

RESEARCH ARTICLE

Complete Plastid Genome Sequencing of Four *Tilia* Species (Malvaceae): A Comparative Analysis and Phylogenetic Implications

Jie Cai^{1,2,3}, Peng-Fei Ma¹, Hong-Tao Li¹, De-Zhu Li^{1*}

1 Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, Yunnan, China, **2** School of Life Science, Yunnan University, Kunming 650091, Yunnan, China, **3** Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming 650204, China

* dzl@mail.kib.ac.cn



OPEN ACCESS

Citation: Cai J, Ma P-F, Li H-T, Li D-Z (2015) Complete Plastid Genome Sequencing of Four *Tilia* Species (Malvaceae): A Comparative Analysis and Phylogenetic Implications. PLoS ONE 10(11): e0142705. doi:10.1371/journal.pone.0142705

Editor: Tongming Yin, Nanjing Forestry University, CHINA

Received: August 22, 2015

Accepted: October 26, 2015

Published: November 13, 2015

Copyright: © 2015 Cai et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Four *Tilia* plastid genomes sequenced in current study are available in NCBI database. The Genbank numbers are KT894772, KT894773, KT894774, KT894775.

Funding: This work was supported by grants from the National Key Basic Research Program of China (No. 2014CB954100) and Kunming Institute of Botany, Chinese Academy of Sciences (No. 2014KIB02) to DZL. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Tilia is an ecologically and economically important genus in the family Malvaceae. However, there is no complete plastid genome of *Tilia* sequenced to date, and the taxonomy of *Tilia* is difficult owing to frequent hybridization and polyploidization. A well-supported interspecific relationships of this genus is not available due to limited informative sites from the commonly used molecular markers. We report here the complete plastid genome sequences of four *Tilia* species determined by the Illumina technology. The *Tilia* plastid genome is 162,653 bp to 162,796 bp in length, encoding 113 unique genes and a total number of 130 genes. The gene order and organization of the *Tilia* plastid genome exhibits the general structure of angiosperms and is very similar to other published plastid genomes of Malvaceae. As other long-lived tree genera, the sequence divergence among the four *Tilia* plastid genomes is very low. And we analyzed the nucleotide substitution patterns and the evolution of insertions and deletions in the *Tilia* plastid genomes. Finally, we build a phylogeny of the four sampled *Tilia* species with high supports using plastid phylogenomics, suggesting that it is an efficient way to resolve the phylogenetic relationships of this genus.

Introduction

Tilia L. (basswood or lime-tree) is a genus of the family Malvaceae in the order Malvales, which contains 23 species of deciduous tree disjunctly distributed in the temperate area across Asia, Europe and North America [1–3]. Trees of *Tilia* are ecologically and economically important. Most species are the dominant element in the broad-leaved temperate forests, and many species are used for timber, honey resources and cultivated worldwide for ornamental purpose.

Traditionally *Tilia* has been placed in its own family Tiliaceae [4, 5]. However, molecular evidence strongly supported the polyphyly of the traditionally circumscribed Tiliaceae [6]. A clade of Tilioideae was resolved and *Tilia* was closely related to the genera *Craigia* and

Competing Interests: The authors have declared that no competing interests exist.

Mortonioidendron in the family of Malvaceae [7, 8]. Within the Malvaceae family, *Tilia* is a distinct genus characterized by woody habit and paddle-shaped bracts of flowers. However, the taxonomy of *Tilia* is difficult and controversial due to limited taxonomic characters and frequently occurred hybridization [3], and species relationships within the genus are poorly known. Previous molecular studies mainly focused on population genetic analyses of some *Tilia* species employing plastid PCR-RFLP markers [9], random amplified polymorphic DNA (RAPD) markers [10–12], microsatellite markers [13]. Recent molecular phylogenetic studies have used chloroplast regions (*rpL32-trnL* and *ndhF-rpL32*) or ribosomal internal transcribed spacer (ITS) to reconstruct the phylogeny of selected species in *Tilia*, but without satisfaction [14, 15]. An expanded attempt of seven plastid regions and three low-copy nuclear regions on comprehensively sampled taxa did also not well resolve interspecific relationships of *Tilia* [16]. The age of *Tilia* trees when they begin to flower and produce seed ranges from six to 40 years old [3]. The slow nucleotide substitution rates of *Tilia* may be attributed to their long generation times.

The recent availability of the next-generation sequencing techniques has enabled generating large amounts of sequence data at relatively low cost [17–19]. Whole plastid genome is increasingly used for phylogenetic analyses and has proven to be effective in resolving difficult phylogenetic relationships [20–22]. Plastid genomes of angiosperms are well known to be highly conserved [23, 24]. They usually have a circular structure of two copies of large inverted repeats (IR) separated by small (SSC) and large (LSC) single-copy regions [25]. With a size ranging from 120 to 160 kb in general, they also exhibit highly conserved gene content and order [23, 24].

The number of sequenced plant plastid genomes increased rapidly during last decade due to the establishment of the next-generation sequencing techniques. Within the Malvaceae family, however, there are only two genera *Gossypium* [26, 27] and *Theobroma* [28] having their plastid genomes sequenced to date. At present, there is no complete plastid genome from *Tilia* sequenced despite of its ecological and economic importance.

To better understand the evolution of plastid genome and explore the potential of phylogenomics basing on plastid genome sequence to clarify interspecific relationships in *Tilia*, four representative *Tilia* species were sequenced using next-generation Illumina sequencing-by-synthesis technology. The main purposes of this study are to (1) gain insights into the structure of plastid genome of *Tilia* as well as in Malvaceae; and (2) explore the feasibility of plastid phylogenomics in reconstructing a solid interspecific relationships of the long-lived tree genus *Tilia*.

Materials and Methods

Plant material

Four species were chosen as representatives of *Tilia* for plastid genome sequencing. Studied species are commonly found in the public woodland where they are native (Table 1) and no specific collecting permits required for sampling. *T. amurensis* is listed in the China’s protected species Grade II, but it is not require permit to collect a few twigs and leaves for vouchers and DNA extraction. Healthy and fresh leaves were collected from a single individual for DNA

Table 1. Sequencing information for the four *Tilia* species used in this study.

Species	Voucher No.	Locality	GenBank numbers	Clean reads	Mean coverage
<i>T. amurensis</i>	11CS2872	Harbin, Heilongjiang, China	KT894772	5,021,460	282x
<i>T. mandshurica</i>	11CS2873	Harbin, Heilongjiang, China	KT894773	10,712,702	602x
<i>T. oliveri</i>	12CS5580	Wushan, Chongqing, China	KT894774	13,141,606	739x
<i>T. paucicostata</i>	13CS6898	Zhongdian, Yunnan, China	KT894775	5,074,944	285x

doi:10.1371/journal.pone.0142705.t001

extraction. The voucher specimens of sampled species were all deposited at the Herbarium of Kunming Institute of Botany, Chinese Academy of Sciences (KUN) ([Table 1](#)).

DNA extraction and template amplification

Total genomic DNA was extracted from ~100 mg of leaf material using a modified CTAB method [[29](#), [30](#)], and their quality was assessed by agarose gel electrophoresis. We amplified the entire plastid genome using long-range PCR and 9 primer pairs as described in Yang et al. 2014 [[30](#)]. Briefly, amplification was performed using Takara PrimeSTAR GXL DNA polymerase (TAKARA BIO INC.) in 25- μ l reaction mixtures with 30–100 ng of DNA template. The PCR amplification conditions were the same as those of Yang et al. 2014 [[30](#)]. All amplifications were successful and amplicon DNA concentrations were determined by visual approximation using gel electrophoresis. Subsequently, the 9 long-ranged PCR products were pooled together in roughly equal mass mixtures for genome sequencing.

Illumina sequencing, assembly, and annotation

Pooled amplified plastid DNAs (6 μ g) were sheared for short-insert (500 bp) sequencing libraries construction according to the manufacturer's instructions (Illumina). The 90 bp paired-end reads were generated on an Illumina Hiseq 2000 at BGI Shenzhen, China. Illumina raw reads were first quality trimmed using NGS QC Tool Kit [[31](#)] (cut-off value for percentage of read length = 80 and for Phred quality score = 30). The plastid genomes from the filtered clean reads were assembled using the CLC Genomics Workbench v. 6.5 (CLC Bio) de novo assembly program at hash length of 63 and with a minimum contig length of 1 kb. The assembled contigs were analyzed by a BLAST of the nucleotide database at the NCBI (<http://www.ncbi.nlm.nih.gov/>), and these aligning to the published plastid DNA sequences were collected for genome finishing. The annotation of completed plastid genomes was carried out using the program DOGMA [[32](#)]. We then manually adjusted the start and stop codons and intron/exon boundaries if necessary. The annotated GenBank files were used to draw the circular plastid genome maps using the OrganellarGenomeDRAW (OGDRAW) [[33](#)].

Comparison of *Tilia* plastid genome with other Malvaceae genera

The complete plastid genomes of *Gossypium hirsutum* (GenBank NC_007944) and *Theobroma cacao* (GenBank HQ336404) were downloaded from NCBI for comparison. The general plastid genome characters were compared between these two and our sequenced four *Tilia* species. Any large structural events such as gene order rearrangement and IR expansion/contraction were recorded. To investigate the difference in genome size between *Gossypium* and *Tilia*, the plastid DNA sequences of *G. hirsutum* and *T. mandshurica* were partitioned into genes, introns and intergenic spacers and then were compared separately.

Sequence analysis for *Tilia* species

The plastid genome sequences of four *Tilia* species and *G. hirsutum* were aligned using MAFFT v. 7.215 [[34](#)] in the default sets and manually adjusted in MEGA 5.0 [[35](#)]. Sequence divergence between the four *Tilia* plastid genome sequences was calculated as uncorrected p-distance using MEGA 5.0. The percentage of variable characters for each noncoding region with an aligned length >200 bp in the genome was calculated as described in Zhang et al [[17](#)]. We scored the number of transitional and transversional substitutions among the four *Tilia* plastid genome sequences and decided the direction of substitutions using *G. hirsutum* as out-group. The indels among them were polarized into insertions and deletions in the same way.

Phylogenetic inference

We downloaded the plastid genome sequences of two species of Brassicaceae, *Arabidopsis thaliana* (GenBank NC_000932) and *Brassica napus* (GenBank NC_016734) as outgroups. The ingroup taxa included one *Theobroma* species, two *Gossypium* species (the second being *G. hirsutum* with GenBank NC_016734) and four *Tilia* species sequenced here. The orientation of the SSC regions from two Brassicaceae species and *Theobroma cacao* were manually reversed for alignment. The alignment of nine plastid genomes with one IR region removed was conducted with MAFFT v. 7.215 [34] in the default sets, followed by manual adjustments in MEGA 5.0 [35]. Phylogenetic analysis using maximum likelihood (ML) method was performed using RAxML v. 8.0.20 [36]. Both the unpartitioned and partitioned ML analyses were performed with the dataset dividing into three partitions corresponding to the LSC, SSC and IR region of the plastid genome. The ML tree was constructed with the combined rapid bootstrap of 500 replicates and search for the best tree in a single run under the GTR + G model as suggested in the RAxML manual. Bayesian analysis was performed using MrBayes v. 3.2 [37] with the GTR + I + G model in the unpartitioned way. The Markov chain Monte Carlo (MCMC) algorithm was run for two million generations with trees sampled every 100 generations. The convergence was reached with the average standard deviation of split frequencies (ASDFs) following 0.01. The first 25% of trees generated were discarded as burn-in and the remaining trees were used to build majority-rule consensus tree. To test the rate of evolution of *Tilia* relative to other Malvaceae species we applied Tajima relative rate test [38] on the plastid genome sequence alignment. The relative rates of evolution between each of the four *Tilia* species and *G. hirsutum* were evaluated using the *Theobroma cacao* as outgroup.

Results

Sequencing of *Tilia* plastid genomes

Illumina 90-bp paired-end sequencing of long-rang PCR amplified plastid DNA generated 5,021,460–13,141,606 clean reads for the four sampled *Tilia* species, with an average sequencing depth from 282× to 739× (Table 1). Using the combination of de novo and reference-guided assembly, we obtained the complete plastid nucleotide sequences for all four species. The determined nucleotide sequences of the four plastid genomes range narrowly from 162,653 bp in *Tilia paucicostata* to 162,796 bp in *Tilia mandshurica* (Table 2). They all have a genome structure resembling those of the vast majority of angiosperms, consisting of a pair of IRs separated by LSC and SSC, and the gene map of *T. mandshurica* plastid genome is presented in Fig 1 as a representative. The four genomes encode an identical set of 130 genes, of which 113 are unique and 17 are duplicated in the IR regions (Table 2), and the arrangements of these 130 genes in them are totally collinear. The 113 unique genes include 79 protein-coding genes, 30 tRNA genes and 4 rRNA genes. They also have an identical GC content of 36.5% that is similar to other angiosperms plastid genomes [23, 24].

Comparisons with other Malvaceae plastid genomes

The comparisons of general genomic features between *Gossypium hirsutum* [26] as a representative species from genus *Gossypium*, *Theobroma cacao* [28] and four sequenced *Tilia* species were presented in Table 2. All plastid genomes except for *G. hirsutum* share identical complements of coding genes with similar order. The *infA* gene in the plastid genome of genus *Gossypium* was inferred to be pseudogene due to the presence of frameshift indels [26]. In addition, the GC content of Malvaceae species is almost the same (Table 2).

Table 2. Comparison of Malvaceae plastid genomes sampled in this study.

	<i>Gossypium hirsutum</i>	<i>Theobroma cacao</i>	<i>Tilia amurensis</i>	<i>Tilia mandshurica</i>	<i>Tilia oliveri</i>	<i>Tilia paucicostata</i>
Size (bp)	160,301	160,604	162,715	162,796	162,734	162,653
LSC (bp)	88,816	89,395	91,124	91,127	91,095	91,139
SSC (bp)	20,269	20,187	20,397	20,371	20,381	20,380
IR (bp)	25,608	25,511	25,597	25,649	25,629	25,567
Number of protein-coding genes ^a	84 (6)	85 (6)	85 (6)	85 (6)	85 (6)	85 (6)
Number of tRNA genes ^a	37 (7)	37 (7)	37 (7)	37 (7)	37 (7)	37 (7)
Number of rRNA genes ^a	8 (4)	8 (4)	8 (4)	8 (4)	8 (4)	8 (4)
GC content (%)	37.3	36.9	36.5	36.5	36.5	36.5

^aThe numbers in parenthesis indicate the genes duplicated in the IR regions.

doi:10.1371/journal.pone.0142705.t002

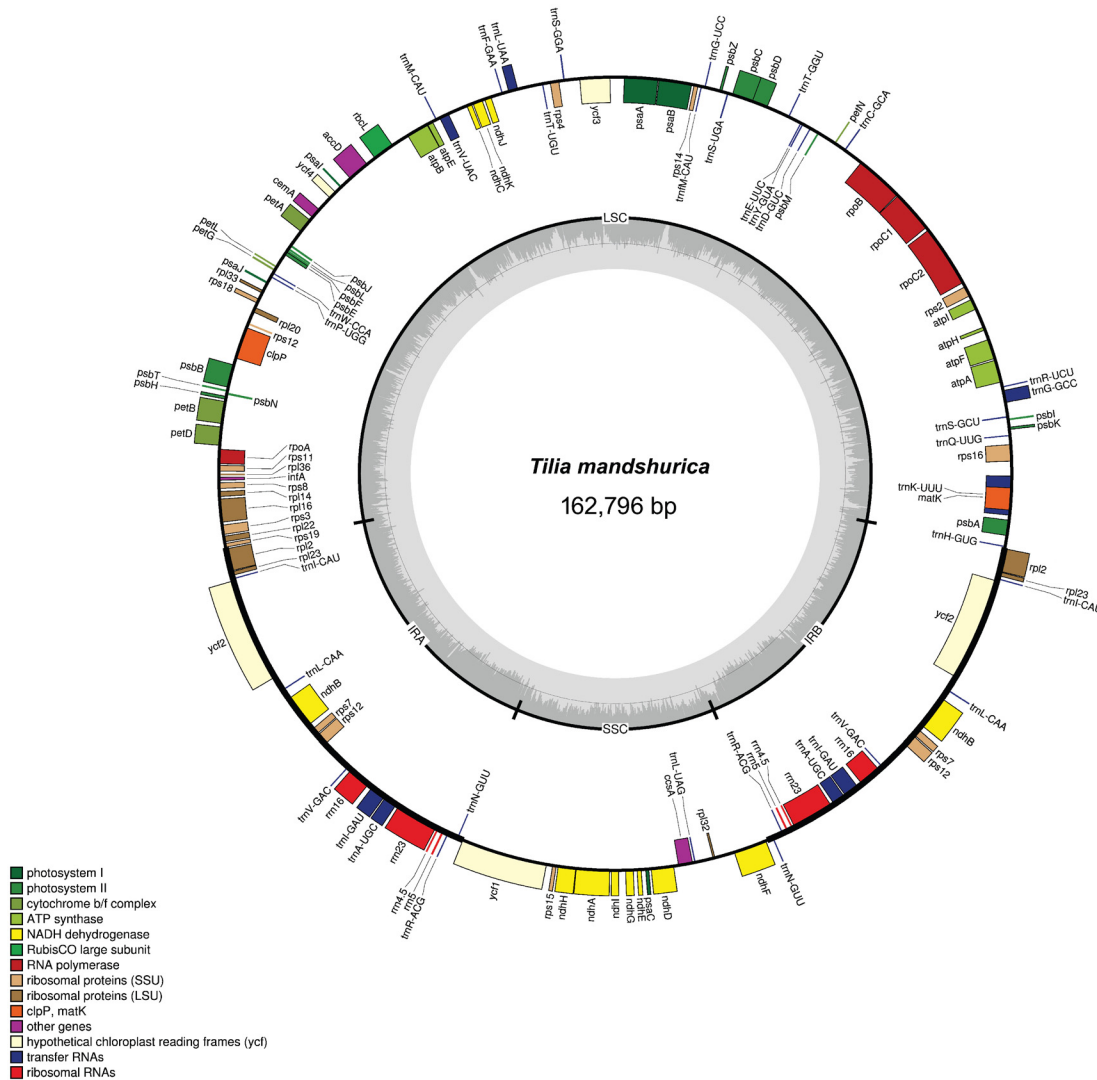


Fig 1. Circular gene map of the plastid genome of *Tilia mandshurica*. Genes drawn within the circle are transcribed clockwise, while those drawn outside are transcribed counterclockwise. Genes are color-coded according to their functional groups.

doi:10.1371/journal.pone.0142705.g001

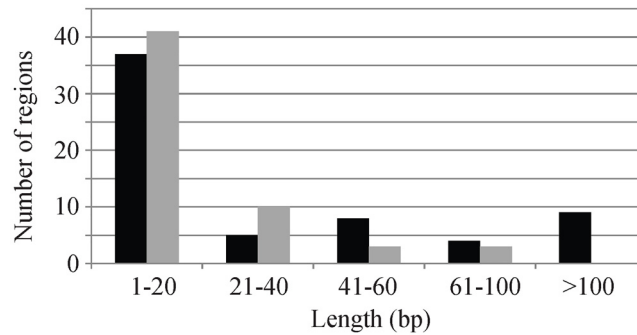


Fig 2. The number of noncoding regions with different sizes between the plastid genomes of *Tilia mandshurica* and *Gossypium hirsutum*. The black bar indicates the regions are larger in *T. mandshurica*, while the gray bar indicates the regions are larger in *G. hirsutum*.

doi:10.1371/journal.pone.0142705.g002

All plastid genomes of Malvaceae possess a typical quadripartite structure of angiosperms. However, the SSC region in the assembled plastid genomes of *Gossypium* and *Tilia* is in the reverse orientation relative to *Theobroma* and most other angiosperms. The four *Tilia* species have the same IR and SSC boundary with 36-bp sequence of the *ycf1* gene extending into the IR regions. The *ycf1* gene also extended into the IR regions with 98-bp sequence duplication in *G. hirsutum*. However, this gene is wholly confined in the SSC region in *Theobroma cacao* while the *ndhF* gene had 6-bp sequence duplicated in the IR regions.

In terms of genome size, the plastid genome of *Tilia* (162,653–162,796 bp) is slightly larger than those of *Gossypium* (159,039–160,433 bp) and *Theobroma* (160,604 bp). We selected *T. mandshurica* and *G. hirsutum* as representative species to investigate the trend toward increased genome size in *Tilia*. The whole plastid genome of *T. mandshurica* is 2495 bp larger than that of *G. hirsutum*, and all three regions LSC, SSC and IR of *T. mandshurica* are larger with LSC (2311 bp) accounting for most variation in genome size (Table 2). Among the 78 common unique protein-coding genes between these two genomes there are 11 genes with difference in length. Four genes (*atpI*, *ccsA*, *rbcl*, and *ycf1*) are larger in *T. mandshurica* while seven genes (*accD*, *matK*, *ndhF*, *petB*, *rpl22*, *rpoA*, and *rpoC2*) are larger in *G. hirsutum*, accounting for 63 bp and 180 bp of the variation in genome size respectively. On the other hand, the majority of noncoding (intergenic and intron) regions (120 of 153) show variations in length. The number of noncoding regions that are larger is similar in *T. mandshurica* and *G. hirsutum* (63 versus 47) (Fig 2). Nevertheless, the 9 regions with length difference above 100 bp are all those larger in *T. mandshurica* than *G. hirsutum*. These 9 regions are *ycf3-trnS-GGA*, *ndhC-trnV-UAC*, *trnH-GUG-psbA*, *trnT-UGU-trnL-UAA*, *trnR-UCC-atpA*, *psbZ-trnG-UCC*, *atpB-rbcL*, *trnC-GCA-petN*, and *trnK-UUU-rps16*. These length differences are mainly caused by the large insertions in the *T. mandshurica* (or deletions in *G. hirsutum*) plastid genome and there are 8 indels (insertions or deletions) in all larger than 100 bp. The total length of these indels is 1990 bp, explaining ~80% of variation in genome size.

Sequence divergence of *Tilia* plastid genomes

We plotted sequence identity among the four *Tilia* plastid genomes using the mVISTA software [39] with *T. mandshurica* as a reference. A genome-wide alignment reveals globally high sequence similarity (> 90% identity) among them (Fig 3). The overall sequence divergence estimated by p-distance among the four genomes was only 0.0013. The pairwise p-distance between the four species ranged from 0.0004 to 0.0021, and the *T. amurensis* has a somewhat larger sequence divergence from the others within the genus. We also compared the sequence

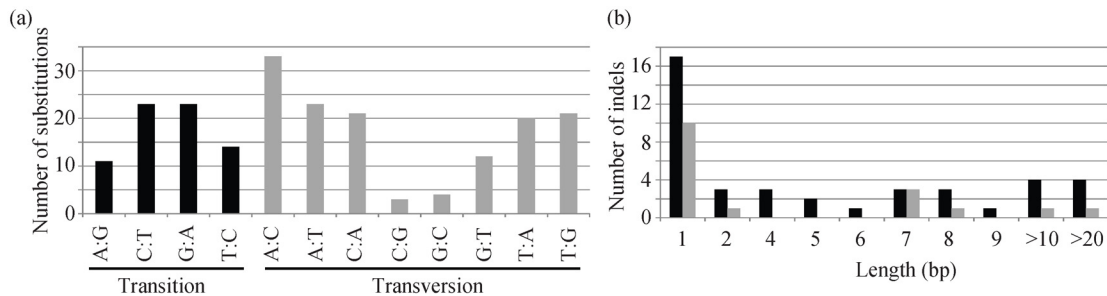


Fig 5. Nucleotide sequence variations identified in the four *Tilia* plastid genomes. (A) The nucleotide substitution patterns in the plastid genomes. **(B)** The length distribution of insertions (black) and deletions (gray) in the plastid genomes.

doi:10.1371/journal.pone.0142705.g005

in the whole genome (Fig 5A). Among them, there were 71 transitions and 137 transversions and the transition/transversion ratio ($Ts/Tv = 0.52$) was nearly identical to the 1:2 ratio expected from equal rates of transition and transversion. More specifically, we observed a significant excess of A to C and lack of G to C and C to G transversions relative to all other substitutions (Fig 5A). The insertions or deletions in plastid genomes of *Tilia* were also inferred with *G. hirsutum* as an outgroup. The number of insertions (41) is much larger than the number of deletions (17) (Fig 5B). And about 41% of the total insertions and 59% of the total deletions are of size 1 bp. The longest insertion and deletion was 76 and 21 bp, respectively. In addition, 52 of these 58 indels were associated with tandem repeats.

Phylogenetic analyses of *Tilia*

The data matrix for phylogenetic analyses contained the whole plastid genome with one IR region removed for nine taxa, including six species of Malvaceae and two outgroups from Brassicaceae. The data set comprised of 145,679 nucleotide positions with 3,377 informative sites for the ingroup taxa. However, there were only 20 informative sites for the four *Tilia* species. Unpartitioned ML analyses resulted in a fully resolved tree with 5 of the 6 nodes supported by 100% bootstrap values (Fig 6). ML analyses partitioned by the three plastid genomic regions (LSC, SSC, and IR) yielded an identical topology with the same 5 nodes 100% supported (data not shown). The remaining one received 81% and 82% bootstrap values from unpartitioned and partitioned analyses, respectively. In Bayesian analysis, the identical topology was obtained with a posterior probability (PP) of 1.0 for all nodes (Fig 6). *T. amurensis* was sister to the other three sampled *Tilia* species, and *T. mandshurica* was then sister to the grouping of *T. oliveri* and *T. paucicostata* (Fig 6).

As can be seen in Fig 6, the branch length leading to *Tilia* and especially those among species were extremely short. This indicated that *Tilia* likely had a slow rate of evolution relative to *Gossypium*. To explore the rate heterogeneity existed between *Gossypium* and *Tilia*, we applied Tajima relative rate test [38] on the sequence alignments of whole plastid genome. This test showed that all the four sampled *Tilia* species were evolving significantly slower than the *Gossypium* ($p < 0.001$; Table 3).

Discussion

Comparison of plastid genome within Malvaceae

In this study, we determined the complete plastid genome sequences from four *Tilia* species using Illumina sequencing technology. In addition to *Gossypium* and *Theobroma* [26–28], *Tilia* is the third genus within the family Malvaceae to have its complete plastid genome sequenced.

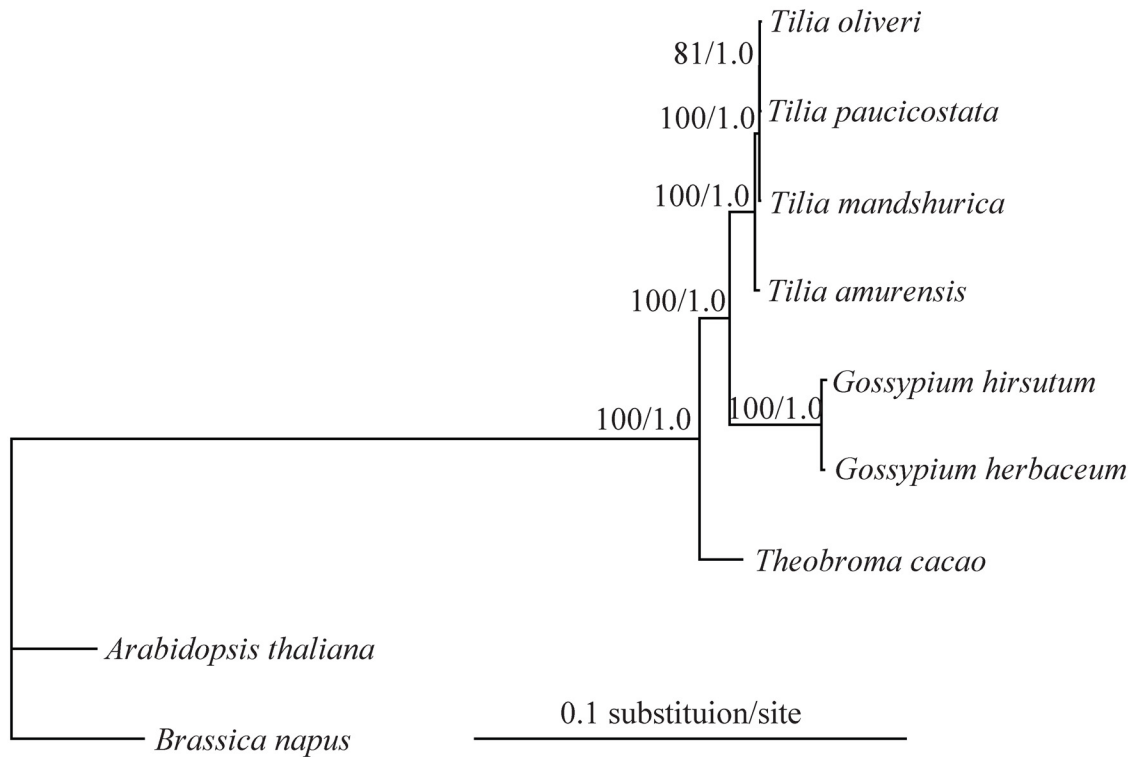


Fig 6. Maximum likelihood phylogeny of the seven Malvaceae species based on the complete plastid genome sequences. The numbers associated with the nodes are bootstrap support and posterior probability values.

doi:10.1371/journal.pone.0142705.g006

All of the four *Tilia* plastid genomes possess a typical quadripartite structure of angiosperms with a pair of inverted repeats dividing the whole genome into two single copy regions (Fig 1). In comparison to *Gossypium* and *Theobroma*, no significant structural reconfigurations such as inversions or gene relocations were detected in the four *Tilia* plastid genomes. The six plastid genomes analyzed here are rather conserved, and only with minor variations in the junctions between the SSC and IRs regions, which are usually different within the same family [40–42]. The four *Tilia* plastid genomes have the same boundary between the SSC and IRs regions with 36-bp sequence of the *ycf1* gene duplicated in IRs and the length of duplication is 98 bp in *G.*

Table 3. Results of Tajima relative rate test.

Ingroup1	Ingroup2	Outgroup	Identical sites	Divergent sites	Ingroup1 specific ^a	Ingroup2 specific	Outgroup specific	Chi-square statistic	p-value	Slow
<i>Gossypium hirsutum</i>	<i>Tilia mandshurica</i>	<i>Theobroma cacao</i>	125,627	113	2456	725	1801	941.96	< 0.001	<i>Tilia mandshurica</i>
<i>Gossypium hirsutum</i>	<i>Tilia paucicostata</i>	<i>Theobroma cacao</i>	125,624	112	2458	727	1801	940.77	< 0.001	<i>Tilia paucicostata</i>
<i>Gossypium hirsutum</i>	<i>Tilia amurensis</i>	<i>Theobroma cacao</i>	125,670	104	2452	694	1813	982.38	< 0.001	<i>Tilia amurensis</i>
<i>Gossypium hirsutum</i>	<i>Tilia oliveri</i>	<i>Theobroma cacao</i>	125,625	112	2451	720	1804	944.93	< 0.001	<i>Tilia oliveri</i>

^a These nucleotide sites were identical in ingroup2 and outgroup but not in ingroup1, and the same idea applies for ingroup2 and outgroup specific.

doi:10.1371/journal.pone.0142705.t003

hirsutum. In the *Theobroma cacao* plastid genome, it is the *ndhF* gene rather than *ycf1* extending into the IRs. However, the SSC region in the *Tilia* plastid genomes was assembled to be in the reverse orientation relative to *Theobroma* and majority of angiosperms while identical to *Gossypium*. The SSC region could exist in two orientations in plastid genomes [43, 44] and this result does not reflect any differences in gene order in these genomes.

In terms of gene content, the *Tilia* plastid genomes share the same set of 85 protein genes with *Theobroma* rather than *Gossypium* (Table 2), although *Tilia* has a closer relationship to *Gossypium* (Fig 6). The *infA* gene that has become a pseudogene in *Gossypium* has intact reading frame and can be functional in *Tilia*. This result indicates that pseudogenization of *infA* independently occurred in the *Gossypium* lineage. In terms of genome size, the *Tilia* plastid genomes show a trend toward increased size within Malvaceae. We divided the whole plastid genome into genes and intergenic spacers and compared their lengths between *Tilia* and *Gossypium*, finding that the variation in genome size is mostly due to length differences in the non-coding regions (Fig 2). Furthermore, a few large indels (> 100 bp) instead of many small indels can be responsible for ~80% of the total length difference between these two genera.

Molecular evolution of *Tilia* plastid genome sequences

In addition to the rather conserved evolution of genome structure, the genetic divergence is extremely low among the four *Tilia* plastid genomes, and there is no mutation hotspot region identified across the genome (Figs 3 and 4) as in other angiosperms [17, 45, 46]. Within the four species, the *T. amurensis* is slight divergent relative to the others. Among a total of 208 substitutions identified among the four *Tilia* plastid genomes (Fig 5A), the transition/transversion ratio of 0.52 is very close to the expected 1:2 ratio in considering equal rates of transition and transversion. The rate of transitions is not significantly elevated as demonstrated recently in many other plant plastid genomes [22, 27, 47]. However, there are a significant excess of A to C and lack of G to C and C to G transversions among the eight types of transversions (Fig 5A). The underlying mechanism under this phenomenon would require more studies to clarify in the future.

In addition to substitutions, the indels are another important class of genetic variation [48–50]. The insertions occur much more frequently than the deletions in the *Tilia* plastid genomes (Fig 5B), consistent with the trend toward larger genome size within Malvaceae. Approximately half of the total indels are of size 1 bp and mainly occur on the homopolymer regions, and almost entirely of the remaining indels > 1 bp are associated with tandem repeats. This result indicates that these indels very likely originated as mutation events formed by slipped strand mispairing [51].

Using plastid phylogenomics to resolve phylogeny of *Tilia*

As suggested by the low sequence divergence observed among the four *Tilia* plastid genomes, the slowdown in evolutionary rates may occur in this long-lived tree genus. Relative rate test does demonstrate a reduced rate of evolution in the *Tilia* relative to its relative genus *Gossypium* (Table 3). The slow evolutionary rate in *Tilia* can be largely attributed to its long generation times [3]. Although there are a few phylogenetic informative sites contained in the *Tilia* plastid genomes, the phylogeny of the four sampled *Tilia* species is well resolved (Fig 6). As previous studies employing multi nuclear and plastid DNA regions have failed to reconstruct the phylogeny of *Tilia* [14, 16], our successful reconstruction of phylogeny for the four *Tilia* species sampled here indicates that plastid phylogenomics holds promise in resolving the inter-specific relationships of this ecologically and economically important genus.

Acknowledgments

We wish to thank Prof. C. Donald Pigott for his support to this study, to Prof. Ting-Shuang Yi for his valuable comments on the manuscript. We are very grateful to Prof. Jun-Bo Yang for his help during the experiments.

Author Contributions

Conceived and designed the experiments: JC DZL. Performed the experiments: JC HTL. Analyzed the data: PFM HTL JC. Contributed reagents/materials/analysis tools: JC PFM HTL. Wrote the paper: PFM JC HTL DZL.

References

1. The Angiosperm Phylogeny Group. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot J Linn Soc.* 2009; 161: 105–121.
2. Mabberley DJ. *Mabberley's plant-book: a portable dictionary of plants, their classifications, and uses.* 3rd ed. Cambridge: Cambridge University Press. 2008.
3. Pigott CD. *Lime-trees and basswoods. A biological monograph of the genus Tilia.* Cambridge: Cambridge University Press. 2012.
4. Engler A. *Syllabus der Pflanzenfamilien.* Berlin: Borntraeger. 1921.
5. Cronquist A. *An intergrated system of classification of flowering plants.* New York: Columbia University Press. 1981
6. Bayer C, Fay MF, de Bruijn AY, Savolainen V, Morton CM, Kubitzki K, et al. Support for an expanded family concept of Malvaceae within a recircumscribed order Malvales: a combined analysis of plastid *atpB* and *rbcl* sequences. *Bot J Linn Soc.* 1999; 129: 267–303
7. Alverson WS, Whitlock BA, Nyffeler R, Bayer C, Baum DA. Phylogenetic analysis of the core Malvales based on sequences of *ndhF*. *Am J Bot.* 1999; 86: 1474–1486.
8. Nyffeler R, Clemens Bayer, Alverson WS, Yen A, Whitlock BA, Chase MW, et al. Phylogenetic analysis of the Malvadendrina clade (Malvaceae s.l.) based on plastid DNA sequences. *Org Divers Evol.* 2005, 5(2): 109–123
9. Fineschi S, Salvini D, Turchini D, Carnevale S, Vendramin GG. Chloroplast DNA variation of *Tilia cordata* (Tiliaceae). *Can J For Res.* 2003; 33: 2503–2508.
10. Liesebach H, Sinkó Z. A contribution to the systematics of the genus *Tilia* with respect to some hybrids by RAPD analysis. *Dendrobiology.* 2008; 59: 13–22
11. Filiz E, Birbilener S, Ozyigit II, Kulac S, Oruc FCS. Assessment of genetic variations of silver lime (*Tilia tomentosa* Moench.) by RAPD markers in urban and forest ecosystems, *Biotechnol Biotechnol Equip.* 2015; 29(4): 631–636
12. Hosseinzadeh Colagar A, Yusefi M, Zarei M, Yousefzadeh A. Assessment of genetic diversity of *Tilia rubra* DC. by RAPD analysis in the Hyrcanian forests, North of Iran. *Pol J Ecol.* 2013; 61: 341–348
13. Logan SA, Phuekvilai P, Wolff K. Ancient woodlands in the limelight: delineation and genetic structure of ancient woodland species *Tilia cordata* and *Tilia platyphyllos* (Tiliaceae) in the UK. *Tree Genet. Genomes.* 2015; 11: 52
14. McCarthy D. Systematics and phylogeography of the genus *Tilia* in North America. PhD Thesis, University of Illinois. 2012.
15. Yousefzadeh H, Hosseinzadeh Colagar A, Tabari M, Sattarian A, Assadi M. Utility of the ITS region sequence and structure for molecular identification of the *Tilia* species from Hyrcanian forest, Iran. *Plant Syst Evol.* 2012; 298: 947–961
16. Phuekvilai P. Relicts, Refugia and Reticulation: A study of population history, hybrids and phylogeny in the long-lived flowering tree genus *Tilia*. PhD Thesis, Newcastle University. 2014.
17. Zhang YJ, Ma PF, Li DZ. High-throughput sequencing of six bamboo chloroplast genomes: phylogenetic implications for temperate woody bamboos (Poaceae: Bambusoideae). *PLoS ONE.* 2011; 6(5): e20596. doi: [10.1371/journal.pone.0020596](https://doi.org/10.1371/journal.pone.0020596) PMID: [21655229](https://pubmed.ncbi.nlm.nih.gov/21655229/)
18. Ruhsam M, Rai HS, Mathews S, Ross TG, Graham SW, Raubeson LA, et al. Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in *Araucaria*? *Mol Ecol Resour.* 2015. doi: [10.1111/1755-0998.12375](https://doi.org/10.1111/1755-0998.12375)

19. Qiao J, Cai M, Yan G, Wang N, Li F, Chen B, et al. High-throughput multiplex cpDNA resequencing clarifies the genetic diversity and genetic relationships among *Brassica napus*, *Brassica rapa* and *Brassica oleracea*. *Plant Biotechnol J*. 2015; doi: [10.1111/pbi.12395](https://doi.org/10.1111/pbi.12395)
20. Parks M, Cronn R, Liston A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol*. 2009; 7: 84 doi: [10.1186/1741-7007-7-84](https://doi.org/10.1186/1741-7007-7-84) PMID: [19954512](https://pubmed.ncbi.nlm.nih.gov/19954512/)
21. Ma PF, Zhang YX, Zeng CX, Guo ZH, Li DZ. Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable bamboo tribe Arundinarieae (Poaceae). *Syst Biol*. 2014. 63(6): 933–950. doi: [10.1093/sysbio/syu054](https://doi.org/10.1093/sysbio/syu054) PMID: [25092479](https://pubmed.ncbi.nlm.nih.gov/25092479/)
22. Carbonell-Caballero J, Alonso R, Ibañez V, Terol J, Talon M, Dopazo J. A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus *Citrus*. *Mol Biol Evol*. 2015; 32(8): 2015–2035 doi: [10.1093/molbev/msv082](https://doi.org/10.1093/molbev/msv082) PMID: [25873589](https://pubmed.ncbi.nlm.nih.gov/25873589/)
23. Raubeson LA, Jansen RK. Chloroplast genomes of plants. In: Henry RJ, editor. *Diversity and evolution of plants-genotypic and phenotypic variation in higher plants*. Wallingford (UK): CABI publishing. 2005. pp. 45–68.
24. Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol*. 2011; 76: 273–297. doi: [10.1007/s11103-011-9762-4](https://doi.org/10.1007/s11103-011-9762-4) PMID: [21424877](https://pubmed.ncbi.nlm.nih.gov/21424877/)
25. Bendich AJ. Circular chloroplast chromosomes: the grand illusion. *Plant Cell*. 2004; 16: 1661–1666 PMID: [15235123](https://pubmed.ncbi.nlm.nih.gov/15235123/)
26. Lee SB, Kaitanis C, Jansen RK, Hostetler JB, Tallon LJ, Town CD, et al. The complete chloroplast genome sequence of *Gossypium hirsutum*: organization and phylogenetic relationships to other angiosperms. *BMC Genomics*. 2006; 7: 61. PMID: [16553962](https://pubmed.ncbi.nlm.nih.gov/16553962/)
27. Xu Q, Xiong G, Li P, He F, Huang Y, Wang K, et al. Analysis of complete nucleotide sequences of 12 *Gossypium* chloroplast genomes: origin and evolution of Allotetraploids. *PLoS ONE*. 2012; 7(8): e37128. doi: [10.1371/journal.pone.0037128](https://doi.org/10.1371/journal.pone.0037128) PMID: [22876273](https://pubmed.ncbi.nlm.nih.gov/22876273/)
28. Jansen RK, Saski C, Lee SB, Hansen AK, Daniell H. Complete plastid genome sequences of three Rosids (*Castanea*, *Prunus*, *Theobroma*): evidence for at least two independent transfers of *rp22* to the nucleus. *Mol Biol Evol*. 2011; 28:835–847. doi: [10.1093/molbev/msq261](https://doi.org/10.1093/molbev/msq261) PMID: [20935065](https://pubmed.ncbi.nlm.nih.gov/20935065/)
29. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bul*. 1987; 19: 11–15.
30. Yang JB, Li DZ, Li HT. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Mol Ecol Resour*. 2014; 14: 1024–1031. doi: [10.1111/1755-0998.12251](https://doi.org/10.1111/1755-0998.12251) PMID: [24620934](https://pubmed.ncbi.nlm.nih.gov/24620934/)
31. Patel RK, Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS ONE*. 2012; 7: e30619. doi: [10.1371/journal.pone.0030619](https://doi.org/10.1371/journal.pone.0030619) PMID: [22312429](https://pubmed.ncbi.nlm.nih.gov/22312429/)
32. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*. 2004; 20(17): 3252–3255 PMID: [15180927](https://pubmed.ncbi.nlm.nih.gov/15180927/)
33. Lohse M, Drechsel O, Kahlau S, Bock R. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res*. 2013; 41(Web Server issue): W575–581. doi: [10.1093/nar/gkt289](https://doi.org/10.1093/nar/gkt289) PMID: [23609545](https://pubmed.ncbi.nlm.nih.gov/23609545/)
34. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002; 30(14): 3059–3066 PMID: [12136088](https://pubmed.ncbi.nlm.nih.gov/12136088/)
35. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011; 28(10): 2731–2739. doi: [10.1093/molbev/msr121](https://doi.org/10.1093/molbev/msr121) PMID: [21546353](https://pubmed.ncbi.nlm.nih.gov/21546353/)
36. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014; 30(9):1312–1313. doi: [10.1093/bioinformatics/btu033](https://doi.org/10.1093/bioinformatics/btu033) PMID: [24451623](https://pubmed.ncbi.nlm.nih.gov/24451623/)
37. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 2012; 61(3): 539–542. doi: [10.1093/sysbio/sys029](https://doi.org/10.1093/sysbio/sys029) PMID: [22357727](https://pubmed.ncbi.nlm.nih.gov/22357727/)
38. Tajima F. Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics*. 1993; 135: 599–607 PMID: [8244016](https://pubmed.ncbi.nlm.nih.gov/8244016/)
39. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res*. 2004; 32(Web Server issue): W273–W279. PMID: [15215394](https://pubmed.ncbi.nlm.nih.gov/15215394/)
40. Saski C, Lee SB, Daniell H, Wood TC, Tomkins J, Kim HG, et al. Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes. *Plant Mol Biol*. 2005; 59: 309–322 PMID: [16247559](https://pubmed.ncbi.nlm.nih.gov/16247559/)

41. Davis JI, Soreng RJ. Migration of endpoints of two genes relative to boundaries between regions of the plastid genome in the grass family (Poaceae). *Am J Bot.* 2010; 97(5): 874–892. doi: [10.3732/ajb.0900228](https://doi.org/10.3732/ajb.0900228) PMID: [21622452](https://pubmed.ncbi.nlm.nih.gov/21622452/)
42. Sun Y-x, Moore MJ, Meng A-p, Soltis PS, Soltis DE, Li J-q, et al. Complete plastid genome sequencing of Trochodendraceae reveals a significant expansion of the inverted repeat and suggests a Paleogene divergence between the two extant species. *PLoS ONE.* 2013; 8(4): e60429. doi: [10.1371/journal.pone.0060429](https://doi.org/10.1371/journal.pone.0060429) PMID: [23577110](https://pubmed.ncbi.nlm.nih.gov/23577110/)
43. Palmer JD. Chloroplast DNA exists in two orientations. *Nature.* 1983; 301: 92–93
44. Martin G, Baurens FC, Cardi C, Aury JM, D'Hont A. The complete chloroplast genome of banana (*Musa acuminata*, Zingiberales): insight into plastid monocotyledon evolution. *PLoS One.* 2013; 8(6): e67350. doi: [10.1371/journal.pone.0067350](https://doi.org/10.1371/journal.pone.0067350) PMID: [23840670](https://pubmed.ncbi.nlm.nih.gov/23840670/)
45. Diekmann K, Hodkinson TR, Wolfe KH, van den Bekerom R, Dix PJ, Barth S. Complete chloroplast genome sequence of a major allogamous forage species, perennial ryegrass (*Lolium perenne* L.). *DNA Res.* 2009; 16(3):165–176. doi: [10.1093/dnares/dsp008](https://doi.org/10.1093/dnares/dsp008) PMID: [19414502](https://pubmed.ncbi.nlm.nih.gov/19414502/)
46. Greiner S, Wang X, Rauwolf U, Silber MV, Mayer K, Meurer J, et al. The complete nucleotide sequences of the five genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*: I. sequence evaluation and plastome evolution. *Nucleic Acids Res.* 2008; 36(7): 2366–2378. doi: [10.1093/nar/gkn081](https://doi.org/10.1093/nar/gkn081) PMID: [18299283](https://pubmed.ncbi.nlm.nih.gov/18299283/)
47. Kim HT, Kim KJ. Chloroplast genome differences between Asian and American *Equisetum arvense* (Equisetaceae) and the origin of the hypervariable *trnY-trnE* intergenic spacer. *PLoS ONE.* 2014; 9(8): e103898. doi: [10.1371/journal.pone.0103898](https://doi.org/10.1371/journal.pone.0103898) PMID: [25157804](https://pubmed.ncbi.nlm.nih.gov/25157804/)
48. Britten RJ, Rowen L, Williams J, Cameron RA. Majority of divergence between closely related DNA samples is due to indels. *Proc Natl Acad Sci USA.* 2003; 100(8): 4661–4665. PMID: [12672966](https://pubmed.ncbi.nlm.nih.gov/12672966/)
49. Anzai T, Shiina T, Kimura N, Yanagiya K, Kohara S, Shigenari A, et al. Comparative sequencing of human and chimpanzee MHC class I regions unveils insertions/deletions as the major path to genomic divergence. *Proc Natl Acad Sci USA.* 2003; 100(13): 7708–7713. PMID: [12799463](https://pubmed.ncbi.nlm.nih.gov/12799463/)
50. Ogurtsov AY, Sunyaev S, Kondrashov AS. Indel-based evolutionary distance and mouse-human divergence. *Genome Res.* 2004; 14(8): 1610–1616 PMID: [15289479](https://pubmed.ncbi.nlm.nih.gov/15289479/)
51. Levinson G, Gutman GA. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol.* 1987; 4(3): 203–221. PMID: [3328815](https://pubmed.ncbi.nlm.nih.gov/3328815/)