

Analysis of Phylogenetic Relationships in The Walnut Family Based on Internal Transcribed Spacer Sequences and Secondary Structures(ITS2)

Zhongzhong Guo

Tarim University

Qiang Jin

Tarim University

Zhenkun Zhao

Tarim University

Wenjun Yu

Tarim University

Gen Li

Tarim University

Yunjiang Cheng

Tarim University

Cuiyun Wu

Tarim University

rui Zhang (✉ zhangrms32@163.com)

Tarim University <https://orcid.org/0000-0002-4360-5179>

Research Article

Keywords: Base sequence, Evolution, Juglandaceae, Ribosomal spacer, Secondary structure

Posted Date: May 13th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-501634/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

This study aims to investigate the phylogenetic relationships within the Juglandaceae family based on the Internal Transcribed Spacer's primary sequence and secondary structures (ITS2). Comparative analysis of 51 Juglandaceae species was performed across most of the defined seven genera. The results showed that the ITS2 secondary structure's folding pattern was highly conserved and congruent with the eukaryote model. Firstly, Neighbor-joining (N.J.) analysis recognized two subfamilies: Platycaryoideae and Engelhardioideae. The Platycaryoideae included the Platycaryeae (*Platycarya*+ (*Carya*+ *Annamocarya*)) and Juglandae (*Juglans*-(*Cyclocarya* + *Pterocarya*)). The Engelhardioideae composed the (*Engelhardia*+*Oreomunnea*+*Alfaroa*). The *Rhoiptelea* genus was generally regarded as an outgroup when inferring the phylogeny of Juglandaceae. However, it is clustered into the Juglandaceae family and showed a close relationship with the Platycaryoideae subfamily. Secondly, the folded 3-helices and 4-helices secondary structure of ITS2 were founded in the Juglandaceae family. Therefore, these ITS2 structures could be used as formal evidence to analyze Juglandaceae's phylogeny relationship. The morphology based on the secondary structure nicely coincided with previous investigations. This study further confirms that ITS2 can serve as a valid basis for parsing evolutionary relationships in higher woody fruit trees.

Introduction

Several molecular (nuclear and plastid) studies have been conducted to infer phylogenetic relationships at a low taxonomic level. The coded ribosomal RNA (rDNA) repeat has been useful in intra- and inter-specific phylogenetic studies within the angiosperms. In eukaryotic cells, the highly conserved rRNA genes are abundant and are arranged in a tandem series at the nucleolar organizer region (NORs) of chromosomes. Each unit is comprised of a non-transcribed spacer known as an intergenic spacer (IGS), the transcription unit coding for the precursor of the rRNA, so the composed basic unit contains the following segments is 5'-IGS-18S-ITS1-5.8S-ITS2-26S-3'[1–4]. The two internal transcribed spacers (ITS1 and ITS2), which are part of the nuclear ribosomal cistron, but are not incorporated into mature ribosomes, have been regarded as valuable targets when used as markers in systematic studies. Sequencing of rDNA regions, including the small ribosomal subunit 5.8S and ITS, have been widely applied in phylogenetic classifications because of the relatively higher mutation rate [5–8].

Furthermore, the ITS2 sequence has been widely used to infer phylogenetic relationships at low intra- and inter-hierarchical levels even at the species and genus ranks based on the predominance nature of hypervariable subregions, biparental inheritance, and intergenomic variability [3]. However, the typical challenge on the alignment of ITS2 sequences arises from closely related and recently radiated taxa due to the highly conserved nature of some ITS2 subregions and other parts that seem to vary high randomly substitution [9]. When phylogenetic inferences are limited to subregions with precise alignment, effective information content may be lost. To gain more insights into the processes shaping expansion segments, studies aimed at comparing the evolution of ITS2 primary sequences and secondary structures are required [10, 11]. The ITS domain brings the large and small ribosomal subunits into proximity for mature

processing; the ITS domain's secondary structures were considered important in defining the cleavage sites for the rRNA release. So, the ITS2 secondary structures did not incorporate into mature ribosomes and were relatively highly conserved [9, 11]. Hence, it is crucial to identify the ITS2 sequence and its secondary structure when performing alignments and reconstruct phylogenies.

The Juglandaceae, also known as the walnut family, comprises woody trees and shrubs and belongs to the order Fagales. The Juglandaceae is a vital nut and timber producer in temperate regions of Europe, Asia, South America, and North America. The Juglandaceae consists of 9 genera and 72 species in the world. Based on Manos's results [12], the Juglandaceae family comprises two major subfamilies; the Juglandoideae and the Engelhardioideae. Within the Engelhardioideae subfamily, the genus *Engelhardia* was paraphyletic to the sub-clade composed of *Oreomunnea* + *Alfaroa* genus. In the Juglandoideae, the full Juglandaceae clade (*Carya*-(*Juglans*-(*Cyclocarya* + *Pterocarya*))) was the sister group to the Platycaryeae genus. However, Schaarschmidt [13] suggested that the Juglandaceae comprised the Platycaryoideae and Juglandoideae. The Platycaryoideae contains two sister groups, the Platycaryeae and Engelhardieae, and the Juglandoideae includes the *Juglans*, *Pterocarya*, and *Cyclocarya* genera.

The germplasm in the *Juglans* genus is of great economic value in the Juglandaceae family. Based on leaf and flower morphology by Dode [14], and fruit and nut morphology by Manning's [15], the *Juglans* genus can be divided into four sections, namely, *Dioscaryon* Dode (traditional *Juglans*), *Rhysocaryon* Dode (black walnuts), *Cardiocaryon* Dode (Asian butternuts), and *Trachycaryon* Dode (American butternut). Aradhya [16] suggested that the *Juglans* section includes the cultivated Persian (English walnut, *Juglans regia* L.), and iron walnut (*Juglans sigillata* Dode), which is originated in the Southern Chinese province of Yunnan and Tibet. Based on results published by Ross-Davis [17], the *J. cinerea* was located close to *J. nigra* and other members of the *Rhysocaryon* section than to the Asian members of the *Cardiocaryon* section.

Previous studies based on fossils, morphology, chemical, chromosome, and sequence analysis have established a framework for the Juglandaceae family [12] [18, 19]). The present cladistic analysis included 51 taxa across six genera (including *Juglans*, *Cyclocarya*, *Pterocarya*, *Platycarya*, *Carya*, *Annamocarya*), with the outgroup *Rhoiptelea* and the partial subfamily Engelhardioideae (*Engelhardia*, *Alfaropsis*, *Oreomunnea*). The present study focused on the features of the secondary structure of ITS2. The objective of the study was to provide new information for the phylogenetic status of the Juglandaceae family. This work will improve our understanding of the phylogeny of the Juglandaceae family.

Materials And Methods

Plant materials

A total of 51 taxa were included in the analysis. A total of 21 accessions were collected in China, which included nineteen ingroup taxa of accessions collected from different locations, including senile

germplasm, wild germplasm, and *Pterocarya hupehensis*. The last was *Juglans nigra* (introduced from the USA). The ITS sequence of other genera was downloaded from the GeneBank of NCBI (<http://www.ncbi.nlm.nih.gov>) (Table 1).

Table 1

Species list, collection site, geographic origin, and GenBank accession number of the 51 Juglandaceae species used for phylogenetic analysis

Species list	Collection site	Species/origin	Description	GenBank accession number/ provider
21 <i>Juglans regia</i>	Shannxi	China	old cultivars	
9 <i>Juglans regia</i>	Hutian	Xinjiang China	old cultivars	
13 <i>Juglans regia</i>	Wensu	Xinjiang China	old cultivars	
16 <i>Juglans regia</i>	Ili	Xinjiang China	old cultivars	
28 <i>Juglans regia</i>	Jining	China	old cultivars	
14 <i>Juglans regia</i>	Ruoqiang	Xinjiang China	old cultivars	
23 <i>Juglans regia</i>	Wenquan	Gansu China	cultivars	
18 <i>Juglans regia</i>	Ili	Xinjiang China	wild walnut	
15 <i>Juglans regia</i>	Kuche	Xinjiang China	old cultivars	
11 <i>Juglans regia</i>	Yecheng	Xinjiang China	old cultivars	
17 <i>Juglans regia</i>	Ili	Xinjiang China	wild walnut	
10 <i>Juglans regia</i>	Hutian	Xinjiang China	old cultivars	
12 <i>Juglans regia</i>	Yecheng	Xinjiang China	old cultivars	
25 <i>Juglans regia</i>	Jinin	China	old cultivars	
19 <i>Juglans regia</i>	Ili	Xinjiang China	wild walnut	
27 <i>Juglans regia</i>	Tibet	China	wild walnut	*
8 <i>Juglans regia</i>				EF141011
24 <i>Juglans regia</i>	Wensu	Xinjiang China	cultivars	

* provided by the National Fruit Germplasm Tai'an Walnut & Chestnut Repositor.

Species list	Collection site	Species/origin	Description	GenBank accession number/ provider
26 <i>Juglans sigiliata</i>	Yunnan	China	landraces	*
<i>Juglans cinerea</i>				AF338480
22 <i>Juglans mandshurica</i>	Shannxi	China	Landraces	
<i>Juglans hopeiensis</i>	Hebei	China	landraces	*
42 <i>Juglans mandshurica</i>				AF179576
<i>Juglans major</i>				AF338480
<i>Juglans neotropica</i>				AF179578
<i>Juglans australis</i>				AF179568
<i>Juglans guatemalensis</i>				AF179573
<i>Juglans hindsii</i>				AF338477
<i>Juglans californica</i>				AF338475
<i>Juglans olanchana</i>				AF179580
<i>Juglans microcarpa</i>				AF338485
5 <i>Juglans nigra</i>				AF338490
20 <i>Juglans nigra</i>	Ili	Xinjiang China	Introduction from USA	
<i>Cyclocarya paliurus</i>				AF303817
<i>Pterocarya hupehensis</i>	Hubei	China	landraces	
<i>Pterocarya macroptera</i>				AF303814
<i>Pterocarya stenoptera</i>				AF303816
<i>Platycarya strobilacea</i>				AF303808
<i>Carya glabra</i>				AF303823
* provided by the National Fruit Germplasm Tai'an Walnut & Chestnut Repositor.				

Species list	Collection site	Species/origin	Description	GenBank accession number/ provider
<i>Carya myristiciformis</i>				AF303821
<i>Annamocarya sinensis</i>				AF303818
<i>Carya tonkinensis</i>				AF303826
<i>Carya cathayensis</i>				AF303819
<i>Rhoiptelea chiliantha</i>				AF303800
<i>Engelhardia serrata</i>				EF141010
<i>Engelhardia spicata</i>				AF303802
<i>Alfaropsis roxburghiana</i>				EF141008
<i>Alfaroa guanacastensis</i>				AF303804
<i>Alfaroa costaricensis</i>				AF303803
<i>Oreomunnea Mexicana</i>				EF141013
<i>Oreomunnea pterocarpa</i>				EF141014
* provided by the National Fruit Germplasm Tai'an Walnut & Chestnut Repositor.				

Plant materials and DNA extraction

Young leaves of 21 accessions growing in China were selected for genomic DNA extraction. The modified CTAB protocols used for DNA extraction were described by Cheng et al. [20].

PCR conditions and ITS sequencing

The ITS2 sequences were explicitly amplified. In the first step, the 18S-5.8S-26S region was amplified and sequenced for all genotypes. PCR amplified the ITS region of the nuclear ribosomal cistron in a 20 µl reaction mixture containing 25 ng DNA as a template, 0.1 µM of forward and reverse primers, 2.5 mM MgCl₂, 2.5 mM dNTPs, 1× *Taq* buffer (100 mM Tris-HCl pH8.0, 500 mM KCl, 0.8% Nonidet P40) and 1 U *Taq* DNA polymerase (Fermentas Inc., MD, USA). Amplification was performed in a PTC-200 DNA Engine Thermal Cycler (Bio-Rad Laboratories Inc., Hercules, California, USA) in 0.2 ml tubes. The universal forward ITS5 primer 5'-GGAAGTAAAAGTCGTAACAAGG-3', and reverse ITS4 primer 5'-

TCCTCCGCTTATTGATATGC-3' (White et al. 1990) were used in the completed ITS segments amplification. The primers were synthesized by Shanghai Sangon Biological Engineering Technology (China). The reaction proceeded under the following conditions: initial denaturing step at 94°C for 3 min and 31 cycles of 60 s denaturing at 94°C, 30 s annealing at 52°C, 45 s elongation at 72°C, and one final cycle of 5 min at 72°C, then storage at 4°C.

The amplified products were cleaned using the UNIQ-10 DNA extraction kit (No SK1132; Shanghai Sangon Biological Engineering Technology) following the manufacturer's instructions. The purified PCR products were cloned into the TaKaRa pMD18-T vector (TaKaRa Bio, Shiga, Japan), and inserts from one to four transformed colonies were sequenced with m13- and m13 + primers by Aoke Biotechnology (Beijing).

Sequence alignments and sequence analysis

A total of 51 sequences were compared with the ITS sequence (GeneBank Accession No. AF399876, NCBI (<http://www.ncbi.nlm.nih.gov/>)) to determine the boundaries of ITS2. The molecular clock hypothesis was tested by software packages such as RRTree [21] and phylip3.67 [22]. To further evaluate the relative rates of substitution. Both these two methods checked for change rates, and the results refused the molecular clock hypothesis to the Juglandaceae ITS2 sequence because of the low substitution rates. The phylogenetic analysis was accomplished using MEGA 4 [23]. A total of 51 ITS2 sequence matrix were analyzed as follows: (1) the alignment was performed online (<http://align.genome.jp/>) with further manual adjustments. (2) N.J. analysis was used to evaluate possible tree topologies by MEGA 4; the rules of the N.J. tree were 500 bootstrap values and using the Kimura 2-parameter correction.

Inference of the Juglandaceae ITS2 secondary structure

For predicting the secondary structures of ITS2, the Mfold web server (<http://frontend.bioinfo.rpi.edu/applications/mfold/cgi-bin/rnaform1.cgi>) was used with default parameters. Structures inferred by Mfold were examined for standard stems, loops, and bulges.

Results

Sequence alignment and analysis

A practical primer pair for the amplification of the ITS region was used for all accessions. The length of the ITS2 region varied from 205 bp (*Oreomunnea Mexicana*) to 226 bp (*Annamocarya Sinensis*). The length variation results from indels (Rafael et al., 2007) or a high proportion of slippage replication of 2–3 bp along the sequence [11]. After alignment, the consensus sequence for the fully aligned length was 250 bp. The number of variable sites was 130, of which 73 were parsimony informative. When the ITS2 sequences were aligned, indels in different evolutionary lineages of the genus were present, corresponding to some mutable positions, such as 48th bp-57th bp, 85th -88th bp, and 223th -233th bp. Biologically, those gaps are assumed to represent insertions or deletions that occurred as the sequences diverged from a common ancestor [4].

Phylogenetic relationships

All analyses based on N.J. phylogeny trees supported *Cyclocarya* as a basal sister genus to *Pterocarya*, while the clade of *Cyclocarya-Pterocarya* was a sister group to the *Juglans* genus. The *Rhysocaryon*, *Juglans*, *Trachycaryon*, and *Cardiocaryon* sections were clustered into the monophyletic *Juglans* genus. The *Juglans* section supported the monophyletic *Regia* and *sigillata* clade (bootstrap = 94%) while demonstrating a sister relationship with the other three sections. The *Trachycaryon* section was clustered as a sister group to the *Cardiocaryon* section (marginal bootstrap = 56%). The topology within the *Rhysocaryon* section, which was supported as a monophyletic group, better resolved the group's relationship concerning the different places, including the temperate Western and the temperate Eastern group and tropical group. *Juglans. olanchana* clustered within the temperate Eastern group; however, the geographical position was the Northernmost in the tropical area. The results of the clustering observation showed the continuity of the population. They supported the fossil witness of black walnut spreading paths of Am Cont West Coast to the East Coast and then to the Southern tropical area in America. *J. major*, as a distinct species, clustered to the base of the *Rhysocaryon* section. Within the tropical group, *J. guatemalensis* was a sister group to the *J. neotropica* and *J. australis* species. The temperate Western group contained *J. californica* and *J. hindsii*.

Xi's results showed that China is one of the origins of *J. Regia* [24, 25]. Therefore, native species with different nut morphology characters were collected from different China locations to enrich the analyses. In our results, the numbered accessions clustered as a monophyletic group. The wild walnut did not show an apparent relationship of immediate forebear to the present cultivar. *Annanocarya* belonged to the *Carya* genus, and *Platycarya* was a basal sister group to the *Carya* and *Annanocarya* clade. *Engelhardia*, *Alfaroa*, and *Oreomunnea* were more distant to the other genera in the Juglandaceae. Based on the above results, the Juglandaceae were divided into two subfamilies, and one was Platycaryoideae and Engelhardioideae. The Platycaryoideae included the Platycaryeae (*Platycarya*+(*Carya* + *Annamocarya*)) and Juglandeae (*Juglans*-(*Cyclocarya* + *Pterocarya*)). The Engelhardioideae composed the (*Engelhardia*+(*Oreomunnea* + *Alfaroa*)). The *Rhoiptelea* genus was regarded as an outgroup because of a close affinity to common ancestry with the Juglandaceae family [12]. However, in our results, the *Rhoiptelea* genus clustered into the *Juglandaceae* family and was close to the Platycaryoideae subfamily clade (Supplemental figure).

Compared to the Juglandaceae family's previous resolution placement [12, 15]), a phylogenetic tree derived from the ITS2 sequence N.J. analysis was constructed from our dataset. However, it was not entirely in agreement with a previous study [12, 13, 15]). In the opinion of Manos [12] and Manning [15], *Platycarya* was the sister group to the Juglandineae (*Platycarya*-((*Carya* + *Annamocarya*)- (*Juglans*-(*Cyclocarya* + *Pterocarya*)))) in the Juglandoideae subfamily, but in our results, the constructed subtree was (*Platycarya* + *Carya* + *Annamocarya*)-(*Juglans*-(*Cyclocarya* + *Pterocarya*)). The close relationship between *Platycarya* and *Carya* was similar to the results of Schaarschmidt [13]; however, in the Schaarschmidt's results, the Juglandaceae family was divided into the *Platycaryoideae* and *Juglandoideae* subfamily, and the *Platycaryoideae* was included ((*Platycarya* + *Carya*) – *Engelhardieae*).

The Engelhardioideae subfamily placement in our phylogenetic N.J. tree was monomorphic and was located at the base of (*Rhoiptelea* + Juglandoideae), differing with Schaarschmidt's results.

Comparison of the ITS2 secondary structure of Juglandaceae to the model in eukaryotes and flowering plants

Comparing the results of the clustering taxa of the phylogenetic tree and the secondary structure of ITS2 sequence, many visible results of the secondary structure frequently coincided with the phylogenetic tree. Based on the rules of minimum free energies, the 3-helices and 4-helices ITS2 secondary structures of the RNA transcript of Juglandaceae structures are highlighted in Fig. 1 The ITS2 secondary structure of the *Juglans* section was considered standard compared to all other genera. A two-steps analysis of the ITS2 folding pattern of the Juglandaceae species was conducted. Our initial aim was to identify the structures of Juglandaceae, and we compared their folding pattern with that of model flowering plants, such as *Photinia pyrifolia*, *Sinapis alba*, and *Oryza sativa* [26] and with the general structure of the proposed model for eukaryotes [9]. The Juglandaceae ITS2 secondary structures (i.e., helices) exhibited significant similarities with the described secondary structures for the flowering plants and eukaryotes. Therefore, following the proposed nomenclature of the ITS structure model, the Juglandaceae family exhibited 3-helices and 4-helices structures. The 3-helices structure lacked the short helix located at the 3'-end of the ITS2 sequence. We named the helices from the 5' sequence to the 3' sequence, and we named the A to C (D) helices. The highly variable region located inside the C helix and only presented 3-helices structure, was named C' (Fig. 1). Analyzed the 3-helices and 4-helices structures of all the genotypes, the entire ITS2 sequence can be divided into a shorter 5'-end conserved region (5'-C.R., containing the A and B helices), reflecting a high nucleotide sequence homology. A longer 3'-end variable region [3'-V.R., containing C (include in 3-helices and 4-helices structures) and D (only in 4-helices structure)], showed little sequence homology. Helices A and B exhibited the least variation among the taxa. The position of the two helices was identical in structures with 3- or 4-helices. The sequence of helices A and B started with conserved base-pair triplets emerging from single-strand regions. Helix A was shaped at the fifth nucleotide (5'-GUU/CAG-3') or the sixth nucleotide (5'-UUG/AAC-3'), and helix B was formed from 58 nucleotides or with slight position variations (5'-GGC/CCG-3'). Secondly, compared to the ITS2 structural models of flowering plants and eukaryotes, helix II was typically short, not branched, and included the U-U mismatch [9]. Helix B of the Juglandaceae ITS2 sequence secondary structure corresponded to helix II in the proposed model, since a U-U mismatch was present in all the examined taxa.

Furthermore, helix C, which occasionally formed with lateral branches, demonstrated the highest level of variability and corresponded to the giant helix III based on several proposed hallmarks described for proposed models. In particular, the region of most significant sequence conservation in the eukaryotic ITS2 was found on the 5' side of helix III, where highly conserved and group-specific consensus sequences were described. As shown in Fig. 1, a highly conserved six bp stretch (5'-GGUGGU-3') sequence representing the ITS2 model of terrestrial plants was found at domain C for the examined Juglandaceae family.

Comparison of the shape of the ITS2 secondary structure in the Juglandaceae family

The ITS2 sequence structures in the *Rhysocaryon*, *Trachycaryon*, and *Cardiocaryon* sections were similar to those in the *Juglans* section. The *Cyclocarya* genus, which is only included in the *C. paliurus* species, folded the standard 3-helices structure. The *Pterocarya* genus structure was similar to the *Juglans* section; however, *P. stenoptera* and *P. macroptera* only folded into the 3-helices structures, while *P. hupehensis* folded both the 3-helices and 4-helices structure. *A. sinensis* is the only species included in the *Annanocarya* genus. It folded only one 3-helices structure, which resembled the 3-helices structure of the *Juglans* section. The *Rhoipteleaceae* family shares a common ancestry with the *Juglandaceae* family and belongs to one *Rhoiptelea* genus and the lower-order level species *R. chiliantha*. The secondary structure was only folded into the standard 3-helices structure. The Engelhardioideae subfamily included seven species; only *E. serrata* and *E. spicata* constructed the 4-helices structure, but an angle located at the 3' end of the sequence (Fig. 7). *Alfaropsis roxburghiana*, *Oreomunnea mexicana*, and *Alfaroa guanacastensis* only include the 3-helices structures similar to those of *Juglans regia*'s structures. However, *Oreomunnea mexicana* shaped the A helix at the fourth nucleotide and started with the base-pair triplets CGU/GCA. *Oreomunnea pterocarpa* and *Alfaroa costaricensis* only constructed 4-helices structures. The *Carya* genus was different in the Juglandaceae family. There were four accessions included in the *Carya* genus; *C. tonkinensis*, *C. glabra*, *C. myristiciformis*, and *C. cathayensis* folded the standard 3-helices structures. *C. tonkinensis* and *C. cathayensis* were folded into exclusive 3-helices while clustered to one taxon with a high bootstrap value (95%); (Fig. 2). *C. glabra* and *C. myristiciformis* constructed the lowest free energy structures, 2-helices (Fig. 3), which were exclusive to species in the Juglandaceae. The analyzed *Platycarya* genus only included the *P. strobilacea*. Its ITS2 sequence formed a whole 2-helices structure of lowest free energy (Fig. 4) in the Juglandaceae family. The exclusive nature of the *Platycarya* genus supported the result of Smith [19], where the *Platycarya* was a divided, separated subfamily.

Comparative analysis of the ITS2 sequences and secondary structures of the Juglandaceae

The A helix was more variable than the B helix in the Juglandaceae family. Within the *Juglans*, *Cardiocaryon*, and *Trachycaryon* section, the typical bulge signal was located at the 20th nucleotide and resulted in a small lateral bulge 5'-CAAUG-3' sequence. However, in the black walnut group, the small bulge was changed to a more prominent lateral bulge because of the transversion generated by U to C (5'-CAAACGCG-3'). The giant lateral bulge was differentiated in the tropical black walnut group. The more prominent lateral bulge of *J. australis* and *J. neotropica* was 5'-AAACGCGU-3'; the marginal accession *J. guatemalensis* folded 5'-CAAUGCG-3'. The apex of the A helix showed the typical 5'-CUUAUG-3' sequence in the *Juglans* section. The typical apex of *Cardiocaryon* and *Trachycaryon* was 5'-CUCAUG-3' but changed to 5'-CUUCUUAUGCUG-3' in the *Rhysocarya* section except for *J. guatemalensis*, which showed the minor variant 5'-CUUCUUAUG-3'.

Comparing the 3-helices and 4-helices structures, but paying close attention to the B helix, we observed the conserved structure and sequence present in the different genera. In the *Juglans* genus, the B helix was short and initially shaped at the GGC/CCG base paired with minimum position change, and the U-U mismatch bulge was located inside it. In different taxa, the apex portions of the B helix contained apparent signature sequences. The U-U or C-U mismatched base-pair were separated by only one base pair in the apex in helix B. (Fig. 1). The mismatched base pair was designated the "marker-site". The marker site showed differences in different sections. In the *Juglans* section, the standard structures of the B helix's distal apex included the conserved 5'-CUUCUG-3' structure, and the marker-site was U-U mismatch. The marker site of the *Cardiocaryon* and *Trachycaryon* section was a U-U mismatch base pair, and the apex was 5'-CUUCUG-3' except for the slight variation in the *J.mandshurica* accession, where the apex was 5'-CUUGUG-3'. From the result of Stanford [27], the black walnut section was separated into two clades, namely the temperate and tropical clades. The temperate clade was separated into the Eastern and Western groups. In our structures, the tropical clade included the C-U mismatch, and the apex was 5'-CUUUUG-3'. The Western temperate group, *J. Californica*, and *J. hindsii* included the C-U mismatch, and the apex was 5'-CUUGUG-3'. The structure of the Eastern temperate group exhibited a weak difference to the Western group but was similar to the tropical clade. The similar secondary structures among the Western, Eastern, and tropical clades were congruent with fossil survey evidence, where the black walnut was distributed from the West to the East Coast and then extended into the Southern area during Oligocene to the Miocene. These similarities inferred that the time of black walnut migration from the East Coast to the Southern area was shorter than from the West to the East Coast. Within the *Pterocarya* genus, the apex of the B helix was a slippage replication region. The apex of *P. macroptera* was 5'-CUCUUG-3', while those of *P. stenoptera* and *P. hupehensis* were 5'-CUCUUCUCG-3'; the genus marker-site was U-U mismatch. Therefore, the *Pterocarya* genus apex of the B helix exhibited a distinctive elongation at the apical portion, the supposed conserved core motif, "C.U.". The apex of the B helix was the typical signal of the Engelhardioideae subfamily, 5'-CUGUCG-3', which was shared among all the subfamily; the marked site changed C-U. In the *Carya* genus, the famous 3-helices structures of *C. tonkinensis*, *C. glabra*, *C. myristiciformis*, and *C. canthayensis* showed the 5'-CGUUG-3' diagnostic apex in the B helix and the marked site transversion to C-U. The apex of the B helix of *Annamocarya sinensis* was 5'-CGUUG-3', and the marker-site was a C-U mismatch. The B helix apex in *Cyclocarya paliurus* was 5'-CUCUUG-3', and the marker-site was a U-U mismatch. *Rhoiptelea chiliantha* included the 5'-CUGUAG-3', and the marker site was a C-U mismatch.

We carefully compared the C helix of 3-helices structure. The C' was the highly variable region and included three bulges in the *Juglans* section. In the C' region of structures folded from the *Cardiocaryon* and *Trachycaryon* sections, the first bulge included one more significant angle. However, it was excluded from the structure of the *Juglans* section. After carefully checked the other species, the structures included the more significant angle in the C' part in the temperate Eastern group of the *Rhysocaryon* section and the *Pterocarya* genus. The C' sequence showed a more significant angle from the alignment results due to the transversion generated at the 205th base pair (the position varied slightly) by the C being changed to U (Fig. 5).

The D helix was only constructed in the 4-helices structure. It was a short helix located at the 3'-end of the ITS2 sequence, and some of the species constructed it. Compared to the standard structure of the *Juglans* section, the *Cardiocaryon* Dode (Asian butternuts) and *Trachycaryon* Dode (American butternut) section constructed the D helix with one bulge (Fig. 6). However, the standard structure of the *Juglans* section had no bulge. Within the Engelhardioideae subfamily, the *Engelhardia* taxon and the species *Oreomunnea pterocarpa* and *Alfaroa costaricensis* only constructed the 4-helices structures.

Discussion

The ITS sequence has proven to be valuable in systematic studies [3, 11, 16]). As part of the rDNA sequence, ITS2 can fold into variable secondary structures depending on its primary sequence, which has been proven useful for resolving closely related taxa phylogenies since they are relatively fast-evolving sequences [11, 28]). Based on the structural parameters of the predicted secondary structure (geometrical features, bond energies, base composition) [29], the ITS2 folded form provide a new source of phylogenetic information which has been used as supplement evidence in sequence alignment for phylogenetic reconstructions [9, 11, 30]). The visible morphological markers of structure contained a remarkable signature sequence, e.g., the bulge and apical portion shared among specific taxa. In the Juglandaceae family, except for the accessions that did not fold the standard structure, the common core 3-helices and 4-helices structures of ITS2 were observed in several taxa; the U-U mismatch at the 5' sequence and the highly 5'-GGUGGU-3' conserved sequence hallmarks at the 3' sequence were also present, both of which are putatively important for ribosomal RNA processing [9].

In the previous works of the evolution of the Juglandaceae family, the accepted clades were (*Juglans*-(*Cyclocarya* + *Pterocarya*)) and (*Carya* + *Annamocarya*) [12, 15]). The folded ITS2 secondary structure is the standard structure popular among all genera, except for *C. cathayensis* and *C. tonkinensi*, which constructed the unique three-helices structure; *Platycarya strobilacea* folded the 2-helices structure. We downloaded two sequences of *Platycarya strobilacea*, the only species included in the genus, and both constructed the excludable 2-helices structure with free energy higher than the standard structure. Therefore, the *Platycarya* genus was highly specialized in the Juglandaceae family and coincided with Manning's classification [15].

The identified marker sites could be considered diagnostic since they were specific to certain groups of species; for example, the U-U positional homologous marker site included in the helix B was identified as a diagnostic marker shared across the *Juglans*, *Cardiocaryon*, *Trachycaryon* section, as well as the *Pterocarya* and *Cyclocarya* genera. Simultaneously, the C-U mismatch base pair was matched in the *Rhysocaryon* section and the *Carya*, *Annamocarya*, *Rhoiptelea*, and *Engelhardia* genera.

Within the genus, the different clades were concordant with the geographical distribution. The taxa of the *Rhysocaryon* section were divided into tropical and temperate region populations. We observed that *J. hindsii* and *J. californica* folded the standard 3-helices and 4-helices structures, but other black walnut species only folded the 3-helices structure. The phenomenon was explained through the clustering

analyses of NCS (chloroplast DNA noncoding spacer) + matK by Aradhya [16]. The explanations included the historical gene introgression from the ancestral *J. cinera* species and black walnut species derived from an independent ancestor in Northwestern America [31].

Data from the ITS2 NJ analyses confirmed that the *Cardiocaryon-Trachycaryon* clade was monophyletic. This result was congruent with results of the distribution of fossils, which revealed that butternuts might have originated and radiated independently from high Northern latitudes while leaving small refuges of disjunction populations in Eastern Asia and North America of geo-climatic change [16]. In the *Cardiocaryon-Trachycaryon* clade, included *J. hopeiensis* which the morphology closely resembles that of *J. Regia*, yet the nut characters are similar to that of *J. mandshurica*. *J. hopeiensis* was considered a subspecies of *J. mandshurica* [16] or an interspecific hybrid between *J. Regia* and *J. mandshurica* [32]. So far, it was argued to the *J. hopeiensis* taxonomic status. In the N.J. analysis, the bootstrap (47%) confirmed the subspecies relationship between *J. mandshurica* and *J. hopeiensis*, as did secondary structure features. These results were similar to those of [33], who used the RAPD marker, and the data clustered *J. hopeiensis* into the *Cardiocaryon* section. The structure of ITS2 typically supported the *Cardiocaryon-Trachycaryon* monophyletic clade, i.e., the marker site U-U mismatch and the D helix with one bulge in the 4-helices structure. The accessions collected in China (coded), including wild walnut located in Ili county, the cultivar (including senility trees and variety), considered *J. Regia* frequently, and the *J. sigillata* native to Yunnan province. In our results, the three groups clustered as a monomorphic clade composed the *Juglans* section. They did not diverge into a new species, thus matching the result of China's ecotype distributions from a morphology study [24].

The structures of the numbered accessions were all examples of standard folding. The typical accessions were collected from Ili county (collection location: latitude 43° 28'N, longitude 82° 08'E) and were considered wild walnut. From the results of Xi [25], the Ili county wild walnut was a remnant of the later Tertiary, and the wild walnut were the immediate ancestors to the cultivar in Central Asia. From the result of Chang [34], the temperate mixed conifer-broadleaf forest during the Oligocene was widely distributed in the Tianshan Mountain area *Juglans* existed in it. N.J. analyses positioned the wild walnut and mixed cultivar together. Therefore, the monomorphic clade result could not validate the direct ancestral relationship to the present cultivar variety. Meanwhile, the monomorphic analysis supported the results of Aradhya [16], which stated that the *Juglans* section was an ancient lineage, and the entire lineage of the *Juglans* section perhaps evolved in isolation in Eurasia from ancestral forms because of the scarcity of stratigraphic data from the places, where the historical distribution of ancestral forms and the domestication were known to have existed.

The palynology evidence in China supported the results of Aradhya [16], which stated the *Juglans* section evolved in isolation in Eurasia and from ancestry forms, such as in Southern and Southwestern China; *J. Regia* pollen existed during the later Eocene to early Oligocene; and Qaidam Basin and Tarim Basin discovered *J.regia* fossil pollen grains dating from the early Oligocene period [35]. In the Tuotuo River area in the Tanggula Mountains, *Juglans* pollinates during the later Eocene to the early Oligocene were reported by Duan [36]. The *J. sigillata*, with many primitive nut characteristics, is widely distributed in

Yunnan and Tibet province. Comparing the *J. sigillata* position to the surrounding of cultivar and wild walnut (bootstrap 94%) showed that the *J. sigillata* species might be more closely related to the ancestor of the *Juglans* section. In the *Carya* genus, four populations that were clustered into two groups were distributed in the Old World (margin bootstrap 97%) (*C. cathayensis* + *C. tonkinensis*) and two in the New World (bootstrap 69%) (*C. glabra* + *C. myristiciformis*). Compared to the structures of the two *Carya* clades' accessions, the accession from the New world folded the lowest free energy structures were unique two-helix structures. In contrast, the accession from the Old world folded the special 3-helices structure and did not have the obvious U-U mismatch at the 5' sequence.

In conclusion, the ITS2 sequence was a double-edged tool for comparing low levels of taxonomic phylogeny relationships. The phylogeny tree from the sequence alignment and parsimony-informative sites, another is the conserved ITS2 typical secondary structure prevalent in the Juglandaceae family, and further comparisons of the structure are the powerful supplement to the sequence alignment results. The following conclusions can be formed from this work. (1) In the Juglandaceae, the 3-helices and 4-helices structures are popular among all genera except the *Platycarya* genus. (2) The ITS2 structure results were arrested for the disjunction of the *Cardiocaryon* Dode (Asian butternuts) and the *Trachycaryon* Dode (American butternut). The clustering and structural features of the butternuts completely supported the close disjunction relationship between *Cardiocaryon* and the *Trachycaryon*. (3) The structure of *J. hopeiensis* supplied the direct evidence of the formal characteristics for taxonomic classification and therefore strengthened the clustering results to the *Cardiocaryon* section.

Note

The ITS2 sequence of walnut in the Supplemental Table will be submitted on NCBI, the submission will be completed before final acceptance, and the accession numbers will be supplied once available.

Declarations

Acknowledgements

This study was funded by the National Natural Science Foundation of China (31460503, 31260463) and The Major Scientific and Technological Projects of the Xinjiang Production and Construction Corps of china (2017DB006).

Conflict of interest

The authors declare that they have no conflict of interest.

Author contributions

Zhongzhong Guo and Qiang Jin prepared the manuscript; Zhenkui Zhao and Wenjun Yu participated in data analysis and experiments; Yunjiang Cheng and Cuiyun Wu performed reference research; Rui Zhang prepared study design.

References

1. Takaiwa F, Oono K, Sugiura M (1985) Nucleotide sequence of the 17S-25S spacer region from rice rDNA. *Plant Mol Biol* 4(6):355–364
2. Barker RF et al (1988) Structure and evolution of the intergenic region in a ribosomal DNA repeat unit of wheat. *J Mol Biol* 201(1):1–17
3. Alvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Mol Phylogenet Evol* 29(3):417–434
4. Witono JR, Kondo K (2007) Molecular phylogeny of *Pinanga* (Palmae) based on internal transcribed spacers (ITS) sequence data. *Chromosome Botany* 2(1):25–37
5. Edger PP et al (2014) Secondary structure analyses of the nuclear rRNA internal transcribed spacers and assessment of its phylogenetic utility across the Brassicaceae (mustards). *PLoS One* 9(7):e101341
6. Koetschan C et al (2014) Internal transcribed spacer 1 secondary structure analysis reveals a common core throughout the anaerobic fungi (Neocallimastigomycota). *PLoS One* 9(3):e91928
7. Giudicelli GC et al (2017) Secondary structure of nrDNA Internal Transcribed Spacers as a useful tool to align highly divergent species in phylogenetic studies. *Genet Mol Biol* 40(1 suppl 1):191–199
8. Zhu S et al (2018) Phylogenetic analysis of *Uncaria* species based on internal transcribed spacer (ITS) region and ITS2 secondary structure. *Pharm Biol* 56(1):548–558
9. Coleman AW (2007) Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. *Nucleic Acids Res* 35(10):3322–3329
10. Gómez-Zurita J, Juan C, Petitpierre E (2000) Sequence, secondary structure and phylogenetic analyses of the ribosomal internal transcribed spacer 2 (ITS2) in the *Timarcha* leaf beetles (Coleoptera: Chrysomelidae). *Insect Mol Biol* 9(6):591–604
11. Trizzino M et al (2009) Comparative analysis of sequences and secondary structures of the rRNA internal transcribed spacer 2 (ITS2) in pollen beetles of the subfamily Meligethinae (Coleoptera, Nitidulidae): potential use of slippage-derived sequences in molecular systematics. *Mol Phylogenet Evol* 51(2):215–226
12. Manos PS, Stone DE, *Evolution, phylogeny, and systematics of the Juglandaceae*. *Annals of the Missouri Botanical Garden* (2001) p. 231–269
13. Schaarschmidt H (1985) The relationship between *Carya* and *Platycarya* (Juglandaceae) and the natural classification of the family. *Feddes Repertorium* 96:345–361
14. Dode LA, *Contribution to the study of the genus Juglans*. 1909
15. Manning WE, *The classification within the Juglandaceae*. *Annals of the Missouri Botanical Garden* (1978) p. 1058–1087
16. Aradhya MK et al (2007) Molecular phylogeny of *Juglans* (Juglandaceae): a biogeographic perspective. *Tree Genetics Genomes* 3(4):363–378

17. Ross-Davis A, Woeste KE (2008) Microsatellite markers for *Juglans cinerea* L. and their utility in other Juglandaceae species. *Conservation genetics* 9(2):465–469
18. Manchester SR (1989) *Early history of the Juglandaceae*, in *Woody plants—evolution and distribution since the Tertiary*. Springer, pp 231–250
19. Smith JF, Doyle JJ (1995) A cladistic analysis of chloroplast DNA restriction site variation and morphology for the genera of the Juglandaceae. *Am J Bot* 82(9):1163–1172
20. Cheng Y-J et al (2003) An efficient protocol for genomic DNA extraction from Citrus species. *Plant Molecular Biology Reporter* 21(2):177–178
21. Tajima F (1993) Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* 135(2):599–607
22. Felsenstein J (1993) PHYLIP (phylogeny inference package), version 3.5 c. Joseph Felsenstein
23. Tamura K et al (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular biology evolution* 24(8):1596–1599
24. Shengke X (1987) GENE RESOURCES OF JUGLANS AND GENETIC IMPROVEMENT OF JUGLANS REGIA IN CHINA [J], 3. *Scientia Silvae Sinicae*
25. Rongting X (1990) Textural Criticism of Walnut (*Juglans regia* L.) Origin in China [J], 1. *Journal of Agricultural University of Hebei*
26. Mai JC, Coleman AW (1997) The internal transcribed spacer 2 exhibits a common secondary structure in green algae and flowering plants. *J Mol Evol* 44(3):258–271
27. Stanford AM, Harden R, Parks CR (2000) Phylogeny and biogeography of *Juglans* (Juglandaceae) based on matK and ITS sequence data. *Am J Bot* 87(6):872–882
28. Schultz J et al (2005) A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. *Rna* 11(4):361–364
29. Billoud B et al (2000) Cirripede phylogeny using a novel approach: molecular morphometrics. *Mol Biol Evol* 17(10):1435–1445
30. Coleman AW (2003) ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *TRENDS in Genetics* 19(7):370–375
31. Fjellstrom R, Parfitt D (1995) Phylogenetic analysis and evolution of the genus *Juglans* (Juglandaceae) as determined from nuclear genome RFLPs. *Plant Syst Evol* 197(1):19–32
32. Mulligan B (1941) *Manual of Cultivated Trees and Shrubs Hardy in North America*. *Nature* 148(3740):7–7
33. Wu Y et al (2000) A study on the genetic relationship among species in *Juglans* L. using RAPD markers. *Acta Horticulturae Sinica* 27(1):17–22
34. Hsin-shi C, *On the eco-geographical characters and the problems of classification of the wild fruit-tree forest in the Ili Valley of Sinkiang*. *Journal of Integrative Plant Biology*, 1973. 15(2)
35. Wang X-M, Wang M-Z, Zhang X-Q (2005) Palynology assemblages and paleoclimatic character of the Late Eocene to the Early Oligocene in China. *Earth Sci* 30(3):309–316

36. Duan Q et al (2007) Sporopollen assemblage from the Totohe Formation and its stratigraphic significance in the Tanggula Mountains, northern Tibet. *Earth Science–Journal of China University of Geosciences* 32:623–637

Figures

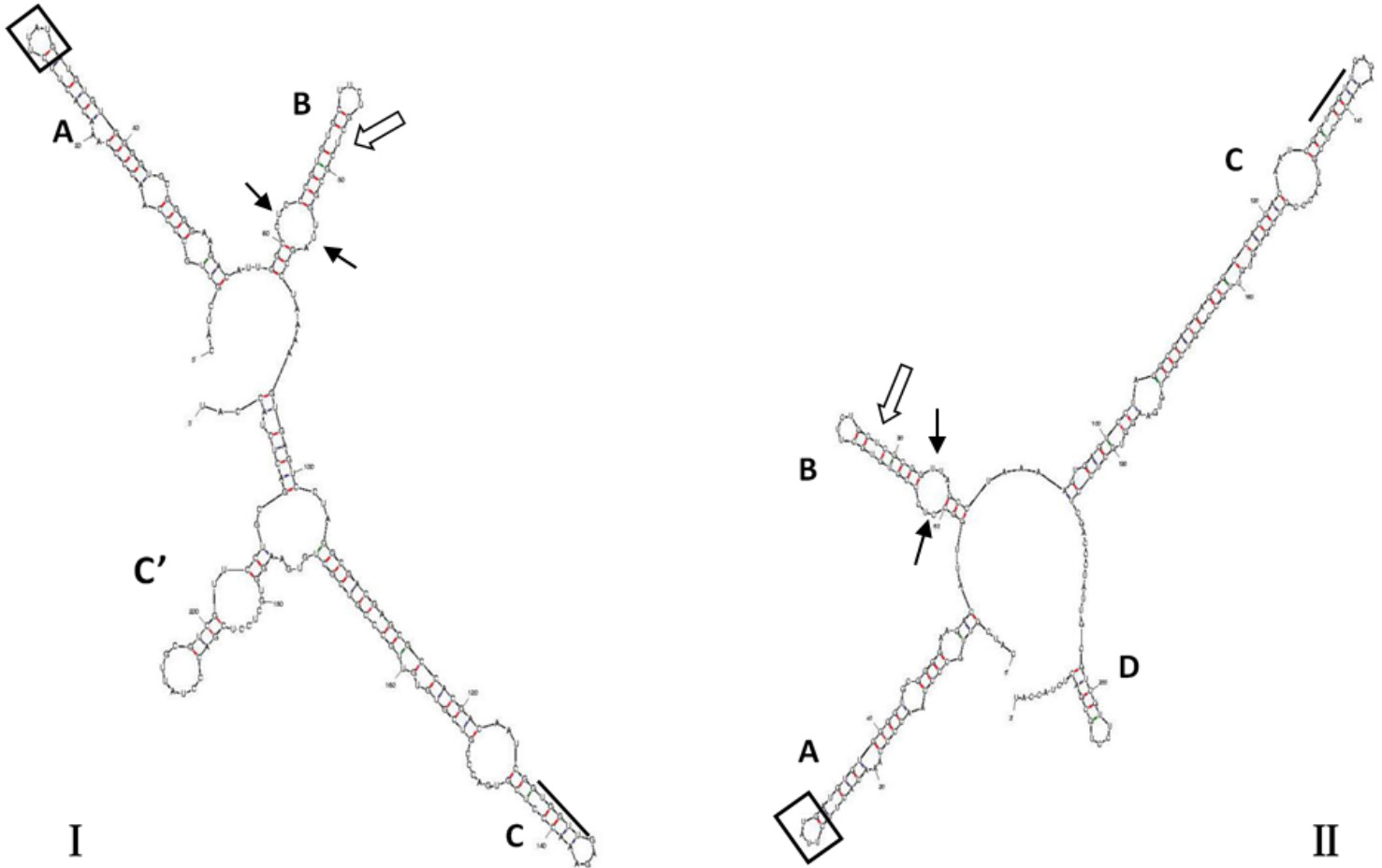


Figure 1

The standard secondary structure of the ribosomal RNA ITS2 transcript of Juglandaceae; (☒) the standard 3-helices structure; (☒) the standard 4-helices structure. Arrows indicate the U-U bulge; a line highlights the consensus sequence -5'GGUGGU-3' on the c helix's 5' side. The walnut's hairpin loop (5'-CUUCUG-3') and black walnut (5'-CUUCUUAUGCUG-3') in a helix are boxed; two hollow arrows indicated the marker site.

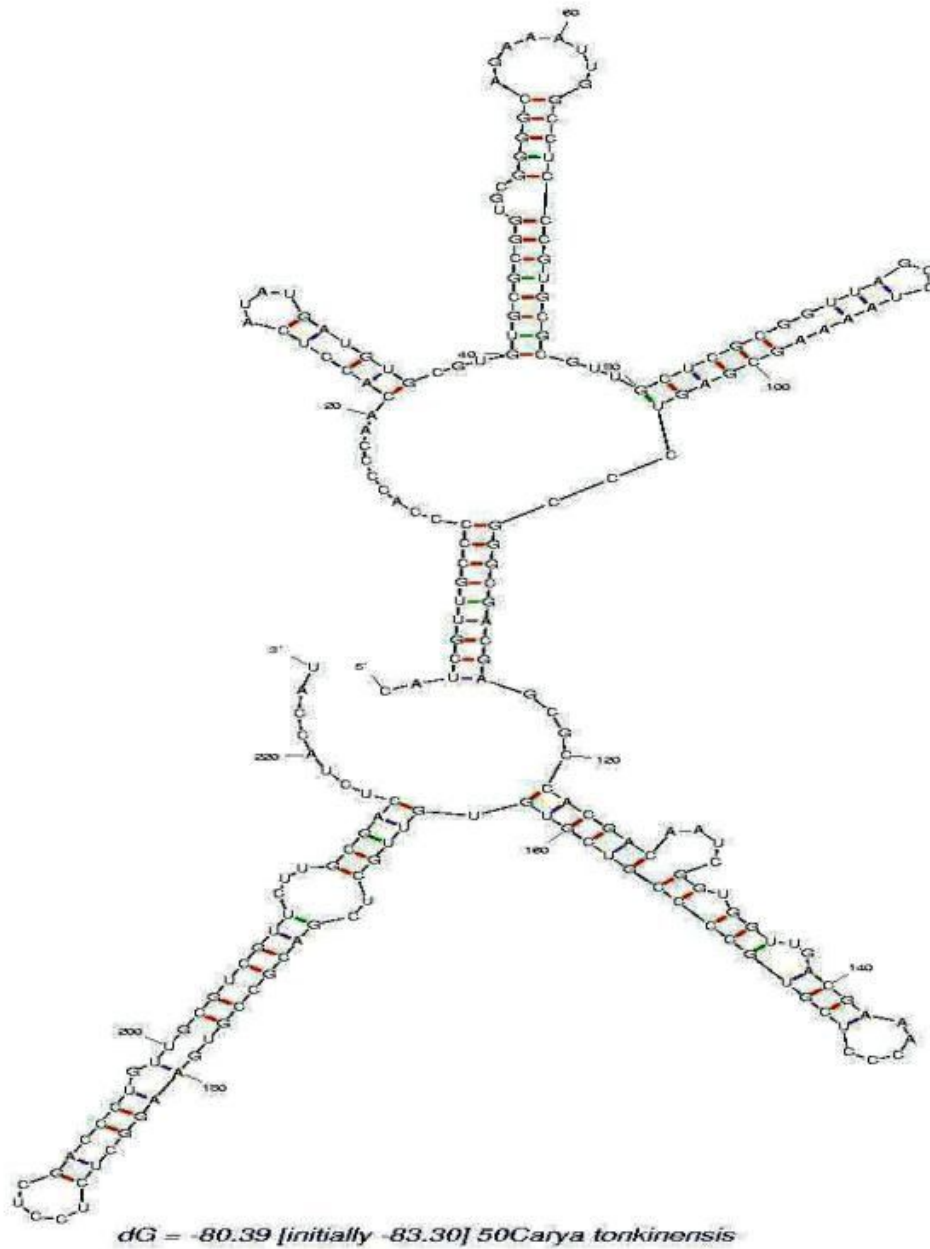
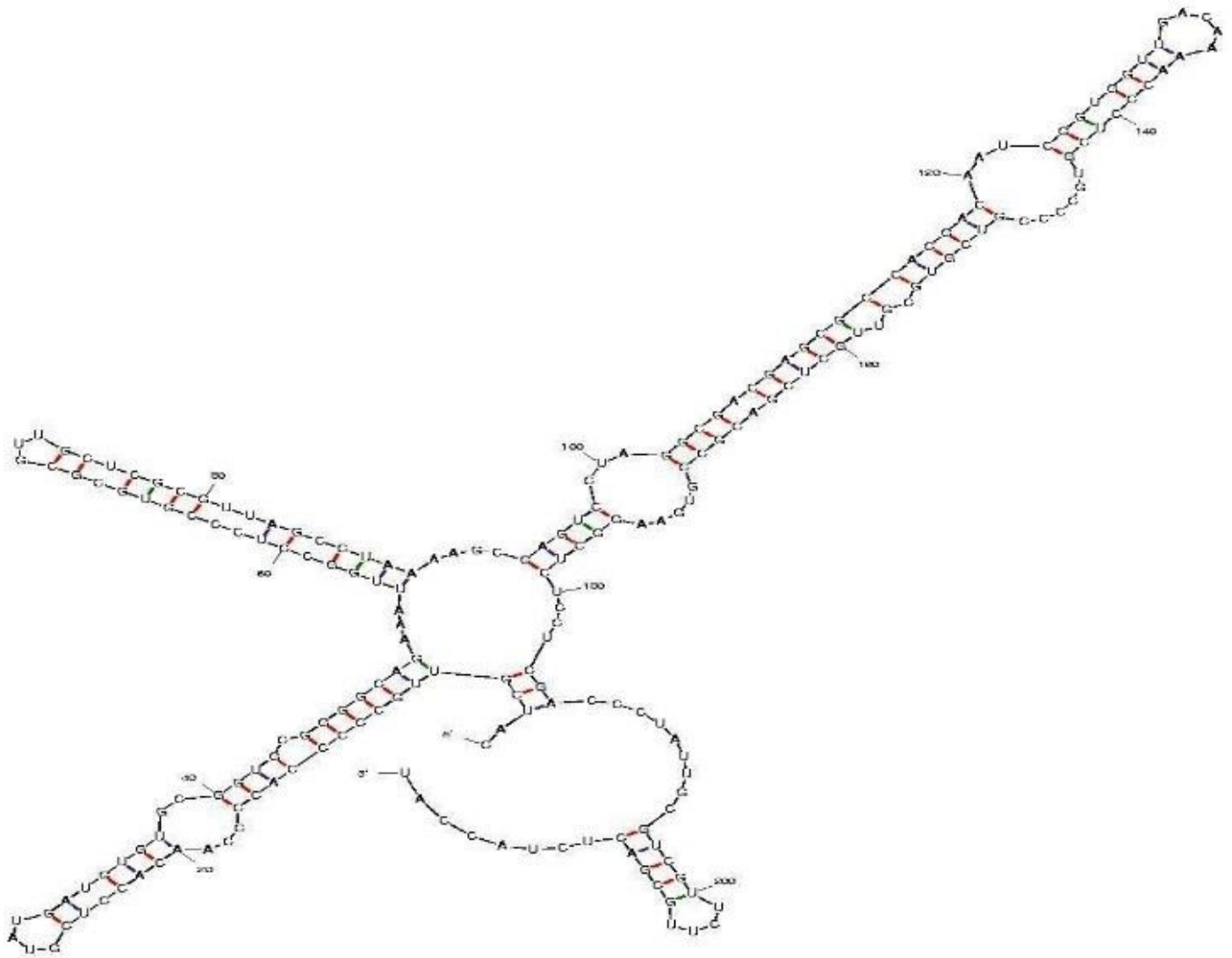


Figure 2

The minimum free energy structures of *C. tonkinensis* and *C. cathayensis* folded the unique 3-helices secondary structures.



$dG = -78.43$ [initially -80.00] 53Carya glabra

Figure 3

The lowest free energy 2-helices structures of *C. glabra* and *C. myristiciformis*.

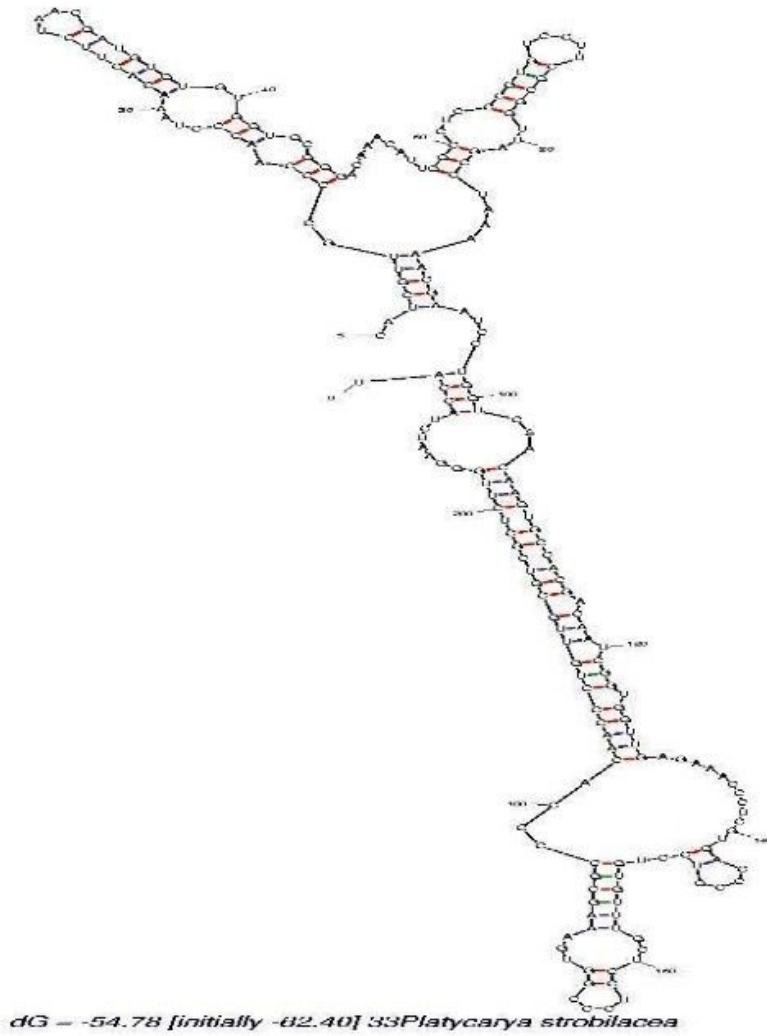


Figure 4

Platycarya strobilacea folded the 2-helices structure.

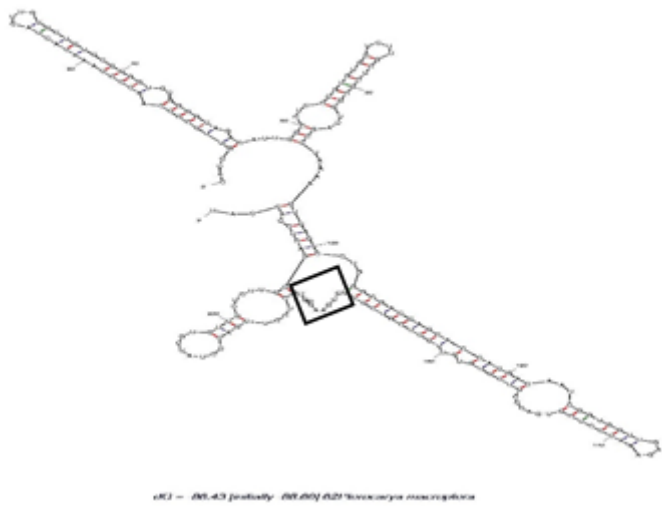


Figure 5

The secondary structure of the genus *Pterocarya*, the larger angle, is highlighted in the box.

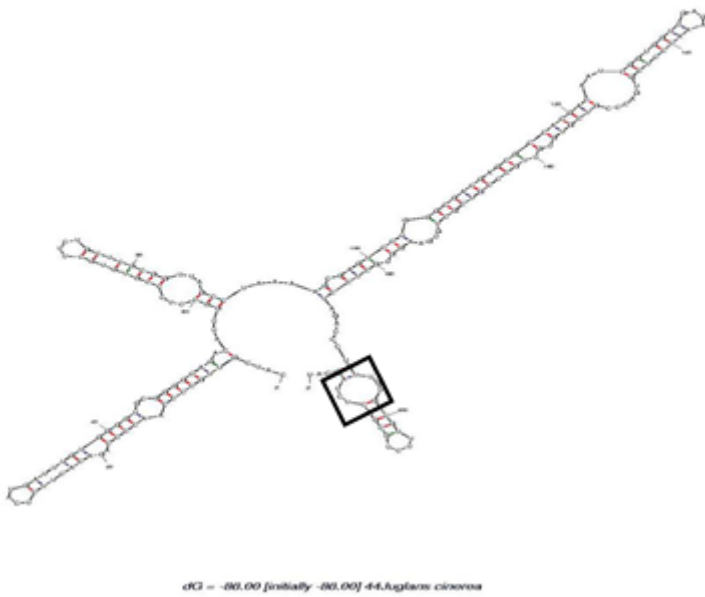


Figure 6

The *J. cinerea* species; the box emphasized particular features of the *Cardiocaryon* Dode (Asian butternuts) and *Trachycaryon* Dode (American butternut) sections.

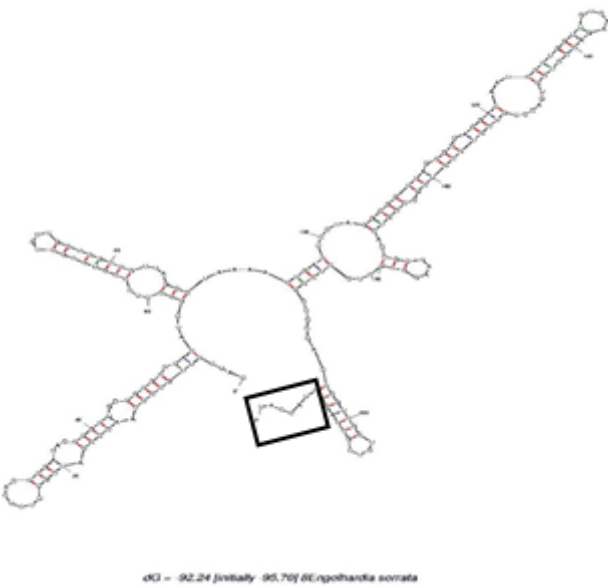


Figure 7

The secondary structure of the *E. serrata* species; *E. spicata* and *E. serrata* only constructed a 4-helices structure at an angle located at the 3' end of the sequence.

Supplementary Files

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

- [supplementaryfigure.png](#)