

Characterization and phylogenetic analysis of the complete chloroplast genome sequences of six species of *Wikstroemia* (Thymelaeaceae) revealed paraphyletic relationship to the monotypic genus *Stellera*.

Liefen He

Yunnan Normal University

Yonghong Zhang (✉ daphnecn@aliyun.com)

Yunnan Normal University

Shiou Yih Lee

Sun Yat-sen University

Research Article

Keywords: Bayesian inference, Daphneae, divergence hotspots, internal transcribed spacer region, sequence repeats

Posted Date: December 1st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-114011/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Wikstroemia (Thymelaeaceae) is a diverse genus, spanning both Asia and Australia and also recorded from the Hawaiian Islands. However, due to its medicinal properties and resource utilization in pulp production, genetic studies on these important species have been overshadowed. In this study, the chloroplast genome sequences of six species of Wikstroemia were sequenced and analyzed. The chloroplast genomes of the six species ranged between 172,610 bp (*W. micrantha*) and 173,697 bp (*W. alternifolia*), and exhibited a typical genome structure consisting of a pair of inverted repeat (IR) regions separated by a large single-copy (LSC) region and a small single-copy (SSC) region. The six chloroplast genomes were similar with a predicted 139 genes that consisted of 92 or 93 protein-coding genes, 38 tRNA genes, and 8 rRNA genes. The overall GC contents were identical (36.7%). Genome comparative analyses were conducted with the inclusion of two additional published species of Wikstroemia in which the sequence divergence and expansion of IRs in the chloroplast genomes were determined. The phylogenetic analysis inferred that Wikstroemia in its current circumscription is paraphyletic to *Stellera chamaejasme*, while the ITS-based tree analyses could not properly resolve the phylogenetic relationship between *Stellera* and Wikstroemia. This finding kindled interest in the proposal to synonymize *Stellera* with Wikstroemia, which was previously brought up, but rejected due to taxonomic conflicts. Nevertheless, this study provides valuable genomic information to aid in the taxonomic implications and phylogenomic reconstruction of Thymelaeaceae.

Introduction

Wikstroemia Engl., consisting of about 70 species, is a diverse genus in the family Thymelaeaceae. Members of *Wikstroemia* are widely distributed in the Asian and Oceanian regions, and also scattered around the Hawaiian Islands[1]. The species are mostly fibrous trees, shrubs or subshrubs with a woody rhizome. Several species are cultivated as raw material for pulp production[2,3], while a handful of them are reported to have medicinal properties[4,5]. However, studies on *Wikstroemia* have been confined to its utilization in pulp production and pharmacological applications; reports on genetic studies of *Wikstroemia* are scarce.

The only report on the genetic diversity to date is of *Wikstroemia ganpi* in Korea using inter simple sequence repeat (ISSR) markers[6] and two, published, complete chloroplast genome sequences of *Wikstroemia chamaedaphne* and *Wikstroemia indica*[7,8]. Due to the lack of molecular evidence, taxonomic studies on *Wikstroemia* have relied solely on morphological characteristics[9]. Ironically, the continuous nature of morphological variation in members of *Wikstroemia* has led to much taxonomic confusion in attempts to distinguish species and has resulted in ambiguities in taxonomic classifications between *Wikstroemia* and its sister genera[9,10]. Furthermore, the difficulty in detecting natural hybridization among the species of *Wikstroemia* due to the possibilities in low reproductive isolation and high genetic similarities suggested that *Wikstroemia* is a large species complex[11].

Chloroplasts are intracellular organelles in plants that contain the entire machinery necessary for the process of photosynthesis[12]. They provide energy through photosynthesis and play an important role in carbon uptake[13]. The chloroplast genome is a circular double-stranded DNA molecule. In plants, the cp genome is specifically maternally inherited and not disturbed by gene recombination[14]. A typical plant cp genome ranges in size from 120 kb to 217 kb[15]. The complete chloroplast genome has a typical quadripartite structure, including a large single-copy (LSC) region, a small single-copy (SSC) region, and two separate inverted regions (IRs)[16]. Due to the ease of availability and advances in next-generation sequencing, the cost of sequencing an organelle genome has become affordable[17]. The trend in plant chloroplast genome sequencing has become increasingly popular, thus contributing to studies of taxonomic classification and molecular evolution. Owing to its slower evolution rate and ease of sequencing and assembly due to its considerably smaller size, the chloroplast genome has been receiving much attention among biologists and taxonomists in exploring possibilities of species identification, species delimitation, phylogenomics, and molecular ecology of plants with challenging taxonomic problems or unique lifestyles[18,19].

The taxonomical placement of *Wikstroemia* has been controversial. It has experienced a complicated classification process when reviewing members of the Thymelaeaceae. However, *Stellera chamaejasme* of the monotypic genus *Stellera* was reported to be sister to *Wikstroemia* based on combined chloroplast DNA sequences (*trnT-trnL*, *trnL-trnF*, *trnL* intron, and *rp16* intron)[20], while *Wikstroemia*, along with the other 14 sister genera based on palynology findings, is taxonomically grouped under the Daphne group of the tribe Daphneae[21,22]. Although the phylogenetic work on members of Thymelaeaceae is actively ongoing[23], phylogenetic relationships in *Wikstroemia* are likely to be understudied. Constituent genera in Thymelaeaceae have experienced similar molecular challenges in which poor phylogenetic resolution is likely to be caused by low genetic variation in selected molecular markers[23]. Such conflicts can be overcome by utilizing genome-scale data sets[24]. At the same time, high divergence regions could be identified through genome comparisons, which could aid in future phylogenetic studies of such a diverse genus as *Wikstroemia*.

In this study, we sequenced the complete chloroplast genomes of six species of *Wikstroemia*, *W. alternifolia*, *W. canescens*, *W. capitata*, *W. dolicantha*, *W. micrantha*, and *W. scytophylla*, to analyze and compare genomes using bioinformatic tools. Our aims were to: (1) characterize the chloroplast genomes of the six species of *Wikstroemia*; (2) examine variations in sequence repeats and codon usage in the six chloroplast genome sequences; (3) identify highly divergent regions in the chloroplast genome sequences; (4) improve the understanding of the intrageneric/intergeneric phylogeny of *Wikstroemia* within the Thymelaeaceae based on chloroplast genome sequences and the nuclear ribosomal DNA internal transcribed spacer (ITS) region.

Results

Chloroplast genome features of six species of *Wikstroemia*

The total length of the chloroplast genomes of the six species of *Wikstroemia* analyzed in this study ranged from 172,610 bp (*W. micrantha*) to 173,697 bp (*W. alternifolia*). All six chloroplast genomes exhibited the typical quadripartite structure (Figure 1) consisting of a pair of IRs regions (41,850–42,073 bp) separated by an LSC region (86,111–86,7017 bp) and an SSC region (2,799–2,871 bp). All six chloroplast genomes had the same 36.7% GC content. However, the GC content in the chloroplast genome of each species of *Wikstroemia* was unevenly distributed. The IR region accounted for the highest GC content (38.8–38.9%), followed by the LSC region (34.7–34.9%), while the SSC region was recorded as having the lowest GC content (26.9–29.5%).

The six chloroplast genomes of *Wikstroemia* displayed identical gene content, gene order and no structural reconfigurations. A total of 139 genes were predicted in six species used in this study, comprising 92 or 93 protein-coding genes, 38 tRNA genes, and 8 rRNA genes (Table 1). However, 28 genes were duplicated in the IR regions, including 16 protein-coding genes (*ccsA*, *ndhA*, *ndhB*, *ndhD*, *ndhE*, *ndhH*, *ndhG*, *ndhI*, *psaC*, *rp12*, *rp123*, *rps7*, *rps12*, *rps15*, *ycf1*, *ycf2*), eight tRNA genes (*trnA*-UGC, *trnI*-CAU, *trnI*-GAU, *trnL*-CAA, *trnL*-UAG, *trnN*-GUU, *trnR*-ACG and *trnV*-GAC) and four rRNAs (*rnm4.5*, *rnm5*, *rnm16*, and *rnm23*) (Table 2). A total of 15 genes were found to contain an intron, with five of them (*ndhB*, *rp12*, *trnA*-UGC and *trnI*-GAU) located in the IR region and the remaining 10 genes (*atpF*, *petB*, *petD*, *rp116*, *rpoC1*, *rps16*, *trnG*-UCC, *trnL*-UAA, *trnK*-UUU and *trnV*-UAC) located in the LSC region (Table S1). However, only the *ycf3* gene, which was present in the LSC region, was detected to contain a pair of introns. Upon comparison, we found that the *trnK*-UUU gene had the longest intron, ranging from 2,498–2,508 bp, in all six genomes.

Repetitive sequence analysis

The total number of SSRs in the chloroplast genome sequences of *W. alternifolia*, *W. canescens*, *W. capitata*, *W. dolicantha*, *W. micrantha*, and *W. scytophylla* were 127, 128, 109, 87, 90, and 110, respectively (Figure 2a). However, no hexanucleotides were detected in the chloroplast genome sequences of *W. alternifolia*, *W. canescens* and *W. scytophylla*. The majority of SSRs (*W. alternifolia*: 70.79%; *W. canescens*: 70.31%; *W. capitata*: 68.81%; *W. dolicantha*: 63.22%; *W. scytophylla*: 63.64%; *W. micrantha*: 61.11%) were located in the LSC regions rather than in the other two regions of the chloroplast genome (Figure 2b).

All six species of *Wikstroemia* contained the same number of long repeats (Figure 3a). In general, all of them contained 24 forward repeats and 25 palindromic repeats, except for *W. canescens* and *W. capitata*. Long forward repeats that ranged between 30 and 40 bp were found most abundant in *W. dolicantha* and *W. micrantha*; while *W. alternifolia*, *W. canescens*, *W. capitata*, and *W. scytophylla* were recorded with higher number of long forward repeats with the lengths of 41 to 60 bp (Figure 3b). Long palindromic repeats were equally abundant in *W. alternifolia* and *W. canescens*, ranging from 40 to 60 bp and above 60 bp (Figure 3c), while long palindromic repeats were abundant in the range of 30 to 60 bp in *W. capitata*, *W. dolicantha*, *W. micrantha*, and *W. scytophylla*. Long reverse repeats were only detected in *W. canescens* and *W. capitata*, and mostly occurred within the range of 30 to 40 bp (Figure 3d).

Analysis of Codon usage

A total of 30 preferred codons (RSCU >1.00) were recorded each in *W. alternifolia*, *W. canescens*, *W. capitata*, *W. dolicantha*, *W. micrantha*, and *W. scytophylla*, in which 10, 10, 11, 11, 11, and 10 were with low preferences (1.00 <RSCU <1.30); 10, 10, 9, 9, 9, and 9 were with moderate preferences (1.30 ≤RSCU ≤1.50); and 12, 11, 13, 12, 12, and 12 were with strong preferences (RSCU>1.50), respectively (Table S2). The stop codon, UAA, was recorded to be most abundant and more preferred when compared to the other two stop codons, UAG and UGA, in all six species. Preferred codons mostly ending with amino acids A or U, except for the leucine (Leu)-encoded codon, UUG. The Leu-encoded codons accounted the highest occurrence (9.38%), while the cysteine (Cys)-encoded codons were recorded as having the least occurrence (3.13%) among all six *Wikstroemia* species.

Sequence divergence analysis

The chloroplast genome sequence alignment of eight species of *Wikstroemia*, using the *W. chamaedaphne* chloroplast genome as the reference, indicated high sequence conservatism across the chloroplast genomes of eight species of *Wikstroemia*, but not in the chloroplast genome of *W. indica* (Figure 4). As a whole, the size and gene order of the chloroplast genomes in *Wikstroemia* are well-conserved, but a distinct large gap was observed beginning within the *ycf1* gene sequence of the IRa to the 5' region of the *trnL*-UAG in the IRb of *W. indica*. Both of the single-copy regions were recorded as having greater sequence divergence than the IR region (Figure 5). With a Pi-value cut off point of 0.025, eight highly variable gene regions were identified; the *ndhD*-*ndhF*, *ndhF*-*rp132*, *ndhJ*, *petL*-*petG*, *psbI*-*trnS*-GCU, *trnG*-UCC, *trnK*-UUU-*rps16* and the *trnL*-UAA-*trnF*-GAA intergenic spacer regions. Six of these highly variable regions were located in the LSC, while two of them were in the SSC region.

Contraction and expansion in IR region

The genes adjacent to the IR borders were consistent across members of *Wikstroemia*, except in *W. indica*, which varied for its adjacent genes at the IRb/SSC (JSB) and IRa/SSC (JSA) border (Figure 6). Instead of the presence of *rp132* and *ndhF* genes in the SSC region, adjacent to JSB and JSA respectively, the *ycf1* gene was located across both the JSA and JSB in the chloroplast genome of *W. indica*. The *trnL*-UAG gene was also placed adjacent to the JSA, in the SSC region of the *W. indica* chloroplast genome. On the other hand, six species (*W. alternifolia*, *W. chamaedaphne*, *W. dolicantha*, *W. indica*, *W. micrantha*, and *W. scytophylla*) had their *rps19* gene crossing the IRa/LSC (JLA) border.

Phylogenetic analysis

The ML and BI trees based on the complete chloroplast genome sequences revealed that all the branch nodes for eight species of *Wikstroemia* included in the phylogenetic tree were supported with high bootstrap values and Bayesian posterior probabilities (ML: ≥90%; BI: ≥ 95%) (Figure 7). In addition, it was suggested a paraphyletic relationship was present in the genus *Wikstroemia*. Two species, *W. alternifolia* and *W. canescens*, were clustered with *Stellera chamaejasme*, while six species of *Wikstroemia* (*W. capitata*, *W. chamaedaphne*, *W. dolicantha*, *W. indica*, *W. micrantha*, and *W. scytophylla*) formed a monophyletic group.

For the ITS sequences in the ML tree revealed a paraphyletic relationship between *Wikstroemia* and *S. chamaejasme*, while most of the branch nodes within the *Wikstroemia* clade were not highly supported (Figure 8a). Strong bootstrap supports were recorded for the sistership between *W. alternifolia* and *W. canescens*, and *W. micrantha* and *W. stenophylla*. Weakly supported sisterships were present between *W. dolicantha* and *W. scytophylla*, and *W. capitata* and *W. ligustrina*. Contrarily, the BI analysis displayed a monophyletic relationship within the *Wikstroemia* clade (Figure 8b). Similar to the ML tree, sisterships were

strongly supported between *W. alternifolia* and *W. canescens*; and *W. micrantha* and *W. stenophylla*, but not between *W. dolicantha* and *W. scytophylla*, and *W. capitata* and *W. ligustrina* in the BI tree.

Discussion

The plastid genomes of the species of *Wikstroemia* in this study were rather conserved, which is similar to other angiosperms[25]. The cp genome lengths of the six species of *Wikstroemia* did not vary much and had cp genome sizes similar to typical angiosperms[26]. The same number and contents of the genes were predicted in this study, suggesting that evolution of the gene sequences was consistent across the six species. Similar to most angiosperms, sequence repeats for A/T were detected to be more abundant than those of G/C in the *Wikstroemia* cp genomes and such probability was affected by the stability between the nucleotides A/T and G/C in the genome[27,28].

The expansion and contraction of the IR region are major evolutionary events that influence the length of the cp genomes[29]. Our study indicated that the contractions and expansions of the IR regions exhibited relatively stable patterns within *Wikstroemia*, with slight variation in which gene recombination between the repetitive sequence or poly-A structure and tRNA could be one of the reasons for the change in length in the IR region[30]. However, *W. indica* indicated dissimilarity in its IR borders, which differed from most angiosperms[31]. We suspect that the chloroplast genome IR contraction and expansion in *W. indica* is severe and may be due to extensive gene transfer and larger IR expansion due to the results of the double strand break repair mechanism[32-34]. Although *W. indica* had a smaller cp genome size (151,731 bp) when compared to other species of *Wikstroemia* sequenced in this study, a higher GC content was detected in its cp genome [8](37.4%). Upon comparison, we found that the chloroplast genome of *W. indica* had a shorter IR region and larger SSC region when compared to other species of *Wikstroemia*. Changes in the placement of IR the borders in the chloroplast genome of *W. indica* was likely due to the contraction of its IR region, causing a loss in number and content of the genes. Among the genes that were not found in *W. indica*, but present in other species of *Wikstroemia* were a pair of *ndhA*, *ndhG*, and *ndhI* that were supposed to be present in the IR region; genes such as *ccsA*, *ndhD*, *ndhE*, *ndhH*, *psaC*, *rps15*, and *trnL-UAG* that were commonly duplicated in the IR regions were reduced to only one copy, and were transferred to the SSC region; while the *ndhF* and *rp32* genes that were common genes in the SSC region were not detected. Therefore, it can be concluded that the contraction of the IR region that caused gene loss has contributed to the difference in cp genome content between *W. indica* and the other seven species of *Wikstroemia*.

Molecular evidence based on chloroplast genome sequences revealed a non-monophyletic relationship between the species of *Wikstroemia* due to *W. alternifolia* and *W. canescens* clustering with *S. chamaedaphne*. Information on the phylogenetic relationships of the species of *Wikstroemia* is scarce. Although taxonomic work is tedious in a genus with diverse species, continuous efforts among taxonomists working on the members of Thymelaeaceae have provided some insights on the taxonomic status of *Wikstroemia*. To provide better insight on the phylogenetic relationships at the nuclear level, we used ITS sequences to perform ML and BI analyses. Unlike phylogenomic tree analyses on the complete chloroplast genome sequences, low bootstrap support and Bayesian posterior probabilities were observed at species level in the genus *Wikstroemia*. However, the molecular placement of the species of *Wikstroemia* were identical in both the ML and BI trees, while the most distinct difference between both phylogenetic trees was the placement of *S. chamaejasme*. In the ML tree based on the ITS sequences, *S. chamaejasme* was clustered within the *Wikstroemia* clade; while *S. chamaejasme* was sister to *Wikstroemia* in the BI tree. The discordance between the chloroplast and nuclear phylogenies in this study may be due to phylogenetic sorting, convergence, unequal rates of evolution, long branch attraction, and introgression[35]. However, low branch node supports in both the ITS-based ML and BI trees suggested that either the inclusion of additional nuclear gene sequences, or the application of the restriction site-associated DNA sequence (RAD-Seq) technique that integrates up to 10% of the nuclear genome[36], could be helpful in resolving the phylogenetic relationship within *Wikstroemia*. Evidently, in this study, the use of a single nuclear gene sequence, i.e. ITS, which was thought to be useful in delimitation of many plants at species level[37], is insufficient for resolving the phylogenetic relationships between *Stellera* and *Wikstroemia*.

Today, members of *Wikstroemia* comprise species previously placed under *Capura* L., *Daphne* L., *Diplomorpha* Meisn., *Daphnimorpha* Nakai, *Lonicera* L., *Passerina* L., *Restella* Pobed., and *Stellera* L. [1,38]. Eventually, the monotypic genus *Stellera*, which exhibits highly similar morphological characteristics, had troubled some taxonomists when compared to *Wikstroemia*. At least five species were placed under *Stellera* before they were transferred to *Wikstroemia*; while many were transferred to allied genera, such as *Daphne*, *Diarthron* and *Thymelaea*, in the tribe Daphneae[38]. This is understandable as *Stellera* has a longer taxonomic history, i.e. back to 1747, when compared to other genera in the Daphneae. Further division of the tribe into several genera emphasized the identity of the species. As a result, the sole species of *Stellera*, *S. chamaejasme*, as the type species, is only species left in the genus. Based on the literatures, we found that *Wikstroemia* has an interesting nomenclatural history, in which two genera, *Diplomorpha* and *Daphnimorpha*, were synonymized and were excused. For *Stellera* to combination with *Wikstroemia* was proposed before by transferring the type species, *S. chamaejasme*, under the monotypic subgenus *Chamaejasme*[39,40]. However, the proposal was rejected as *Stellera* has priority over *Wikstroemia*[41] and based on the Rules of Nomenclature, the combination can only be accepted if *Stellera* is proposed as a nomen genus rejiciendum [42](nov. gen. rejic.). Therefore, we do not exclude the possibility that *Stellera* could be synonym to *Wikstroemia*, however, based on the phylogenetic trees in this study, *Wikstroemia* is polyphyletic and *Stellera* is unquestionably a sister to *Wikstroemia*.

A comprehensive history of *Wikstroemia* was well-reviewed[43]. *Wikstroemia* was named after the Swedish botanist Johan Emanuel Wikström to commemorate his contributions in botany. However, the use of his name, *Wikstroemia*, was first proposed in 1821 by Heinrich Schrader for a genus in the family Theaceae[44]. In the same year, Kurt Sprengel described a new species with a new genus name, *Wikströmia glandulosa*, which was later corrected to a species of *Eupatorium*[43]. Neither of those names applied to a genus in the Thymelaeaceae. Eventually, members of Thymelaeaceae were once grouped under the genus *Capura*, which was established since 1771. Yet, in 1833, Endlicher proposed the genus name *Wickstroemia* for these species to honor Wikström, causing several nomenclature conflicts to occur. Thus the genus *Capura* was rejected and synonymized under *Wickstroemia* despite being earlier[45]. The spelling *Wickstroemia* was corrected to *Wikstroemia* in 1841, and is retained until today. However, changes in the genus name still occurred after that. Heller did not agree with the use of *Wikstroemia* Endl. to be practical and he treated it as a homonym of *Wikstroemia* Schrad[43]. As a consequence, a new genus, *Diplomorpha*, was established to replace *Wikstroemia* in Thymelaeaceae in 1841[46]. Unfortunately, the use of *Diplomorpha* did not last long as

members of *Diplomorpha* were either corrected to *Sauropus* or synonymized under *Wikstroemia*, including *W. canescens*[1]. Although many species of *Diplomorpha* were treated as unresolved due to incomplete description of their origin [38,47], based on their epithets, we speculate that these unresolved names could be members of *Wikstroemia*. It is noteworthy that the genus *Daphnimorpha* was similarly treated and is synonymized under *Wikstroemia* due to the high similarity in morphological characteristics when compared to the latter[22], while *W. australis* is still recognized as the type species for *Wikstroemia*.

Subgeneric classification in *Wikstroemia* was first proposed to include only two sections, *Euwikstroemia* and *Diplomorpha*[48]. The two sections were then included as subgenus *Wikstroemia* and subgenus *Diplomorpha*, along with the proposal of a new subgenus, *Chamaejasme*. However, subgenus *Chamaejasme* was then rejected due to taxonomic conflict [39-41]. Alongside the proposal of a novel genus, *Daphnimorpha*, which consisted of two species, *Daphnimorpha capitellata* (synonym *Diplomorpha capitellata*), and *D. kudoii* [49], the suggestion to retain subgenus *Diplomorpha* as an independent genus instead of a subgenus was brought up[50]. However, the suggestion was not heeded; *Daphnimorpha* was synonymized under *Wikstroemia*[22] and the subgeneric classification of *Wikstroemia*, consisting only of the subgenera *Wikstroemia* and *Diplomorpha*, is generally accepted currently[42,51].

One of the key morphological characteristics proposed to differentiate *Wikstroemia* from allied genera is the presence of petaloid scales in the flower [39]. However, the presence of disc in the flowers was not emphasized in *Wikstroemia*[10]. Owing to that, some species of *Daphne* may be misidentified as *Wikstroemia* due to incomplete morphological descriptions in both genera, causing an overlap in features of classification. A study conducted on the variations in morphological characteristics for members of Thymelaeaceae revealed that species of *Wikstroemia* and *Diplomorpha* have a scaly, subulate or clapper-shaped disc, while species of *Daphnimorpha* have a fan-shaped or half-cylindrical disc[49]. The three genera share only the same morphological feature of the petaloid scales instead of a disc, which occurs in *Daphne*, *Edgeworthia*, and *Eriosolena*. It was emphasized that discs and petaloid scales should not be treated as homologous structures, but to be considered as mutually exclusive in members of the Thymelaeaceae[52].

Conclusion

To the best of our knowledge, this study presents the first genome-scale analysis on species of *Wikstroemia*. The findings revealed high conservation of genes in the chloroplast genomes. The identification of highly variable gene regions in the chloroplast genome sequences of *Wikstroemia* could potentially be useful in resolving phylogenetic relationship in the genus. A strong sistership between *Wikstroemia* and the monotypic genus *Stellera* was present. The ML and BI trees based on the complete chloroplast genome sequences revealed that all the branch nodes for eight species of *Wikstroemia* included in the phylogenetic tree were supported with high bootstrap values and Bayesian posterior probabilities (ML: $\geq 90\%$; BI: $\geq 95\%$), while the ITS-based tree analyses could not properly resolve the phylogenetic relationship between *Stellera* and *Wikstroemia*. Nevertheless, the molecular data obtained in this study will serve as a valuable resource for providing greater insights into the taxonomy and phylogeny of Thymelaeaceae.

Materials And Methods

Plant Materials and DNA Extraction

Fresh leaf materials of six species of *Wikstroemia*, *W. alternifolia*, *W. canescens*, *W. capitata*, *W. dolicantha*, *W. micrantha* and *W. scytophylla*, were collected from botanical gardens and natural populations in China (Table 1). Voucher specimens were deposited in the Herbarium of Yunnan Normal University[53]. The total genomic DNA was extracted using the Axygen® AxyPrep Multisource Genomic Miniprep DNA kit (Corning, USA), following the manufacturer's protocol.

Chloroplast Genome Sequencing, Assembly and Annotation

A sequence library was constructed and sequencing was performed on an Illumina HiSeq 2500-PE150 platform (Illumina, USA). All raw reads were filtered using NGS QC Toolkit version 2.3.3 with default parameters to obtain clean reads[54]. The plastome was de novo assembled using NOVOPlasty[55] with the *rbcl* gene sequence of *Daphne kiusiana* (GenBank accession KY991380) as the seed sequence. Gene annotation was performed in Geneious Prime (Biomatters, New Zealand) using the complete chloroplast genome sequence of *W. chamaedaphne* (GenBank accession MN563132) as the reference genome. The circular physical map of the chloroplast genome was visualized using OGDRAW[56].

Repeat analyses

Short sequence repeats (SSRs) were identified using MISA-web[57], in which parameters for identification of perfect mono-, di-, tri-, tetra-, penta-, and hexanucleotide motifs were set for a minimum of 10, 5, 4, 3, 3, and 3 repeats, respectively. Long repeats, including forward, palindrome, reverse and complement repeats, were determined using REPuter[58] with a hamming distance of 3 and a minimal repeat size of 30 bp.

Codon Usage

Protein-coding sequences of each chloroplast genome were extracted and the relative synonymous codon usage (RSCU) was analyzed using MEGA 7[59].

Comparative genome and divergence analyses

The complete chloroplast genome sequences of two species of *Wikstroemia*, *W. chamaedaphne* (GenBank accession MN563132) and *W. indica* (GenBank accession MN453832), that were available in the NCBI GenBank, were downloaded and included in subsequent analyses. By using the chloroplast genome sequences of *W. chamaedaphne* as the reference genome, nucleotide variation in the chloroplast genome sequence alignment of the eight species of *Wikstroemia* were visualized using mVISTA[60] in Shuffle-LAGAN mode. To detect the expansion and contraction of the IR region in the chloroplast genomes across the eight species, the IR/SC boundaries of the chloroplast genomes were visualized using IRscope. To detect the mutational hotspots and divergence regions in the chloroplast genomes of the eight species, sequence alignment of the chloroplast genome sequences was carried out using Geneious Prime

(Biomatters, New Zealand). Calculations of the nucleotide variability (Pi) among the eight chloroplast genomes were performed using DnaSP v5[61] with window length of 1,000 bp and a step size of 500 bp.

Polymerase chain reaction and Sanger sequencing

Polymerase chain reaction (PCR) amplification was carried out in a 20 µL volume reaction using the ITS universal primer set: 5F: 5'-GGAAGTAAAAGTCGTAA-CAAGG-3' (forward) and 4R: 5'-TCCTCCGCTTATTGATATGC-3' (reverse). The PCR reactions for the nuclear ribosomal DNA ITS region contained 10 µL of 2× Taq PCR Starmix with loading dye (Genstar Biosolutions, China), 0.4 µM of each primer and 20 ng of genomic DNA as a template. PCR amplifications was conducted on a T100™ Thermal Cycler (Bio-Rad, USA), with an initial denaturation at 93°C for 5 min; 40 cycles of denaturation at 93°C for 30 s, annealing at 60°C for 30 s, extension at 72°C for 30 s; and a final extension at 72°C for 5 min. PCR products were sent for direct Sanger sequencing at both ends using an ABI 3730 DNA Analyzer (Applied Biosystems, USA).

Phylogenetic Analyses

The phylogenetic analysis was conducted on 17 complete chloroplast genome sequences of the Thymelaeaceae. Two species, *Psidium guajava* (Myrtaceae; GenBank accession KY635879) and *Gossypium gossypoides* (Malvaceae; GenBank accession HQ901195) were included as outgroups. Sequence alignment was carried out using MAFFT[62]. Maximum-likelihood (ML) tree was constructed using RAxML 8.2.11[63], in which the general-time-reversible (GTR) and gamma distributed (+G) (+GTR+G) DNA substitution model was selected and all branch nodes were calculated under 1,000 bootstrap replicates; while the Bayesian inference (BI) was conducted using MrBayes[64], in which the Markov Chain Monte Carlo (MCMC) was conducted with 2,000,000 generations and sampling was collected at every 100 cycles. Both tree analyses were conducted through the online programs available in the CIPRESS Science Gateway web portal[65]. The final tree results were visualized under FigTree[66].

The ITS sequences were aligned and manually trimmed for their primer sequences to obtain clean sequences. A total of 26 additional ITS sequences derived from members of the Thymelaeaceae were downloaded from the NCBI GenBank and MUSCLE-aligned against the ITS sequences of the six species of *Wikstroemia* used in this study using MEGA 7[59]. Two species, *P. guajava* (Myrtaceae; GenBank accession MN295360) and *Gossypium australe* (Malvaceae; GenBank accession AF057763), were included as outgroups. The alignment was trimmed using trimAL v1.2[67] with the gappout method in order to reduce the systematic errors produced by poor alignment. The optimal DNA substitution model for the ML analysis using the “Find Best DNA/ Protein Model (ML)” function embedded in MEGA 7[59] was calculated to be the Kimura two-parameter (K2P) with discrete Gamma model (+G4) and invariant included (+I) (=K2P+G+I). ML analysis was performed using MEGA 7[59] with 1,000 bootstrap replicates. The BI analysis was conducted with the method previously described [64].

Declarations

Acknowledgements

This work was supported by the National Natural Science Foundation of China (31760048) and the Fundamental Research Funds for the Central Universities (33000-31611215).

Author contributions

Y.H.Z. performed the experiments. L.F.H. assembled the sequences and analyzed the data. L.F.H. wrote the manuscript. Y.H.Z. collected the plant material. Y.H.Z. and X.Y.L. conceived the research and revised the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

References

1. Wang, Y. & Gilbert, M. G. *Wikstroemia* in *Flora of China*. 215–229 (Science Press and Missouri Botanical Garden Press, 2007).
2. Yum, H., Singer, B. W. & Bacon, A. Coniferous wood pulp in traditional Korean paper between the 15th and 18th centuries AD. *Archaeometry* **51**, 467-479, doi:10.1111/j.1475-4754.2008.00448.x (2010).
3. Lin, L. D., Chang, F. C., Ko, C. H., Wang, C. Y. & Wang, Y. N. Properties of enzyme pretreated *Wikstroemia sikokiana* and *Broussonetia papyrifera* bast fiber pulps. *Bioresources* **10**, 3625-3637, doi:10.15376/biores.10.2.3625-3637 (2015).
4. Li, Y. M., Zhu, L., Jiang, J. G., Yang, L. & Wang, D. Y. Bioactive components and pharmacological action of *Wikstroemia indica* (L.) C. A. Mey and its clinical application. *Curr Pharm Biotechnol* **10**, 743-752, doi:10.2174/138920109789978748 (2009).
5. Helmi, H., Susanti, I., Agung, N. A. & Kusen, S. Antibacterial activity of belilik (*Brucea javanica* (L.) Merr) and benta (*Wikstroemia androsaemifolia* Decne) to inhibit the growth of enteropathogenic bacteria. *J Biol Res* **21**, 35-40, doi:10.23869/bphjbr.21.1.20154 (2016).
6. Kim, E.-H., Kim, J.-Y., Jung, J.-Y. & Son, S.-W. Impact of habitat damage on *Wikstroemia ganpi* (Siebold & Zucc.) Maxim. genetic diversity and structure. *J Agr Life Sci* **52**, 33-44, doi:10.14397/jals.2018.52.2.33 (2018).
7. Qian, S. J., Zhang, Y. H. & Li, G. D. The complete chloroplast genome of a medicinal plant, *Wikstroemia chamaedaphne* (Thymelaeaceae). *Mitochondrial DNA B* **5**, 648-649, doi:10.1080/23802359.2019.1711228 (2020).

8. Qian, S. J. & Zhang, Y. H. Characterization of the complete chloroplast genome of a medicinal plant, *Wikstroemia indica* (Thymelaeaceae). *Mitochondrial DNA B* **5**, 83-84, doi:10.1080/23802359.2019.1696249 (2020).
9. Mayer, S. S. Morphological Variation in Hawaiian *Wikstroemia* (Thymelaeaceae). *Syst Bot* **16**, 693-704, doi:10.2307/2418871 (1992).
10. Zhang, Y.-Z., Sun, W.-G., Jiang, X., Li, Z.-M. & Zhang, Y.-H. Numerical taxonomy of the genera *Daphne* and *Wikstroemia*. *Guihaia* **36**, 61-72, doi:10.11931/guihaia.gxzw201504020 (2016).
11. Mayer, S. S. Artificial Hybridization in Hawaiian *Wikstroemia* (Thymelaeaceae). *Am J Bot* **78**, 122-130, doi:10.1002/j.1537-2197.1991.tb12578.x (1991).
12. Sugiura, M. The chloroplast genome. *Plant Mol Bio* **19**, 149-168, doi:10.1007/BF00015612 (1992).
13. Zhang, T. *et al.* Comparative analysis of the complete chloroplast genome sequences of six species of Pulsatilla Miller, Ranunculaceae. *Chin Med* **14**, 53, doi:10.1186/s13020-019-0274-5 (2019).
14. Douglas, S. E. Plastid evolution: origins, diversity, trends. *Curr Opin Genet Dev* **8**, 655-661, doi:10.1016/S0959-437X(98)80033-6 (1998).
15. Xiong, Y., Xiong, Y., Jia, S. & Ma, X. The complete chloroplast genome sequencing and comparative analysis of Reed Canary Grass (*Phalaris arundinacea*) and Hardinggrass (*P. aquatica*). *Plants* **9**, 748, doi:10.3390/plants9060748 (2020).
16. Ravi, V., Khurana, J. P., Tyagi, A. K. & Khurana, P. An update on chloroplast genomes. *Plant Syst Evo* **271**, 101-122, doi:10.1007/s00606-007-0608-0 (2008).
17. Mader, M. *et al.* Complete chloroplast genome sequences of four meliaceae species and comparative analyses. *Int J Mol Sci* **19**, 701, doi:10.3390/ijms19030701 (2018).
18. Graham, S. W., Lam, V. K. Y. & Merckx, V. S. F. T. Plastomes on the edge: the evolutionary breakdown of mycoheterotroph plastid genomes. *New Phytol* **214**, 48-55, doi:10.1111/nph.14398 (2017).
19. Li, X. *et al.* Plant DNA barcoding: from gene to genome. *Biol Rev Camb Philos Soc* **90**, 157-166, doi:10.1111/brv.12104 (2015).
20. Zhang, Y.-H., Volis, S. & Sun, H. Chloroplast phylogeny and phylogeography of *Stellera chamaejasme* on the Qinghai-Tibet Plateau and in adjacent regions. *Mol Phylogenet Evo* **57**, 1162-1172, doi:10.1016/j.ympev.2010.08.033 (2010).
21. Herber, B. E. Pollen morphology of the Thymelaeaceae in relation to its taxonomy. *Plant Syst Evo* **232**, 107-121, doi:10.1007/s006060200030 (2002).
22. Herber, B. E. *Flowering Plants · Dicotyledons* (eds K. Kubitzki & C. Bayer) 373-396 (Springer, 2003).
23. Foster, C. S. P. *et al.* Molecular phylogenetics provides new insights into the systematics of Pimelea and Thecanthes (Thymelaeaceae). *Aust Syst Bot* **29**, 185-196, doi:10.1071/SB16013 (2016).
24. Foster, C. S. P., Henwood, M. J. & Ho, S. Y. W. Plastome sequences and exploration of tree-space help to resolve the phylogeny of riceflowers (Thymelaeaceae: *Pimelea*). *Mol Phylogenet Evo* **127**, 156-167, doi:10.1016/j.ympev.2018.05.018 (2018).
25. Cheon, K.-S., Kim, K.-A., Kwak, M., Lee, B. & Yoo, K.-O. The complete chloroplast genome sequences of four *Viola* species (Violaceae) and comparative analyses with its congeneric species. *Plos One* **14**, e0214162, doi:10.1371/journal.pone.0214162 (2019).
26. Li, W., Zhang, C., Guo, X., Liu, Q. & Wang, K. Complete chloroplast genome of *Camellia japonica* genome structures, comparative and phylogenetic analysis. *PLoS ONE* **14**, e0216645, doi:10.1371/journal.pone.0216645 (2019).
27. Yang, Y. *et al.* Complete Chloroplast Genome Sequence of Poisonous and Medicinal Plant *Datura stramonium*: Organizations and Implications for Genetic Engineering. *Plos One* **9**, e110656, doi:10.1371/journal.pone.0110656 (2014).
28. Raveendar, S. *et al.* The Complete Chloroplast Genome of *Capsicum annum* var. *glabriusculum* Using Illumina Sequencing. *Molecules* **20**, 13080-13088, doi:10.3390/molecules200713080 (2015).
29. Liu, X. *et al.* Complete Chloroplast Genome Sequence and Phylogenetic Analysis of *Quercus bawanglingensis* Huang, Li et Xing, a Vulnerable Oak Tree in China. *Forests* **10**, 0587, doi:10.3390/f10070587 (2019).
30. Plunkett, G. M. & Downie, S. R. Expansion and Contraction of the Chloroplast Inverted Repeat in Apiaceae Subfamily Apioideae. *Syst Bot* **25**, 648-667, doi:10.2307/2666726 (2000).
31. Raubeson, L. A. *et al.* Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics* **8**, 174, doi:10.1186/1471-2164-8-174 (2007).
32. Goulding, S. E., Olmstead, R. G., Morden, C. W. & Wolfe, K. H. Ebb and flow of the chloroplast inverted repeat. *Mol Gen Gener* **252**, 195-206, doi:10.1007/BF02173220 (1996).
33. Wang, R.-J. *et al.* Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evol Bio* **8**, 36, doi:10.1186/1471-2148-8-36 (2008).
34. Peery, R. M. *Understanding angiosperm genome interactions and evolution: insights from sacred lotus (Nelumbo nucifera) and the carrot family (Apiaceae)* Doctor Dissertation thesis, University of Illinois at Urbana-Champaign, (2015).
35. Soltis, D. E. & Kuzoff, R. K. Discordance between nuclear and chloroplast phylogenies in the Heuchera group (Saxifragaceae). *Evolution* **49**, 727-742, doi:10.2307/2410326 (1995).
36. Zimmer, E. A. & Wen, J. Using nuclear gene data for plant phylogenetics: Progress and prospects II. Next-gen approaches. *J Syst Evo* **53**, 371-379, doi:10.1111/jse.12174 (2015).
37. Poczai, P. & Hyvönen, J. Nuclear ribosomal spacer regions in plant phylogenetics: problems and prospects. *Mol Biol Rep* **37**, 1897-1912, doi:10.1007/s11033-009-9630-3 (2010).
38. Allkin, B. The Plant List 2013. <http://www.worldfloraonline.org>. (2020).
39. Domke, W. Zur Kenntnis einiger Thymelaeaceen. . *Notizblatt des Botanischen Gartens und Museums zu Berlin-Dahlem*, 348-363 (1932).

40. Domke, W. *Untersuchungen über die systematische und geographische Gliederung der Thymelaeaceen nebst einer Neubeschreibung ihrer Gattung.* (Schweizerbart Science Publishers, 1934).
41. Rehder, A. Notes on the ligneous plants described by Léveillé from eastern Asia. *J Arnold Arboretum***15**, 267-326, doi:<https://www.jstor.org/stable/43782260> (1934).
42. Hou, D. Thymelaeaceae. *Flora Malesiana - Series 1, Spermatophyta***6**, 1-48 (1960).
43. Peterson, B. Johan Emanuel Wikstrom, with Historical Notes on the Genus Wikstroemia. *Pac Sci***50**, 77-83, doi:<http://hdl.handle.net/10125/2606> (1996).
44. Kobuski, C. E. Studies in the theaceae, xvi bibliographical notes on the genus laplacea. *J Arnold Arboretum***28**, 435-438, doi:10.5962/bhl.part.25572 (1947).
45. Blake, S. F. New spermatophytes collected in Venezuela and Curaçao by messrs. Curran & Haman. *Contributions from the Gray Herbarium of Harvard University***53**, 30-55, doi:<https://www.jstor.org/stable/41764336> (1918).
46. Heller, A. Observations on the ferns and flowering plants of the Hawaiian Islands. *Minnesota Bot Stud***1**, 760-922 (1897).
47. Meyer, K. A. Bemerkungen über die Gattungen der Daphnaceen ohne perigynische Schuppen, nebst einer Charakteristik derselben. **1**, 354-359 (1843).
48. Meisner, C. F. in *Prodromus* Vol. 14, (ed A. DeCandolle) 493-605 (V.Masson, 1857).
49. Hamaya, T. A dendrological monograph of the Thymelaeaceae plants of Japan. *Bull Tokyo Univ Forest***50**, 45-96 (1955).
50. Hamaya, T. Dendrological Studies of the Japanese and Some Foreign Genera of the Thymelaeaceae: Anatomical and Phylogenetic Studies. *Report on forest exercise of Department of agriculture, Tokyo University*, 1-80 (1959).
51. Huang, S. Taxa nova Thymelaeacearum Sinicarum. *Acta Bot Yunnan***7**, 277-291 (1985).
52. Heinig, K. H. Studies in the Floral Morphology of the Thymelaeaceae. *Am J Bot***38**, 113-132, doi:10.1002/j.1537-2197.1951.tb14801.x (1951).
53. Thiers, B. E. Index Herbariorum: A global directory of public herbaria and associated staff. (2020).
54. Patel, R. K. & Jain, M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One***7**, e30619, doi:10.1371/journal.pone.0030619 (2012).
55. Dierckxsens, N., Mardulyn, P. & Smits, G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res***45**, e18, doi:10.1093/nar/gkw955 (2017).
56. 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***52**, 267-274, doi:10.1007/s00294-007-0161-y (2007).
57. Beier, S., Thiel, T., Münch, T., Scholz, U. & Mascher, M. MISA-web: a web server for microsatellite prediction. *Bioinformatics***33**, 2583–2585, doi:10.1093/bioinformatics/btx198 (2017).
58. Kurtz, S. *et al.* REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res***29**, 4633-4642, doi:10.1093/nar/29.22.4633 (2001).
59. Kumar, S., Stecher, G. & Tamura, K. Mega7: MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evo***33**, 1870-1874, doi:10.1093/molbev/msw054 (2016).
60. Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. & Dubchak, I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res***32**, W273-279, doi:10.1093/nar/gkh458 (2004).
61. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics***25**, 1451-1452, doi:10.1093/bioinformatics/btp187 (2009).
62. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evo***30**, 772-780, doi:10.1093/molbev/mst010 (2013).
63. Guindon, S. *et al.* New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst Bio***59**, 307-321, doi:10.1093/sysbio/syq010 (2010).
64. Ronquist, F. *et al.* MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst Bio***61**, 539-542, doi:10.1093/sysbio/sys029 (2012).
65. Miller, M. A., Pfeiffer, W. & Schwartz, T. in *2010 Gateway Computing Environments Workshop (GCE)* 1-8 (IEEE, New Orleans, LA, USA, 2010).
66. Rambaut, A. *FigTree v1.4*, <http://tree.bio.ed.ac.uk/software/figtree> (2018).
67. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics***25**, 1972-1973, doi:10.1093/bioinformatics/btp348 (2009).

Additional Information

Supplementary Materials

Table S1: Information of introns and exons for protein-coding genes in six species of *Wikstroemia* used in this study.

Table S2: Relative synonymous codon usage of protein-coding genes in chloroplast genomes of six species of *Wikstroemia* used in this study.

Tables

Table 1. Chloroplast genome features of six species of *Wikstroemia*

Species	Origin	Coordinates (longitude, latitude)	Chloroplast genome								Chloroplast genes		
			Total (bp)	GC content (%)	LSC (bp)	GC content (%)	SSC (bp)	GC content (%)	IR (bp)	GC content (%)	Total	CDS	tRNA
<i>Wikstroemia alternifolia</i>	Batang, Sichuan	29°19'27"N, 99°18'40"E	173,697	36.7	86,694	34.7	2,857	29.5	42,073	38.8	139	93	38
<i>Wikstroemia canescens</i>	Batang, Sichuan	29°19'27"N, 99°18'40"E	173,667	36.7	86,701	34.8	2,854	26.9	42,056	38.8	139	93	38
<i>Wikstroemia capitata</i>	Yin Tiao Ling Nature Reserve, Sichuan	31°28'02"N, 109°55'53"E	172,849	36.7	86,154	34.8	2,871	29.4	41,912	38.9	139	92	38
<i>Wikstroemia dolicantha</i>	Kunming, Yunnan	25°07'48"N, 102°42'24"E	172,804	36.7	86,230	34.8	2,854	28.7	41,860	38.9	139	92	38
<i>Wikstroemia micrantha</i>	Changshou, Chongqing	30°10'20"N, 30°10'20"E	172,610	36.7	86,111	34.9	2,799	29.5	41,850	38.9	139	93	38
<i>Wikstroemia scytophylla</i>	Kunming Botanical Garden	25°08'36"N, 102°44'27"E	173,2544	36.7	86,338	34.8	2,840	29.4	42,038	38.8	139	93	38

Table 2. Genes present in the chloroplast genomes of six species of *Wikstroemia* used in this study.

Genes	
RNAs, ribosomal	<i>rrn4.5</i> (×2), <i>rrn5</i> (×2), <i>rrn16</i> (×2), <i>rrn23</i> (×2)
RNAs, transfer	<i>trnA</i> -UGC(×2), <i>trnC</i> -GCA, <i>trnD</i> -GUC, <i>trnE</i> -UUC, <i>trnF</i> -GAA, <i>trnI</i> M-CAU, <i>trnG</i> -GCC, <i>trnG</i> -UCC, <i>trnH</i> -GUG, <i>trnI</i> -CAU(×2), <i>trnI</i> -GAU(×2), <i>trnK</i> -UUU, <i>trnL</i> -CAA(×2), <i>trnL</i> -UAA, <i>trnL</i> -UAG(×2), <i>trnM</i> -CAU, <i>trnN</i> -GUU(×2), <i>trnP</i> -UGG, <i>trnQ</i> -UUG, <i>trnR</i> -ACG(×2), <i>trnR</i> -UCU, <i>trnS</i> -GCU, <i>trnS</i> -GGA, <i>trnS</i> -UGA, <i>trnT</i> -GGU, <i>trnT</i> -UGU, <i>trnV</i> -GAC(×2), <i>trnV</i> -UAC, <i>trnW</i> -CCA, <i>trnY</i> -GUA
Transcription and splicing	<i>matK</i> , <i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> , <i>rpoC2</i>
Translation, ribosomal proteins	
Small subunit	<i>rps3</i> , <i>rps4</i> , <i>rps7</i> (×2), <i>rps8</i> , <i>rps11</i> , <i>rps12</i> (×2), <i>rps14</i> , <i>rps15</i> (×2), <i>rps16</i> , <i>rps18</i> , <i>rps19</i>
Large subunit	<i>rpl2</i> (×2), <i>rpl14</i> , <i>rpl16</i> , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> (×2), <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>
Photosynthesis	
ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> , <i>atpH</i> , <i>atpI</i>
Photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> (×2), <i>psal</i> , <i>psaJ</i>
Photosystem II	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbN</i> , <i>psbT</i> , <i>psbZ</i>
Calvin cycle	<i>rbcL</i>
Cytochrome complex	<i>petA</i> , <i>petB</i> , <i>petD</i> , <i>petG</i> , <i>petL</i> , <i>petN</i>
NADH dehydrogenase	<i>ndhA</i> (×2), <i>ndhB</i> (×2), <i>ndhC</i> , <i>ndhD</i> (×2), <i>ndhE</i> (×2), <i>ndhF</i> , <i>ndhG</i> (×2), <i>ndhH</i> (×2), <i>ndhI</i> (×2), <i>ndhJ</i> , <i>nahK</i>
Others	<i>accD</i> , <i>ccsA</i> (×2), <i>cemaA</i> , <i>ycf1</i> (×2), <i>ycf2</i> (×2), <i>ycf3</i> , <i>ycf4</i>

Figures

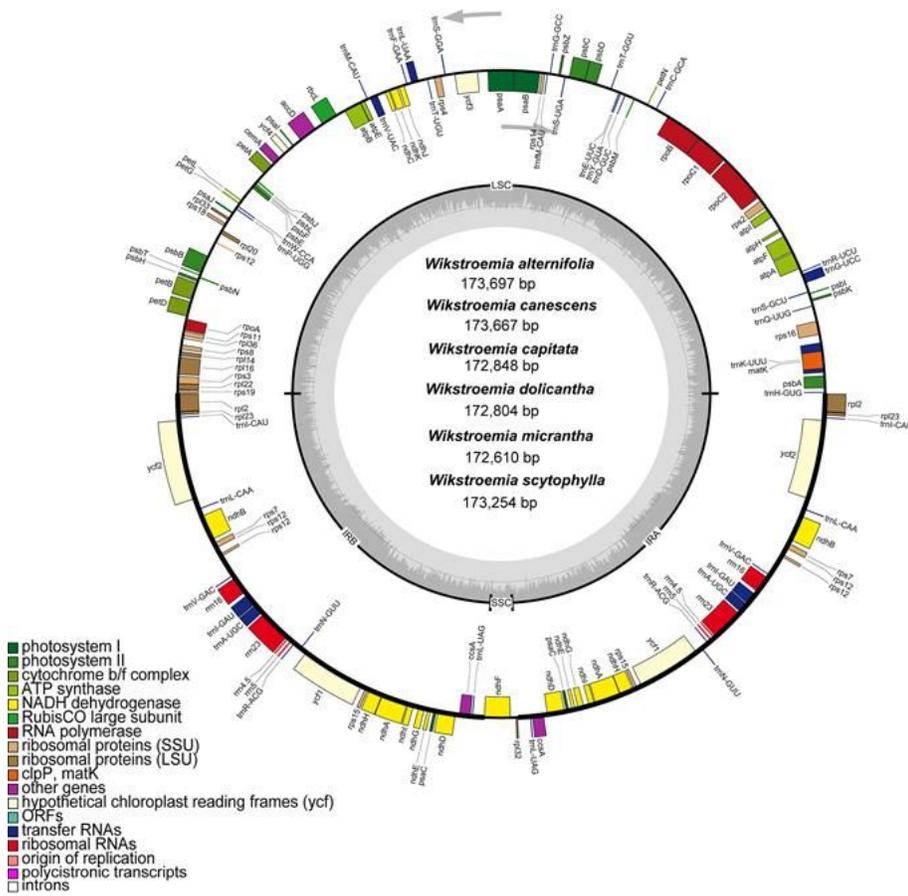


Figure 1
 Gene map for chloroplast genomes of six species of Wikstroemia used in this study. Genes on inside of map are transcribed in clockwise direction; genes on outside of map are transcribed in counterclockwise direction. Darker gray in inner circle represents to GC content whereas light gray corresponds to AT content. Different functional groups of genes are shown in different colors.

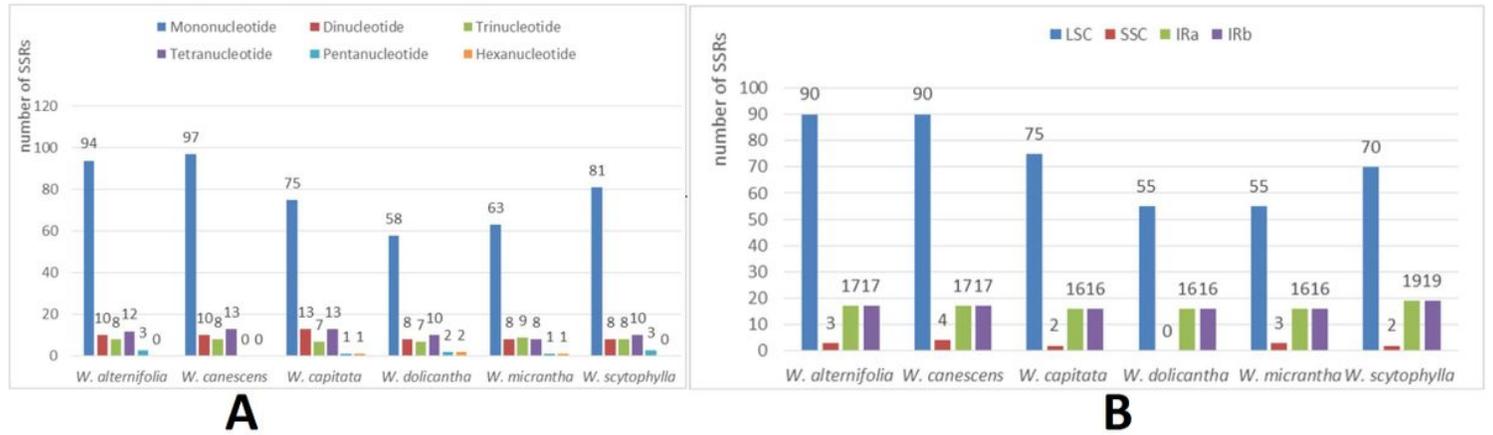


Figure 2
 Distribution of small sequence repeats (SSRs) in the chloroplast genomes of six accessions of Wikstroemia. (A) Number of different SSR types detected in six chloroplast genomes of Wikstroemia; (B) Frequencies of identified SSRs in large single-copy (LSC), small single-copy (SSC) and inverted repeats (IRs) regions.

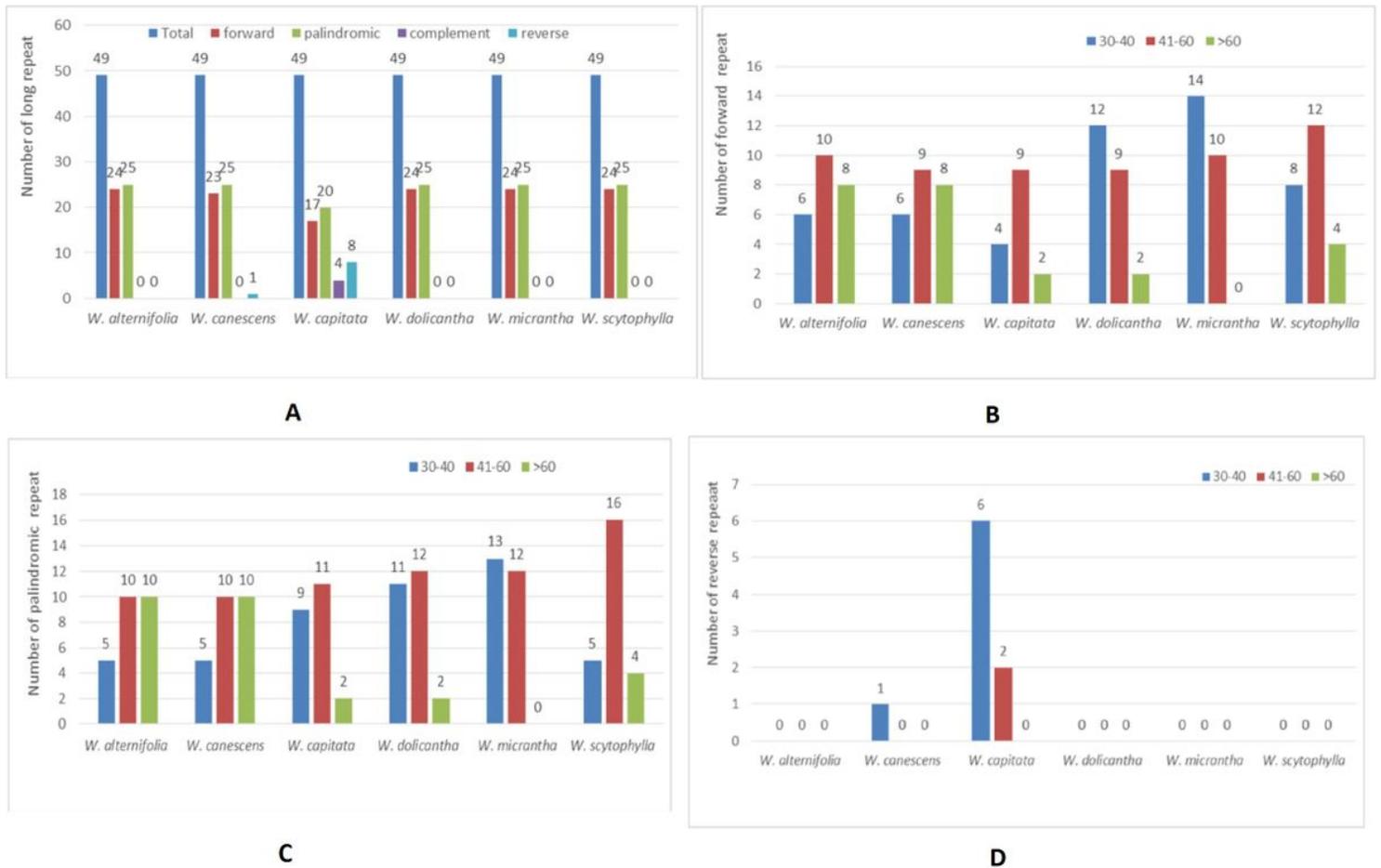


Figure 3

Analysis of long repeat sequences in six chloroplast genomes of Wikstroemia. (A) Quantities of long repeats based on types; (B) frequencies of forward repeats by length; (C) frequency of the palindromic repeats by length; and (D) frequency of reverse repeats by length.

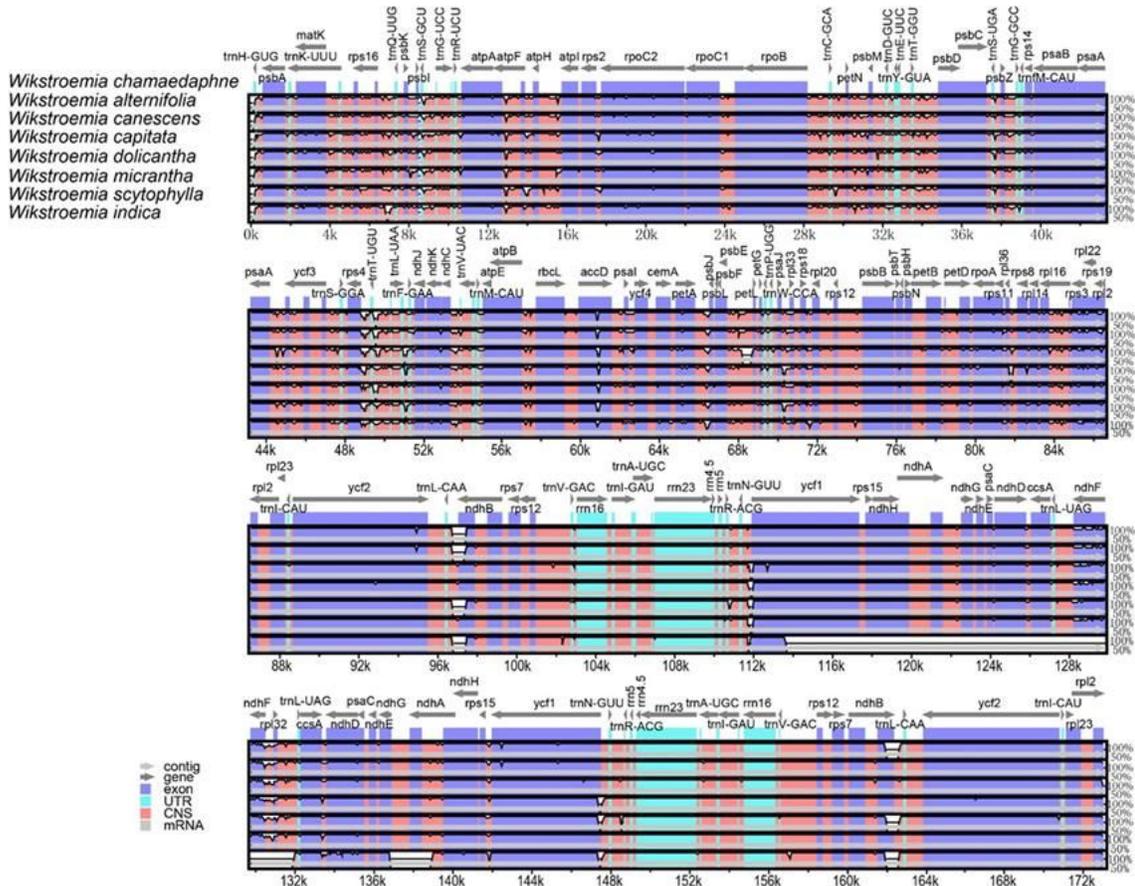


Figure 4

Complete chloroplast genome comparison of eight species of Wikstroemia using chloroplast genome of *W. chamaedaphne* as reference.

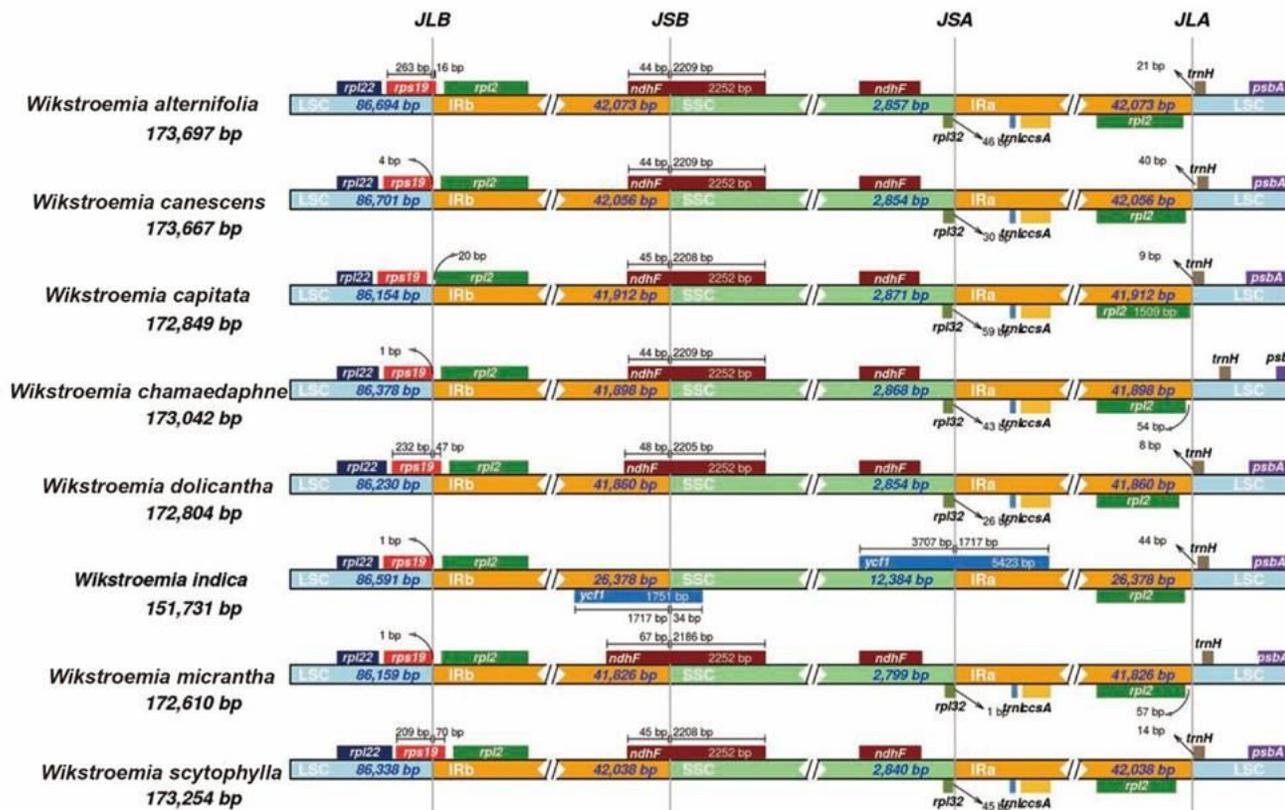


Figure 5

Sliding window analysis of complete chloroplast genome sequences among eight species of *Wikstroemia* (Window length: 1000 bp; step size: 500 bp).

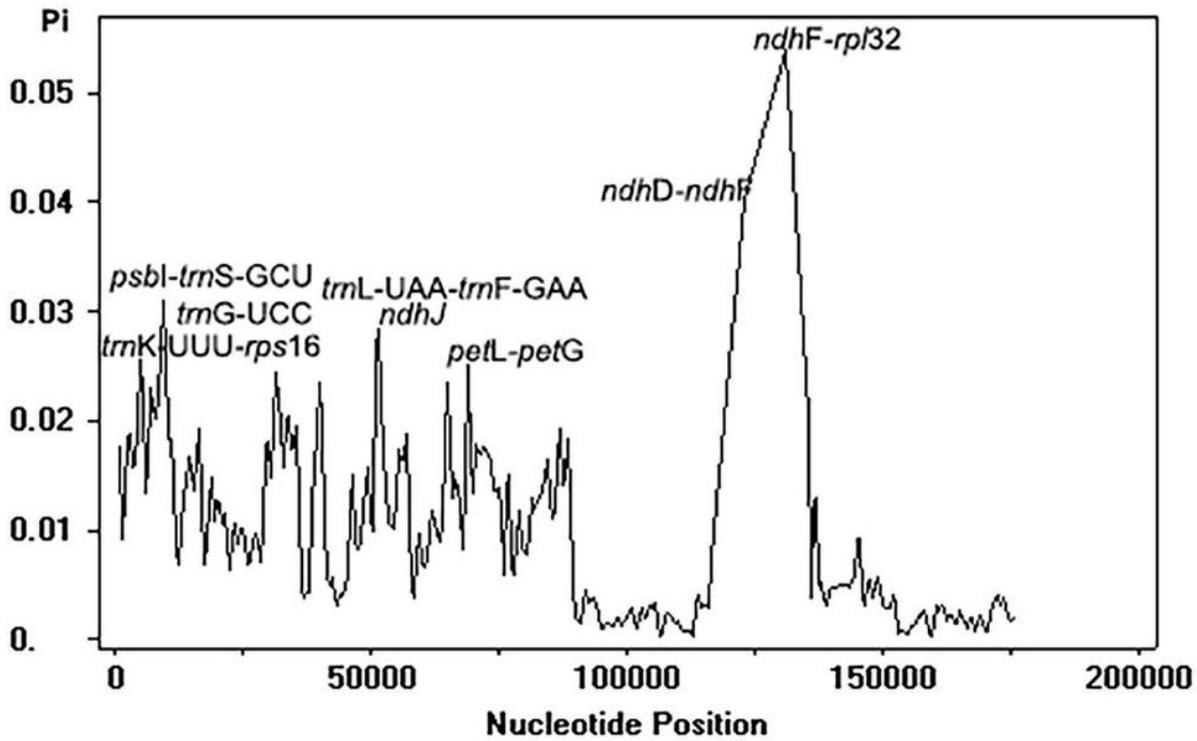


Figure 6

Comparison of borders between LSC, SSC and IR regions across eight chloroplast genomes of *Wikstroemia*.

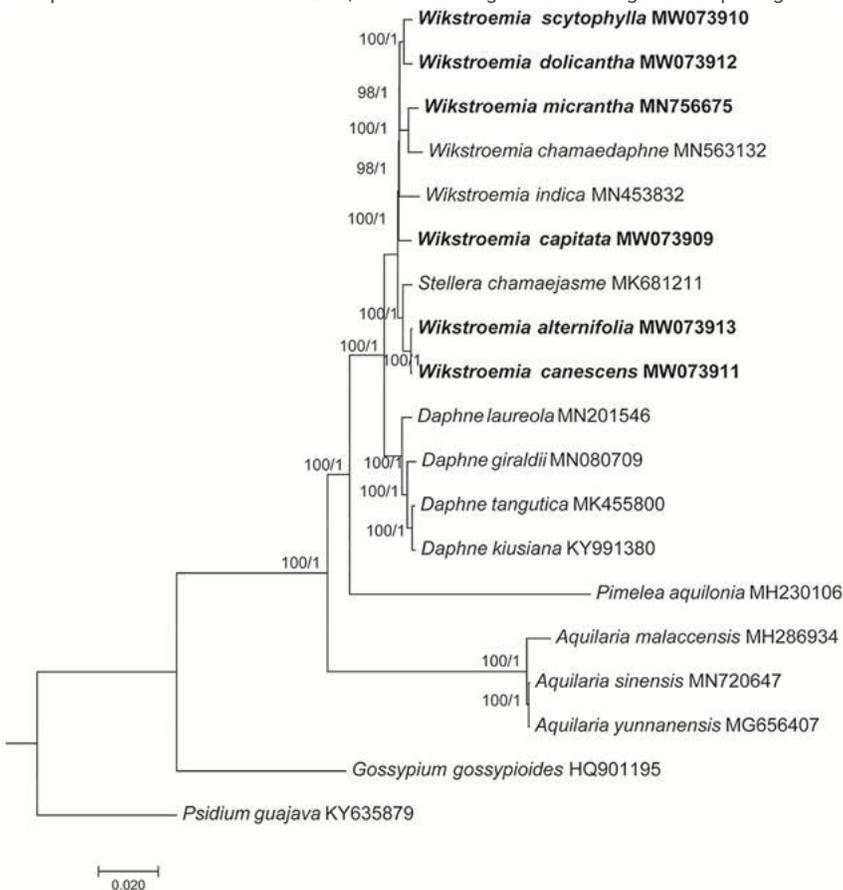


Figure 7

Maximum-likelihood (ML) and Bayesian inference (BI) of *Wikstroemia* and allied genera based on chloroplast genome sequences of 17 taxa representing 5 genera of Thymelaeaceae. Bootstrap support values and Bayesian posterior probabilities are indicated at each branch node. Sequences obtained through this study are indicated in bold; two species, *Psidium guajava* (KY635879) and *Gossypium gossypioides* (HQ901195) were included as outgroups.

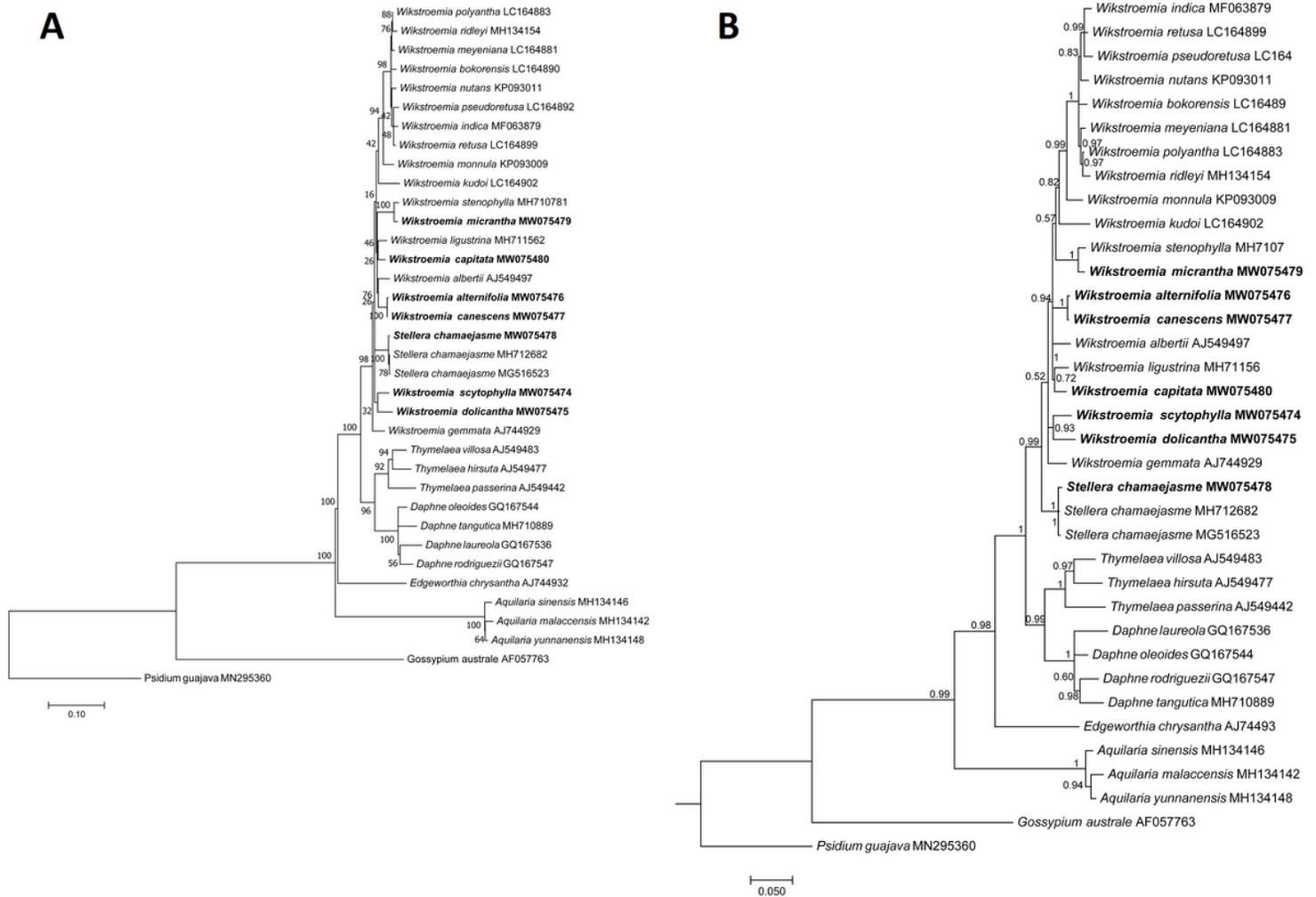


Figure 8

Phylogenetic analyses of Thymelaeaceae based on nuclear ribosomal DNA internal transcribed spacer (ITS) gene sequences of 34 taxa representing 6 genera of Thymelaeaceae. (A) maximum-likelihood (ML), and (B) Bayesian inference (BI) tree analyses were conducted in 1,000 bootstrap replicates. Two species, *Psidium guajava* (MN2953604) and *Gossypium australe* (AF057763) were included as outgroups.

Supplementary Files

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

- [SupplementaryMaterials.zip](#)