE1558524 (PE) | X18021 (ZJU) | 6462 (PE) | X18024 (ZJU) | X18050 (ZJU) |
| Locality | Kochi, Japan | Shizuoka, Japan | Carolina, America | Hunan, China | Beijing, China | Henan, China | Miyazaki, Japan |
| GenBank numbers | MN311167 | MN311168 | MN311169 | MN311170 | MN311171 | MN311172 | MN311163 |
| Clean Reads | 23,791,782 | 20,641,240 | 28,744,390 | 20,472,262 | 20,829,928 | 20,453,560 | 21,438,940 |
| Chloroplast genome size (bp) | 131,179 | 130,905 | 130,116 | 130,976 | 131,505 | 130,347 | 132,547 |
| GC content | 34.20% | 34.30% | 34.20% | 34.30% | 34.30% | 34.50% | 34.2% |
| Total number of genes | 110 | 110 | 110 | 110 | 110 | 110 | 110 |
| Protein coding genes | 76 | 76 | 76 | 76 | 76 | 76 | 76 |
| tRNA genes | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| Rrna genes | 4 | 4 | 4 | 4 | 4 | 4 | 4 |

Table 1 The basic characteristics of six *Wisteria* species and *Millettia japonica* plastomes.

# Additional Files

Additional file 1: Table S1-S6 Table S1. Information of species used for in this study. Table S2. List of synonymous substitution rates ($d_S$). Table S3. List of genes in *Wisteria* and *Millettia japonica* plastomes. Table S4. Statistical information of simple

sequences repeats (SSRs) and repeat types within *Wisteria*. Table S5. Genetic diversity information for each sequence region of *Wisteria*. Table S6. Divergence time estimates in millions of years (Ma) before for selected nodes (44.3 KB).

Additional file 2: Figure S1. Gene map of six *Wisteria* and *Millettia japonica* plastomes (1.72 MB).

Additional file 2: Figure S2. MAUVE alignment of *Wisteria* plastomes. Within each of the alignment, local collinear blocks are represented by blocks of the same color connected by lines (4.22 MB).

Additional file 2: Figure S3. Sequence identity plots among six Wisteria plastomes. CNS: conserved non-coding sequences; UTR: untranslated regions (1.15 MB).

# Figures
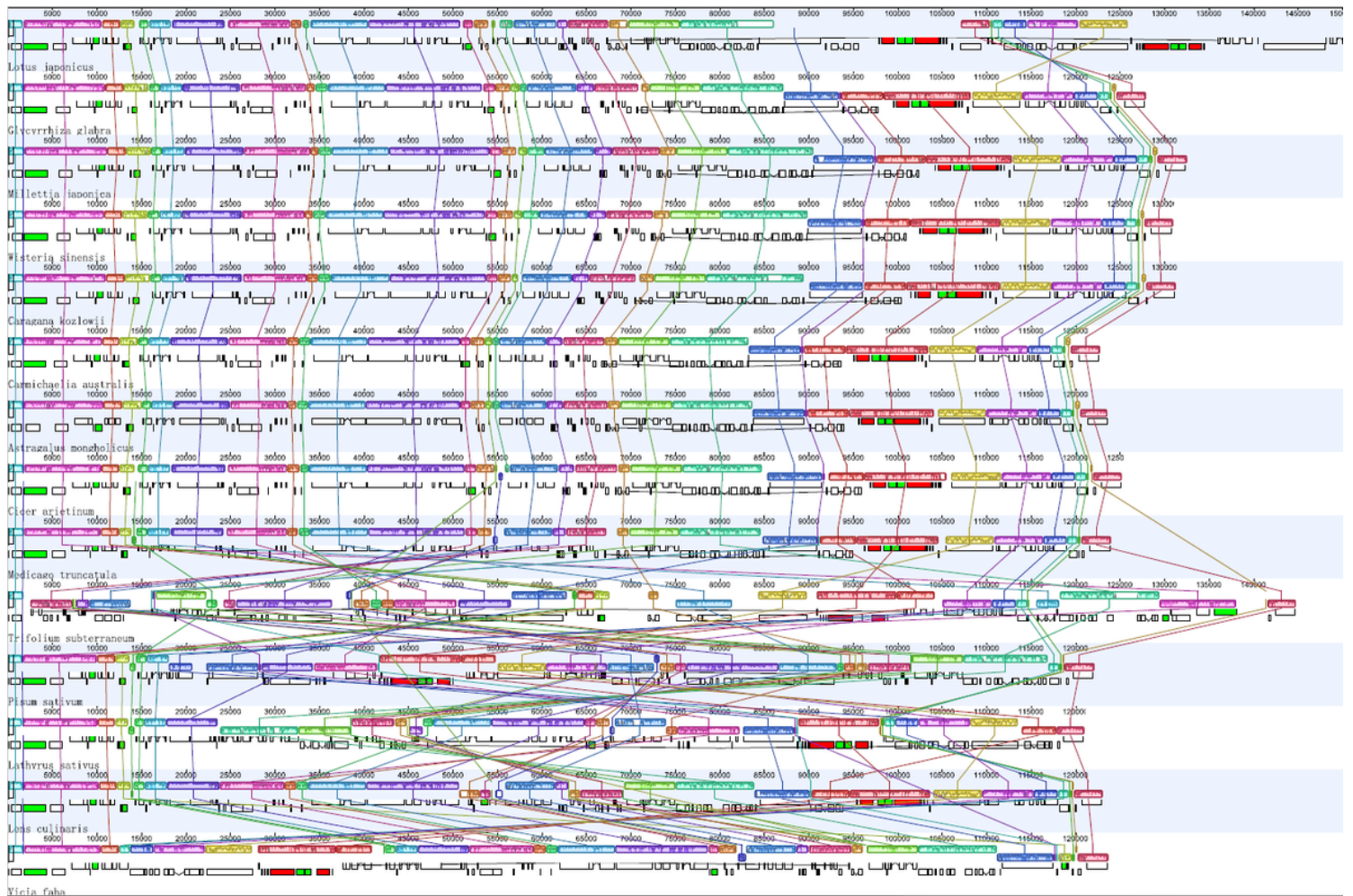


### Figure 1

Mauve alignment of IRLC species using Lotus japonicus as reference. The local collinear blocks are represented by blocks of the same color connected by lines.
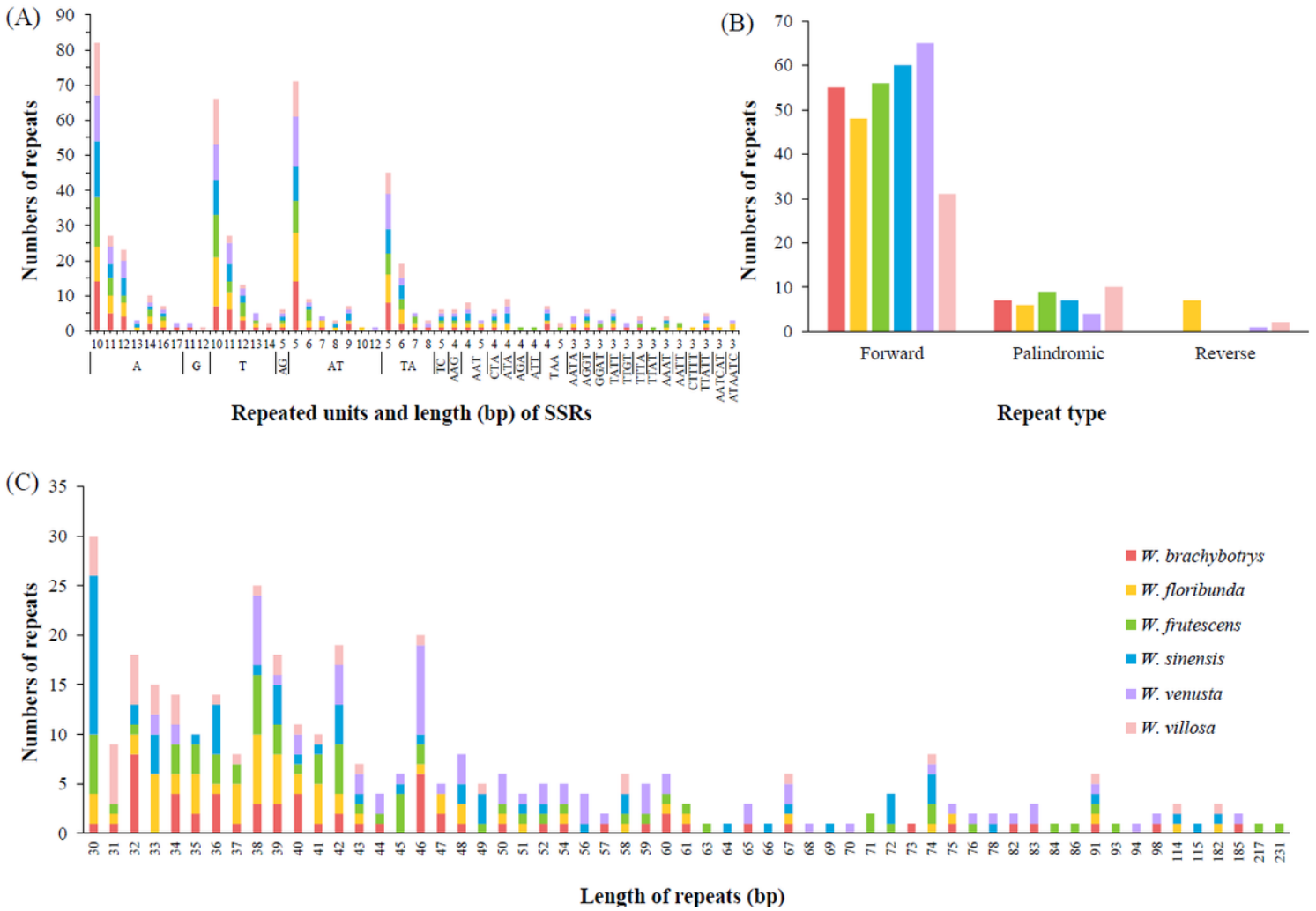
**Figure 2**

The length, type, and number of simple sequence repeats (SSRs) and repeated sequences of six Wisteria plastomes: (a) Number of SSRs in different types, (b) Frequency of repeat types, (c) Frequency of repeat types.
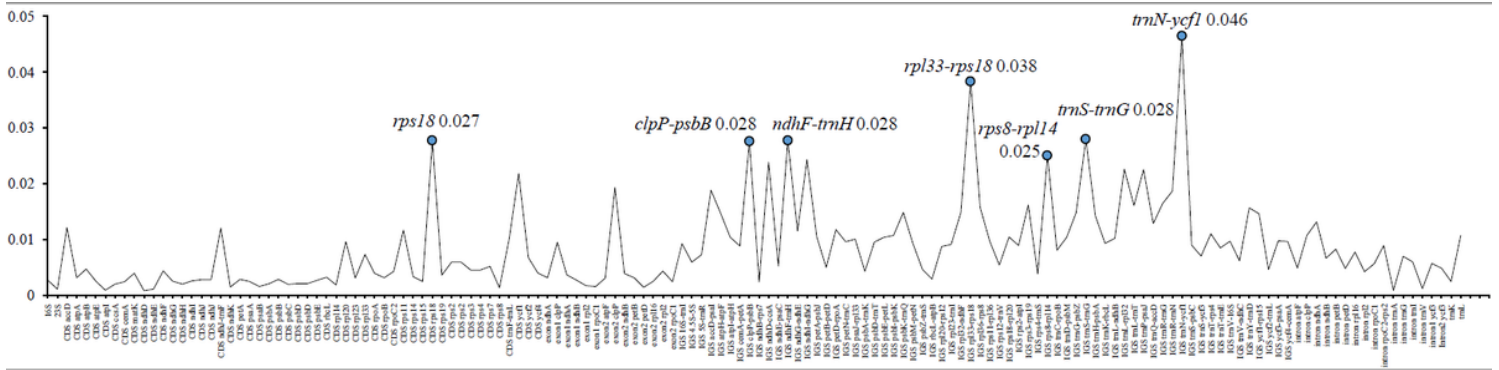


**Figure 3**

Nucleotide variability (Pi) values of candidate regions among six Wisteria plastomes. The blue point indicate the hotspot.
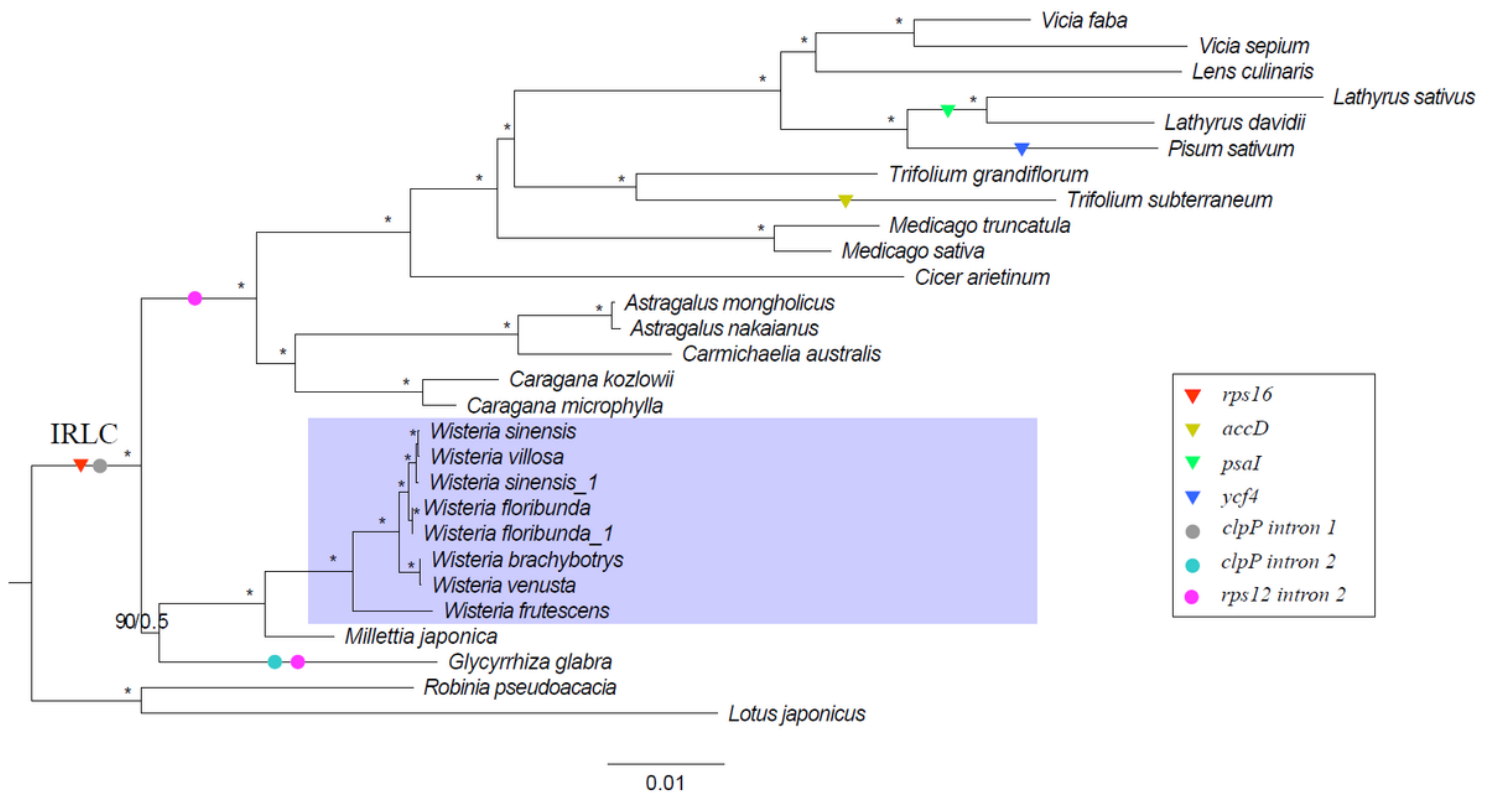
**Figure 4**

Phylogenetic relationships of IRLC taxa inferred from Maximum Likelihood (ML) and Bayesian Inference (BI) analyses of the 74 protein coding genes. ML topology shown with ML bootstrap support value/Bayesian posterior probability given at each node. Nodes with 100 ML bootstrap support value/1.0 Bayesian posterior probability are marked as asterisk. The loss of genes/introns is marked on the branch.
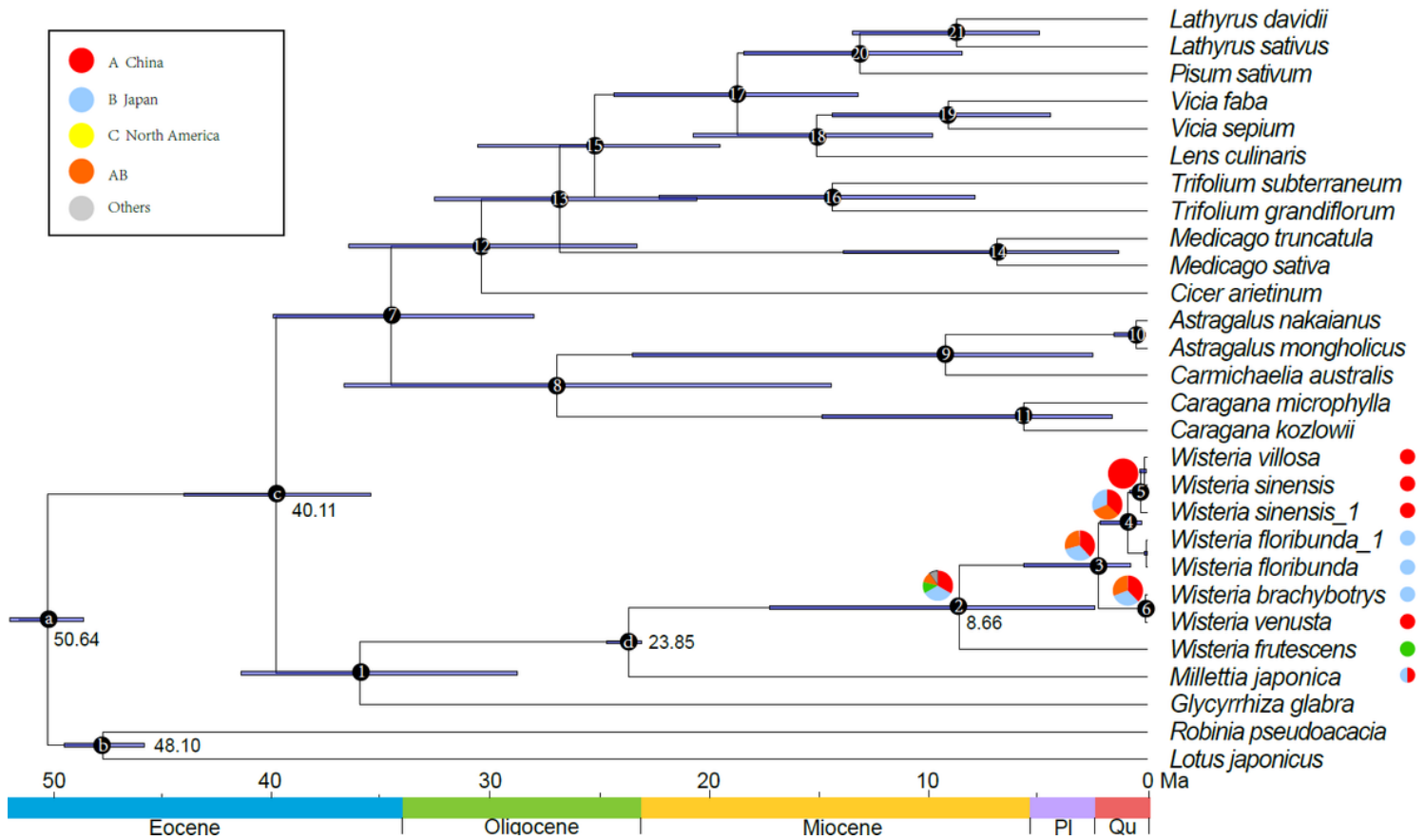
**Figure 5**

Phylogenetic chronogram and distribution range of IRLC species as inferred from Beast analyses based on four calibration points (nodes a-d). The divergence times for selected nodes are shown in Table S6.

## Supplementary Files

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

- FigureS3.pdf
- FigureS2.pdf
- additionalfile1.docx
- FigureS1.pdf