Journal of Molecular Evolution (2020) 88:361–371
https://doi.org/10.1007/s00239-020-09937-1

ORIGINAL ARTICLE

Horizontal Gene Transfer has Impacted *cox1* Gene Evolution in *Cassytha filiformis*

Canyu Zhang[1,2] · Hui Ma[1] · M. Virginia Sanchez-Puerta[3,4] · Lang Li[1,5] · Jianhua Xiao[1,2] · Zhifang Liu[1,2] · Xiuqin Ci[1,5] · Jie Li[1,5]

Received: 9 October 2019 / Accepted: 6 March 2020 / Published online: 18 March 2020

Abstract

The gene *cox1* is one of the most reported mitochondrial genes involved in horizontal gene transfer among angiosperms. However, whether different *cox1* copies exist in different populations of a species and whether any other novel way except intron homing exists for *cox1* intron acquisition is less understood. In this study, we chose *Cassytha filiformis*, a parasitic plant from the angiosperm family Lauraceae, as an example to study *cox1* variation and evolution. We identified the stable and inheritable co-occurrence of two copies of *cox1* genes, which were different in base composition and insertion/deletion among samples of a single species, *C. filiformis*. The bioinformatic analyses revealed that Type I copy had intact open reading frames, but type II copy had premature stop codons and was a pseudogene. Further INDEL characterization, phylogenetic analyses, and CCT comparisons consistently support two different origins for the two types of *C. filiformis cox1* genes. Type I *cox1* was likely vertically inherited within the magnoliids but it has captured an intron from another species, whereas the entire type II intron-containing *cox1* has most likely been transferred integrally from *Cuscuta* or other Convolvulaceae species. The finding of the two independent horizontal gene transfer events associated with *C. filiformis cox1* genes not only promotes our understanding of the evolutionary history of *C. filiformis*, but also leaves intriguing evolutionary questions that merits further efforts.

Keywords *cox1 intron* · *Cassytha filiformis* · Horizontal gene transfer · Parasitism

Abbreviations

CCT	Co-conversion tract
HGT	Horizontal gene transfer
INDEL	INsertion/DELetion

Introduction

Horizontal gene transfer is the transmission of genetic material between different organisms or between different cytoplasmic organelles and nucleus through asexual processes. It plays an important role in the evolution of many organisms. For instance, HGT is the main cause of rapid antibiotic resistance circulation among bacteria (Koonin et al. 2001; Gyles and Boerlin 2014; Kay et al. 2002). HGT is not only broadly present in the prokaryotic world, but also increasingly reported in eukaryotes (Keeling and Palmer 2008). In land plants, massive HGT has been discovered through

Handling editor: **Alan Christensen**.

Canyu Zhang and Hui Ma have contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00239-020-09937-1) contains supplementary material, which is available to authorized users.

✉ Jie Li
jieli@xtbg.ac.cn

[1] Plant Phylogenetics and Conservation Group, Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Kunming 650223, Yunnan, People's Republic of China

[2] University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China

[3] IBAM, Universidad Nacional de Cuyo, CONICET, Facultad de Ciencias Agrarias, Almirante Brown 500, M5528AHB Chacras de Coria, Argentina

[4] Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Cuyo, 5500 Mendoza, Argentina

[5] Center of Conservation Biology, Core Botanical Gardens, Chinese Academy of Sciences, Mengla 666303, People's Republic of China

the use of in-depth sequencing in a few species, such as *Amborella trichopoda*, *Geranium brycei*, *Rafflesia cantleyi*, *Sapria himalayana,* and *Lophophytum mirabile* (Bergthorsson et al. 2004; Molina et al. 2014; Park et al. 2015; Rice et al. 2013; Sanchez-Puerta et al. 2017; Xi et al. 2012, 2013). Besides, extensive HGT has been shown to promote plant colonization of land (Yue et al. 2012). With the sequencing of more plant genomes, the widespread footprints of HGT in additional plants will be gradually uncovered. The frequency of identified HGT is much higher in mitochondria than in plastids and nuclei, and a large fraction of the HGT reports come from parasitic plants and their hosts (Davis and Xi 2015; Keeling and Palmer 2008; Sanchez-Puerta 2014). Among mitochondrial genes, *cox1* is the most frequently implicated in HGT (Cho et al. 1998; Sanchez-Puerta et al. 2008).

The mitochondrial gene *cox1* encodes the cytochrome c oxidase subunit I required to constitute the respiratory complex IV and is essential for oxidative phosphorylation (Toffaletti et al. 2003). In most vascular plants, it does not have an intron, while in a sizeable fraction of angiosperms its exon is interpolated by an intron sequence. Indeed, *cox1* introns have invaded nearly every angiosperm parasitic lineage, leaving *Krameria* and *Schoepfia* as the only two sampled parasitic plants lacking *cox1* introns (Barkman et al. 2007). The *cox1* intron encodes a site-specific DNA endonuclease, which facilitates its propagation (Delahodde et al. 1989). Homing is used to describe such a phenomenon. Intron homing, the introduction of an intron into a homologous allele lacking it, has been proposed to proceed by the double strand-break repair pathway (Lambowitz and Belfort 1993). During this process, part of the foreign exonic regions immediately flanking the invading intron often engages in a gene conversion activity that replaces part of the recipient exonic sequence (Delahodde et al. 1989; Lambowitz and Belfort 1993; Mueller et al. 1996; Wenzlau et al. 1989). A region of converted exonic sequence is called a "co-conversion tract" (CCT). If the flanking exon sequences of the donor and recipient plants differ, then the repair process will create a "footprint" (CCT) that can remain even after the intron itself is lost again (Cho and Palmer 1999).

The sporadic distribution of the *cox1* intron among angiosperms is attributed to HGT via the above intron homing mechanism in most cases (Barkman et al. 2007; Cho et al. 1998; Sanchez-Puerta et al. 2011, 2008), and intron loss usually via a retrotranscribed copy of a mature *cox1* transcript (Sanchez-Puerta et al. 2008). As *cox1* introns are more frequently found in parasitic plants, we choose *Cassytha*, the only parasitic genus with 10–20 species in the family Lauraceae, as the target system to study *cox1* evolution. According to the Flora of China, only one pantropical species, the hemiparasite *C. filiformis*, distributes in China. *C. filiformis* has a wide range of hosts, with more than 100 host species in Guangxi alone (Li et al. 1992). The known hosts in China according to our field surveys and from the literature are summarized in Table S1. The intimate connection of *C. filiformis* with its host through haustoria and the wide host range grant a large potential of genetic flow.

According to NCBI nucleotide databases, *cox1* sequences from *C. filiformis* and other Lauraceae species have been reported. A comparison of their sequences suggested that *C. filiformis* has an intron in its *cox1*, while other Lauraceae species have not. It is not clear whether the *cox1* intron was acquired exclusively by *C. filiformis*, or whether it was lost in other members of the family. In this study, we generated *cox1* sequences from different *C. filiformis* samples collected from three distant places and from 32 other species from different lineages within the family Lauraceae, and analyzed the *cox1* sequences from a wide diversity of angiosperms. We aimed to achieve the following objectives: (1) to test whether different *cox1* copies exist in *C. filiformis*; (2) to examine whether *cox1* intron in *C. filiformis* has been retained from a common Lauraceae ancestor or was horizontally transferred from other non-Lauraceae species; and (3) to understand the evolutionary history of the *cox1* genes in *Cassytha*.

Methods

Sampling and Sequencing

We collected *C. filiformis* stem samples at least two centimeters away from the host to prevent contamination. Leaf and stem samples from other 32 Lauraceae species were also collected (Table S2). Total genomic DNA was extracted with the Plant Genomic DNAKit (Tiangen Biotech, China). The *cox1* genes in *C. filiformis* were amplified by PCR using two primers, *cox1* intron-F (5′-CATCTCTTTYTGTTCTTC GGT-3′) and *cox1* intron-R (5′-AGCTGGAAGTTCTCC AAAAGT-3′) (Sanchez-Puerta et al. 2008). Another set of primers designed by Primer Premier 5 (Lalitha 2000), *cox1* exon-F (5′-GTATGGAATTAGCACGACCCG-3′) and *cox1* exon-R (5′-TACGACCACGAAG GAACGAC-3′), were used to amplify *cox1* genes in *C. filiformis* as well as 32 Lauraceae species under study. The PCR mixtures for *cox1* amplification were 2.5 µl of 10 × PCR reaction buffer (Takara, Japan), 1.5 µl of 25 mM $MgCl_2$, 1 µl of each primer (Shanghai Sangon, China) at 10 ng/µl, 1 µl of 2.5 mM dNTP solution in an equimolar ratio, 0.2 µl of Taq DNA-polymerase (5 U/µl, Takara, Japan), 2 µl of genomic DNA at 5 ng/µl, and ddH_2O to reach a total volume of 25 µl. The amplified products were purified using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). In the genus *Cassytha*, PCR products of *Cassytha* sp. and *C. pubescens* were successfully sequenced, but the *C. filiformis* appeared polymorphisms. In order to

assess whether different *cox1* genes exist in *C. filiformis*, we further cloned PCR products from *C. filiformis* using the pEASY-T3 Cloning Kit (TransGen Biotech, Beijing, China). Between 2 and 6 clones were sequenced for each individual. All fragments were sequenced in both directions using BigDye 3.1 reagents with an ABI 3770 automated sequencer (Applied Biosystems, Carlsbad, California, USA). All sequences are deposited in GenBank (Table S2).

Sequence Alignment and Phylogenetic Analyses

All sequence alignments were manually edited using Geneious v6 (Kearse et al. 2012). Homologous *cox1* sequences were identified using BLASTN against NCBI Non-Redundant Nucleotide Database (Tables S3 and S4). In order to remove the influence of CCTs and editing sites on phylogenetic analyses, we excluded 30 bp of the exon downstream the intron insertion site and the predicted editing sites. Multiple sequence alignments of the *cox1* coding sequences and introns were performed with MAFFT (Katoh et al. 2017) and manually adjusted. Phylogenetic analyses were performed on the aligned sequences of *cox1* exons and *cox1* introns, respectively. The maximum likelihood (ML) analyses were performed in RAxML-HPC BlackBox via CIPRES (Miller et al. 2010) and RAxML under the general time reversible model with parameters for invariable sites and gamma-distributed rate heterogeneity (GTR + I + G; 4 rate categories, 1000 bootstraps).

Results

Co-existence of Two *cox1* Copies in *C. filiformis*

In order to investigate whether different forms of the mitochondrial gene *cox1* exist in distinct populations of *C. filiformis*, we collected stem samples of *C. filiformis* from three geographically distant places in China, i.e., Shenzhen, Nanning, and Xishuangbanna, amplified the *cox1* sequences, and sequenced them. Since *cox1* variation across angiosperms is mainly found in the presence/absence of the *cox1* intron and the co-conversion tract, we amplified this region (~ 1140 bp fragments). As the initial sequencing of *cox1* introns was quite unsuccessful due to the presence of multiple peaks, we resorted to gene cloning of the PCR products. The subsequent sequencing of the cloned fragments revealed clearly that two distinct copies of *cox1* introns were present in the same samples.

Although we expected *cox1* variation in terms of the absence or presence of the intron, the identification of two different *cox1* intron sequences in the same samples was surprising. We wondered whether the exons of the two *cox1* also differed, thus we amplified and cloned the whole *cox1* (~ 2221 bp) from additional samples for further sequencing. Two distinct copies of *cox1* genes were identified in most samples of *C. filiformis* collected from three different places, suggesting a stable, inheritable nature of both *cox1* alleles in *C. filiformis*. Moreover, the frequency of co-existence of these two *cox1* alleles was very high as they were detected in 18 out of 20 samples (Table 1). Only one type of *cox1* was detected in samples M6 and M48, probably due to insufficient clone sample. We then amplified the DNA and directly sequenced the PCR products from these two samples and found no polymorphisms when sequencing the non-clonal PCR products, which verified the existence of only type I *cox1* gene (Table 1).

The two alleles of *cox1* in *C. filiformis* differ strikingly in their exon and intron sequences and length (Table 2). Their intron sequence identity is only 84.6%, suggesting that the two copies of *cox1* should come from completely different origins. Further bioinformatic analysis on the two *cox1* genes of *C. filiformis* revealed that the exons of type I *cox1* have an intact open reading frame, whereas the exons of type II *cox1* have premature stop codons, which likely produces a much shorter malfunctional protein. The intron of type I *cox1* is 967 bp in length and has a full-length open reading frame of 921 bp, encoding a homing endonuclease. Similar to the *cox1* exons, the intron of type II *cox1* is 912 bp in length but contains several nonsense mutations (Table 2).

A BLAST search against Genbank databases showed that type I *cox1* is similar to *cox1* in magnoliids species, whereas type II *cox1* displays very high similarity to *cox1* in *Cuscuta japonica*, a Convolvulaceae species. Moreover, the two copies of *cox1* in *C. filiformis* show contrasting Insertion/Deletion (INDEL) in multiple sequence alignments. These two *cox1* differ in 12 INDEL loci, whereas type II *cox1* in *C. filiformis* and *cox1* in *C. japonica* are nearly identical at 10 of these loci (Fig. 1). These ten shared INDELs are unique as they are absent from the *cox1* in all other species under study. Therefore, INDEL comparisons further support the different origins of the two copies of *cox1*.

The Different Origins of the Two *C. filiformis cox1* Genes

In order to track down the exact origins of the *cox1* genes in *C. filiformis*, we carried out phylogenetic analyses on *cox1* from many other Lauraceae species as well as a diverse range of angiosperms. We systematically studied *cox1* genes from many other Lauraceae species. We took leaf samples from 30 other species, representing 16 Lauraceae genera distributed in China. Besides, we also included four Australian *Cassytha* stem samples, three from *C. pubescens* and another from an unidentified *Cassytha* sp. (Table S2). We also downloaded *cox1* sequences of seven other Lauraceae species from the NCBI Nucleotide databases (Table S3).

Table 1 Occurrence of the two types of *cox1* in the sequenced clones of *C. filiformis* samples

Sample ID	*N*	type I	type II	Host	Location	Latitude	Longitude
CaM	4	1	3	*Aporosa* sp.	Xishuangbanna, China	N21° 41′	E101° 25′
CaT	6	1	5	*Litsea rotundifolia*	Xishuangbanna, China	N21° 41′	E101° 25′
N3B	5	4	1	*Melicope pteleifolia*	Nanning, China	N108° 18′ 42″	E22° 44′ 51″
N9B	6	5	1	*Dicranopteris pedata*	Nanning, China	N108° 18′ 42″	E22° 44′ 51″
N13B	5	2	3	*Elaeocarpus decipiens*	Nanning, China	N108° 18′ 42″	E22° 44′ 51″
N20B	6	5	1	*Melastoma malabathricum*	Nanning, China	N108° 18′ 42″	E22° 44′ 51″
N21B	6	2	4	*Lygodium japonicum*	Nanning, China	N108° 18′ 42″	E22° 44′ 51″
N23B	3	1	2	*Ficus* sp.	Nanning, China	N108° 23′ 23″	E22° 47′ 12″
N27B	6	5	1	Melicope pteleifolia	Nanning, China	N108° 23′ 23″	E22° 47′ 12″
N28B	5	1	4	*Melicope pteleifolia*	Nanning, China	N108° 23′ 23″	E22° 47′ 12″
CP	5	4	1	*Psychotria asiatica*	Nanning, China	N108° 23′ 23″	E22° 47′ 12″
M3	6	4	2	*Stephania macrantha*	Shenzhen, China	N114° 12′ 58″	E22° 35′ 14″
M6	6	6	0	*Salix myrtillacea*	Shenzhen, China	N114° 12′ 58″	E22° 35′ 14″
M13	5	2	3	*Melastoma sanguineum*	Shenzhen, China	N114°12′ 58″	E22° 35′ 14″
M21	3	2	1	*Aporosa dioica*	Shenzhen, China	N114° 12′ 58″	E22° 35′ 14″
M24	4	2	2	*Melicope pteleifolia*	Shenzhen, China	N114° 12′ 58″	E22° 35′ 14″
M39	4	2	2	*Bridelia tomentosa*	Shenzhen, China	N113° 57′ 53″	E22° 39′ 8″
M48	6	6	0	*Acacia* sp.	Shenzhen, China	N113° 57′ 53″	E22° 39′ 8″
M50	2	1	1	*Engelhardia roxburghiana*	Shenzhen, China	N113° 57′ 53″	E22° 39′ 8″
M60	6	5	1	*Schima superba*	Shenzhen, China	N113° 57′ 53″	E22° 39′ 8″

[a]*N* (The number of successfully sequenced clones), type I (The number of successfully sequenced clones of *C. filiformis* type I), type II (The number of successfully sequenced clones of *C. filiformis* type II)

Table 2 Sequence comparison of the two types of *cox1* in *C. filiformis*

	Sequence length (bp)		Sequence identity between the two types of *cox1* (%)
	Type I *cox1*	Type II *cox1*	
Complete sequence	2187	2136	89.90
Exon1	573	572	97.03
Intron	967	912	85.04
Exon2	647	652	93.53

Unlike the other 16 genera of the family Lauraceae, the genus *Cassytha* is the only one that harbors introns in its *cox1* genes. The difference is that *C. filiformis* has two types of *cox1* genes, while the four Australian *Cassytha* samples have no polymorphisms when sequencing non-clonal PCR products and were shown to contain only the type I *cox1*, suggesting that the introduction of type I *cox1* intron probably took place before the speciation of *C. filiformis*. It is unsure whether type II *cox1* is unique to some local Chinese *C. filiformis* populations, transferred horizontally after the split of *C. filiformis* and *C. pubescens*, or it is found in other *Cassytha* species and was lost randomly in certain populations. The sequencing of more *Cassytha* samples is required to answer this question.

We analyzed the exon and intron trees individually because *cox1* introns are frequently involved in horizontal gene transfer and often show significant phylogenetic incongruence in comparison to *cox1* exons. The phylogenetic tree based on the exon sequences (Fig. 2) suggests that the two alleles of *cox1* of *C. filiformis* have completely different origins. The *C. filiformis* type II displays very high affinity to *cox1* in *Cuscuta* spp. and *Ipomoea* spp., two Convolvulaceae species, suggesting a foreign origin of this allele. On the other hand, the *C. filiformis* type I is phylogenetically close to those in magnoliids, consistent with a vertical inheritance of this *cox1* coding sequence.

In contrast to the observed in the exon phylogeny, the *cox1* introns of all *Cassytha* species cluster in a single clade including both *cox1* alleles in *C. filiformis*, although not as sister taxa (Fig. 3). According to Figs. 2 and 3, it is clear that the phylogenetic positions of exons and intron of type II *C. filiformis cox1* do not change much, as both form a monophyletic clade with *Cuscuta* spp. and *Ipomoea* spp. However, the intron tree shows a sister relationship between *C. filiformis* type II *cox1* and *C. japonica* with 100% of bootstrap support. These results suggest that both the exons and intron of type II *C. filiformis cox1* might share an origin with *cox1* genes from *Cuscuta* spp.

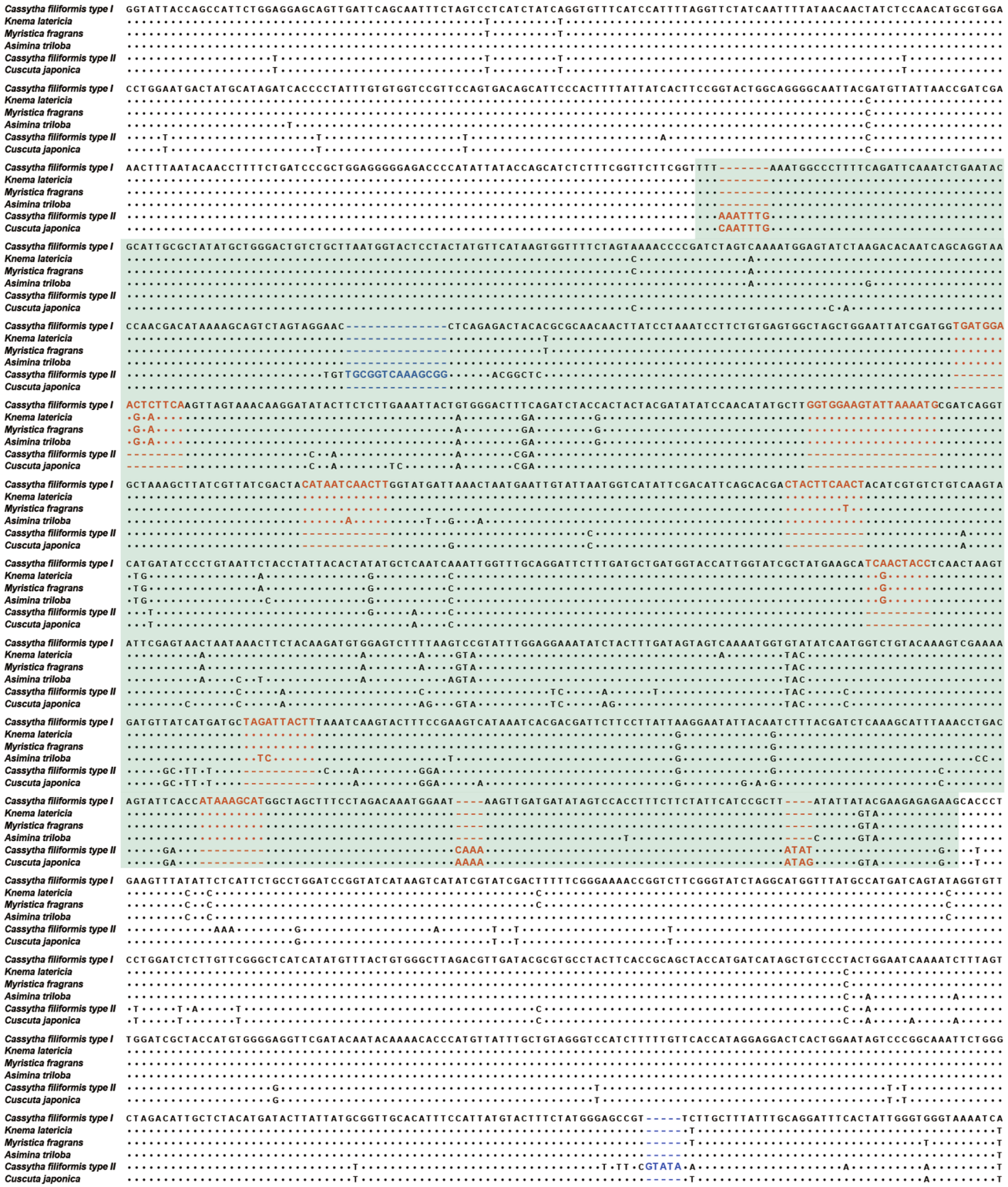


Fig. 1 Partial sequence alignment of the two types of *cox1* in *Cassytha filiformis* and four other *cox1* homologs. For comparison, we included *cox1* in *Cuscuta japonica* and three Magnoliales species, *Asimina triloba*, *Knema latericia,* and *Myristica fragrans*. The ten shared INDEL positions by type II *cox1* in *C. filiformis* and *cox1* in *C. japonica* are in red. The two INDEL positions unique to *C. filiformis* type II *cox1* are in blue. A shaded box indicates the intron sequence (Color figure online)

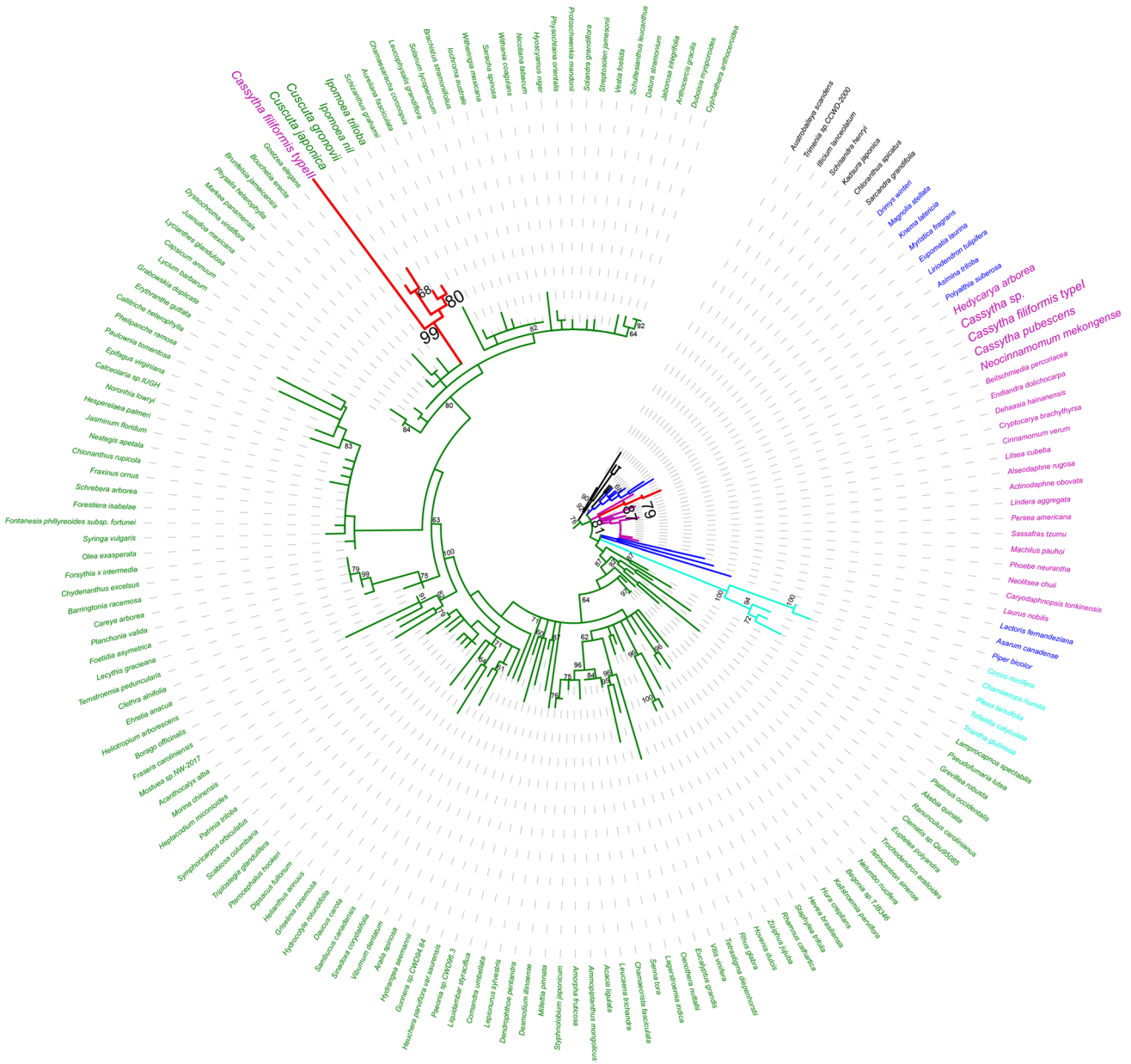


Fig. 2 ML tree of 173 species based on *cox1* exons analyzed under a GTR + I + G model. Only ML bootstrap values > 60% are displayed. Species that belong to monocots, eudicots, Lauraceae, and other magnoliids except Lauraceae are in cyan, green, magenta, and blue, respectively. The branches leading to the two types of *C. filiformis cox1* are colored red (Color figure online)

Interestingly, the phylogenetic position of *C. filiformis* type I *cox1* and other *Cassytha* species *cox1* differs considerably in the exon tree and the intron trees. *C. filiformis* type I *cox1* and other *Cassytha* species *cox1* cluster in the clade of Lauraceae and other basal magnoliids species in the exon tree, while they are related to diverse eudicots, in particular, *Cuscuta japonica*, *Ipomoea* spp. and *Calceolaria* sp. with 80% of bootstrap support, according to the intron tree. The phylogenetic incongruence of *C. filiformis* type I *cox1* and other *Cassytha* species *cox1* exons and introns are similar to other cases of horizontal gene transfer, where an exogenous *cox1* intron invaded the native *cox1* copy via intron homing.

Furthermore, *cox1* introns are accompanied by a characteristic co-conversion tract (CCT) when the exon of the donor is different from that of the recipient plant (Cho and Palmer 1999; Cho et al. 1998; Sanchez-Puerta et al. 2011, 2008), Cusimano et al. (2008) grouped CCTs of all available

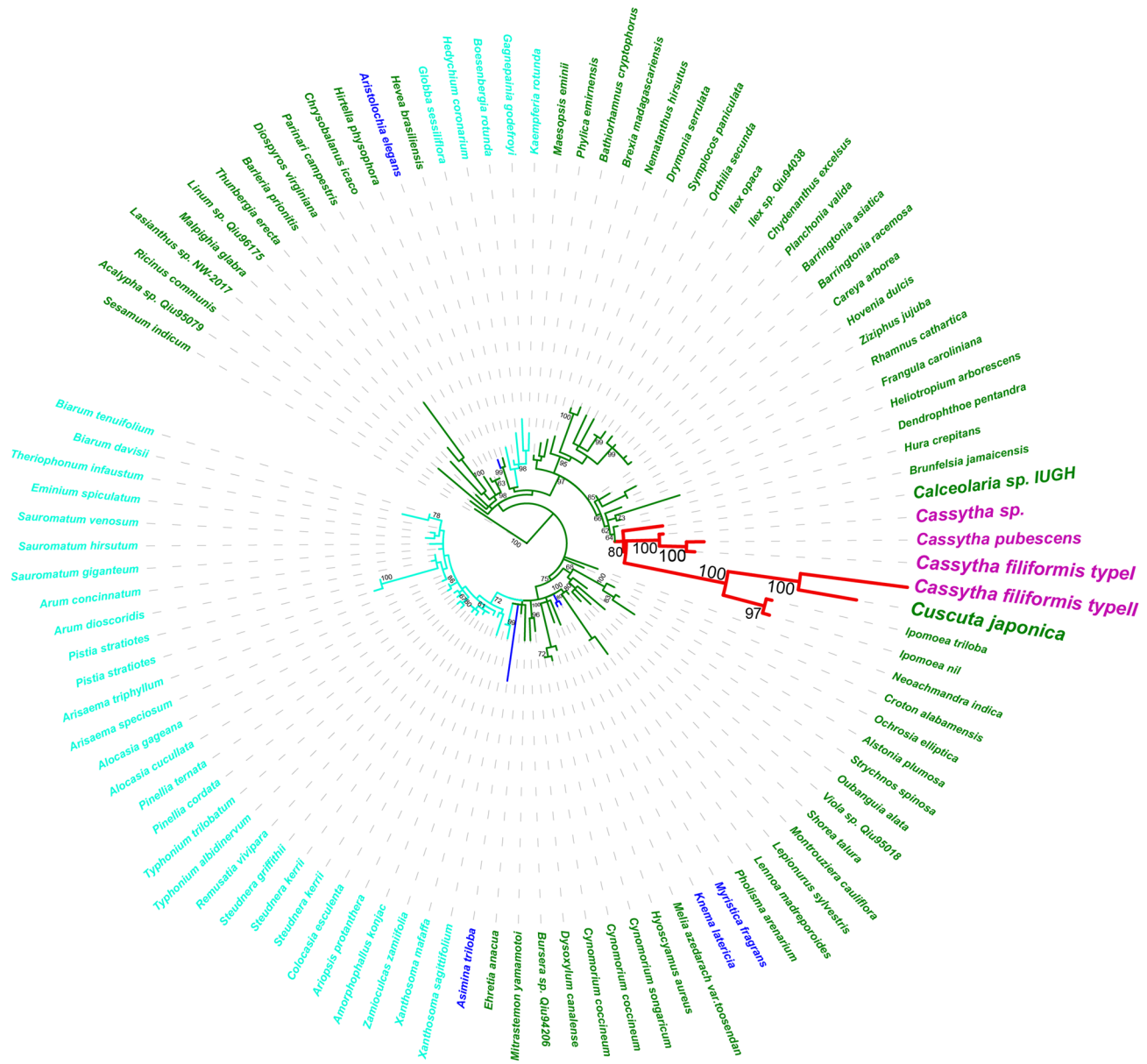
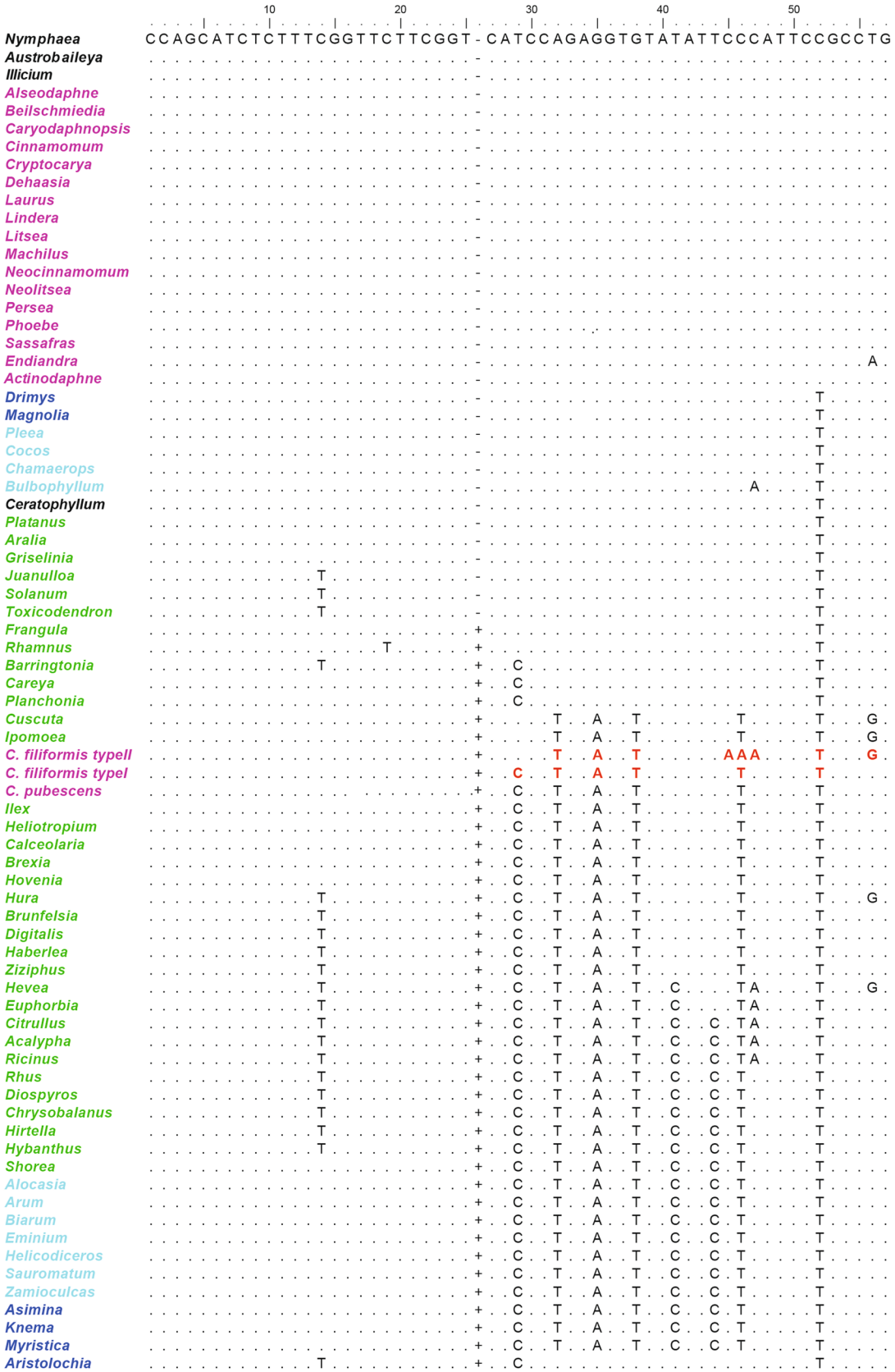


Fig. 3 ML tree of 103 species based on *cox1* introns analyzed under a GTR + I + G model. Only ML bootstrap values > 60% are displayed. Species that belong to monocots, eudicots, Lauraceae, and other magnoliids except Lauraceae are in cyan, green, magenta, and blue, respectively. The branches leading to the two types of *C. filiformis cox1* are colored red (Color figure online)

angiosperm *cox1* sequences into 20 types. We thoroughly compared the exon sequences flanking the intron insertion site of all the newly sequenced *cox1*, as well as homologous *cox1* sequences downloaded from NCBI databases (Fig. 4). It is clear from Fig. 4 that the species of the family Lauraceae, except for *Cassytha* spp., lack the *cox1* intron and the characteristic CCT. These observations suggest that the absence of the *cox1* intron in these Lauraceae species is not due to intron loss and that *C. filiformis* obtained the *cox1* introns by HGT after its divergence from other Lauraceae. The type I *cox1* of *C. filiformis* has an intron and a 26 bp-long CCT that it is also observed in *Cassytha* spp., as well as in a few other species that show affinity to *Cassytha* spp. in the intron phylogeny. The type II *cox1* of *C. filiformis* has an extended CCT of 30 bp shared with *C. japonica* and *Ipomoea* spp. Therefore, the two types of CCTs and the intron phylogeny support the two independent evolutionary origins of *C. filiformis cox1* genes.


10
20
30
40
50
Nymphaea CCAGCATCTCTTTCGGTTCTTCGGT-CATCCAGAGGTGTATATTCCCATTCCGCCTG
Austrobaileya
Illicium
Alseodaphne
Beilschmiedia
Caryodaphnopsis
Cinnamomum
Cryptocarya
Dehaasia
Laurus
Lindera
Litsea
Machilus
Neocinnamomum
Neolitsea
Persea
Phoebe
Sassafras
Endiandra
Actinodaphne
Drimys
Magnolia
Pleea
Cocos
Chamaerops
Bulbophyllum
Ceratophyllum
Platanus
Aralia
Griselinia
Juanulloa
Solanum
Toxicodendron
Frangula
Rhamnus
Barringtonia
Careya
Planchonia
Cuscuta
Ipomoea
C. filiformis typell
C. filiformis typel
C. pubescens
Ilex
Heliotropium
Calceolaria
Brexia
Hovenia
Hura
Brunfelsia
Digitalis
Haberlea
Ziziphus
Hevea
Euphorbia
Citrullus
Acalypha
Ricinus
Rhus
Diospyros
Chrysobalanus
Hirtella
Hybanthus
Shorea
Alocasia
Arum
Biarum
Eminium
Helicodiceros
Sauromatum
Zamioculcas
Asimina
Knema
Myristica
Aristolochia

◂**Fig. 4** Sequence comparisons of *cox1* sequences flanking the intron insertion site in 75 angiosperms. Species that belong to monocots, eudicots, Lauraceae, and other magnoliids except Lauraceae are in cyan, green, magenta, and blue, respectively. The co-conversion tracts (CCTs) of two types of *C. filiformis cox1* are colored red. Plus (+) and minus (−) symbols indicate *cox1* intron presence and absence, respectively (Color figure online)

Discussion

The Structure and Frequent Co-existence of Two *cox1* Alleles in *C. filiformis*

In this study, we identified two different copies of the gene *cox1* in individual samples of *C. filiformis*. The exon sequences of *C. filiformis* type I *cox1* have an intact open reading frame that encodes the cytochrome c oxidase subunit I, in agreement with earlier studies (Barkman et al. 2007). The intron of the type I *cox1* encodes a putative functional homing endonuclease, because of the presence of two LAGLI-DADG motifs and an intact open reading frame, which may be involved in intron propagation and splicing (Belfort and Perlman 1995). In contrast, *C. filiformis* type II *cox1* is a pseudogene given that we identified several stop codons within the exons and also in the intron sequence. All sequenced samples of *C. filiformis* contain the type I copy, and a few samples lack the type II copy, indicating that only type I *cox1* is essential and type II copies might have escaped from functional constraints. Also, the *cox1* type II copy was not found in other species of the same genus. The presence of two *cox1* alleles, one of which is a pseudogene has been previously described in *Geranium brycei* (Park et al. 2015).

In this study, a population-level study found that the co-existence of these two alleles in *C. filiformis* was quite frequent, as they were found in more than 90% of the 20 individuals of *C. filiformis* analyzed. Anyway, the co-occurrence of two *cox1* alleles, either in a single mitochondrial genome or in different mitochondria or cells of the stem of *C. filiformis* is outstanding and deserves further investigation. The origin of each of the *cox1* alleles in *C. filiformis* may be explained by its increased chance as a parasitic plant to exchange genetic information with its hosts and a greater flexibility in genome evolution after adopting a parasitic lifestyle (Davis and Xi 2015; Sanchez-Puerta 2014). It is also possible that the co-existence of different *cox1* alleles, or other mitochondrial genes, in other species is underestimated due to the limited sampling at the population level or difficulty to detect additional gene copies at lower stoichiometries. In either case, deeper sequencing on a wider range of plants and increased population sampling are required to evaluate the co-occurrence of *cox1* alleles in different species as well as the ecological and evolutionary importance of *cox1* heterozygosity.

Two Independent HGT Events

The acquisition of foreign DNA has been predicted to be a key event in the evolution of angiosperms (Atsatt 1973), and *cox1* intron could represent a marker of a genomically more widespread historical transformation (Barkman et al. 2007). The numerous angiosperm-to-angiosperm transfers of *cox1* intron and its outstanding evolutionary history have sparked the interest of several researchers. Cho et al. (1998) and Sanchez-Puerta et al. (2008) analyzed all available *cox1* data from angiosperms and confirmed that the *cox1* intron has been horizontally acquired numerous times during angiosperm evolution. For example, the *cox1* intron was acquired by horizontal transfer in at least three separate occasions during the evolution of the Solanaceae (Sanchez-Puerta et al. 2011). The opposite view argued that *cox1* intron loss is a predominant factor in *cox1* evolutionary history in Araceae (Cusimano et al. 2008). Moreover, for the first time, two copies of the *cox1* gene which differ in intron content were found in *Geranium brycei* mitochondria and supported the notion of repeated, independent HGT (Park et al. 2015).

In our study, we also found two different copies of the *cox1* gene in *C. filiformis*. In addition to the well-documented *cox1* intron homing, we also identified exons involved in horizontal gene transfer. In fact, the phylogenetic analyses of *cox1* exons and introns revealed a different origin of the two intron-containing *cox1* alleles of *C. filiformis*. One full-length copy had been clearly acquired by horizontal gene transfer from the lineage Convolvulaceae and it is a pseudogene in *C. filiformis*. The *cox1* coding regions of *Cuscuta* spp. and *Ipomoea* spp. have intact open reading frames and encode the cytochrome c oxidase subunit I. Therefore, the pseudogenization of type II *cox1* of *C. filiformis* may have taken place after the horizontal gene transfer event. The other copy has vertically inherited exons and a horizontally transferred intron. The intron phylogeny shows a close relationship to the *cox1* introns of *Calceolaria* spp., *Ipomoea* spp., *Cuscuta japonica*, and the foreign copy of *C. filiformis* and this copy shares a 20 bp-long CCT with them and other angiosperms. In addition, the other genera of Lauraceae analyzed show a single *cox1* allele that lacks the intron and CCT. The *cox1* exons of all Lauraceae, including *C. filiformis* type I and *Cassytha* spp., are highly similar to each other.

The two independent HGT events reveal a highly dynamic mitochondrial genome in *C. filiformis* and raised more questions. For instance, did the intron of type II *cox1* of *C. filiformis* invade the native *cox1* gene? Did recombination take place between the two different copies of *cox1* genes? The fact that both *C. filiformis cox1* introns are found in a single clade in the intron phylogeny and are associated to a similar CCT opens the possibility of an intracellular intron invasion from the type II *cox1* allele to the type I *cox1*

allele. However, the high sequence divergence observed in the type II *cox1* intron, shared with *Cuscuta* and *Ipomoea* in comparison to the more conserved type I *cox1* intron sequence argues strongly against it. In contrast, it suggests a second horizontal acquisition from a donor containing an intron related to those of the Convolvulaceae. By analyzing the *cox1* alignment in detail, we could not find evidence of recombination between the two *cox1* alleles in *C. filiformis*.

Further questions remain unanswered, such as when did these two HGT events happen? Were *cox1* genes in other *Cassytha* species acquired from additional donors? How did the HGT events influence the evolution of *C. filiformis*? Since *Cassytha* and *Cuscuta* are both parasitic plants and HGT can promote adaptation to parasitism, it is tempting to ask whether any parasitism-related gene would be exchanged in the HGT event. The answers to the above questions will greatly contribute to our understanding of the mitochondrial dynamics in *C. filiformis* and historical events during its parasitism evolution.

Conclusions

In this study, we investigated *cox1* evolution in a parasitic Lauraceae species, *C. filiformis*. We found consistent co-existence of two different *cox1* alleles in 90% of the samples of *C. filiformis* collected from distant locations around China, and demonstrated clearly the different origins of the two types of *cox1* genes as well as the implications of two independent horizontal transfer events. Our study deepens our understanding of the complicated evolutionary histories of *C. filiformis cox1* and the highly dynamic mitochondrial genome in this parasitic plant.

Acknowledgements The authors thank Jianmei Wu (nature photographer, Tongya Seedbank, Shenzhen, China) for sample collection and Jipei Yue (Kunming Institute of Botany, Chinese Academy of Sciences) for manuscript proofreading. This work was supported by the CAS 135 program (Grant No. 2017XTBG-T03), National Natural Science Foundation of China (Grant Number 31500311), and CAS "Light of West China" Program.

Data Availability The *cox1* fragments sequenced in our study were deposited in NCBI nucleotide database under the accession numbers MH093679-MH093710 and MT010850-MT010851. GenBank accession numbers of *C. filiformis* type I and type II *cox1* are MH093709 and MH093710, respectively.

References

Atsatt PR (1973) Parasitic flowering plants: how did they evolve? Am Nat 107:502–510

Barkman TJ, McNeal JR, Lim SH, Coat G, Croom HB, Young ND, dePamphilis CW (2007) Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. BMC Evol Biol 7:248

Belfort M, Perlman PS (1995) Mechanisms of intron mobility. J Biol Chem 270:30237–30240

Bergthorsson U, Richardson AO, Young GJ, Goertzen LR, Palmer JD (2004) Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*. Proc Natl Acad Sci USA 101:17747–17752

Cho Y, Palmer JD (1999) Multiple acquisitions via horizontal transfer of a group I intron in the mitochondrial *cox1* gene during evolution of the Araceae family. Mol Biol Evol 16:1155–1165

Cho Y, Qiu YL, Kuhlman P, Palmer JD (1998) Explosive invasion of plant mitochondria by a group I intron. Proc Natl Acad Sci USA 95:14244–14249

Cusimano N, Zhang LB, Renner SS (2008) Reevaluation of the *cox1* group I intron in Araceae and angiosperms indicates a history dominated by loss rather than horizontal transfer. Mol Biol Evol 25:265–276

Davis CC, Xi Z (2015) Horizontal gene transfer in parasitic plants. Curr Opin Plant Biol 26:14–19

Delahodde A, Goguel V, Becam AM, Creusot F, Perea J, Banroques J, Jacq C (1989) Site-specific DNA endonuclease and RNA maturase activities of two homologous intron-encoded proteins from yeast mitochondria. Cell 56:431–441

Gyles C, Boerlin P (2014) Horizontally transferred genetic elements and their role in pathogenesis of bacterial disease. Vet Pathol 51:328–340

Katoh K, Rozewicki J, Yamada KD (2017) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform 20:1160–1166

Kay E, Vogel TM, Bertolla F, Nalin R, Simonet P (2002) In situ transfer of antibiotic resistance genes from transgenic (transplastomic) Tobacco plants to bacteria. Appl Environ Microbiol 68:3345–3351

Kearse M et al (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649

Keeling PJ, Palmer JD (2008) Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet 9:605

Koonin EV, Makarova KS, Aravind L (2001) Horizontal gene transfer in prokaryotes: quantification and classification. Annu Rev Microbiol 55:709–742

Lalitha S (2000) Primer premier 5. Biotech Softw Internet Rep 1:270–272

Lambowitz AM, Belfort M (1993) Introns as mobile genetic elements. Annu Rev Biochem 62:587–622

Li Q, Yao D, Cai J, Huang L (1992) A field survey of *Cassytha filiformis* host range. Guangxi Plant Protection 4:21–24

Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: grid computing environments. Accessed 14 Nov 2010

Molina J et al (2014) Possible loss of the chloroplast genome in the parasitic flowering plant *Rafflesia lagascae* (Rafflesiaceae). Mol Biol Evol 31:793–803

Mueller JE, Smith D, Belfort M (1996) Exon coconversion biases accompanying intron homing: battle of the nucleases. Genes Dev 10:2158–2166

Park S, Grewe F, Zhu A, Ruhlman TA, Sabir J, Mower JP, Jansen RK (2015) Dynamic evolution of *Geranium* mitochondrial genomes through multiple horizontal and intracellular gene transfers. New Phytol 208:570–583

Rice DW et al (2013) Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. Science 342:1468–1473

Sanchez-Puerta MV (2014) Involvement of plastid, mitochondrial and nuclear genomes in plant-to-plant horizontal gene transfer. Acta Soc Bot Pol 83:317–323

Sanchez-Puerta MV, Cho Y, Mower JP, Alverson AJ, Palmer JD (2008) Frequent, phylogenetically local horizontal transfer of the *cox1* group I intron in flowering plant mitochondria. Mol Biol Evol 25:1762–1777

Sanchez-Puerta MV, Abbona CC, Zhuo S, Tepe EJ, Bohs L, Olmstead RG, Palmer JD (2011) Multiple recent horizontal transfers of the *cox1* intron in Solanaceae and extended co-conversion of flanking exons. BMC Evol Biol 11:277

Sanchez-Puerta MV, García LE, Wohlfeiler J, Ceriotti LF (2017) Unparalleled replacement of native mitochondrial genes by foreign homologs in a holoparasitic plant. New Phytol 214:376–387

Toffaletti DL, Poeta MD, Rude TH, Dietrich FS, Perfect JR (2003) Regulation of cytochrome c oxidase subunit 1 (COX1) expression in *Cryptococcus neoformans* by temperature and host environment. Microbiology 149:1041–1049

Wenzlau JM, Saldanha RJ, Butow RA, Perlman PS (1989) A latent intron-encoded maturase is also an endonuclease needed for intron mobility. Cell 56:421–430

Xi Z et al (2012) Horizontal transfer of expressed genes in a parasitic flowering plant. BMC Genomics 13:227–227

Xi Z, Wang Y, Bradley RK, Sugumaran M, Marx CJ, Rest JS, Davis CC (2013) Massive mitochondrial gene transfer in a parasitic flowering plant clade. PLoS Genet 9:2

Yue J, Hu X, Sun H, Yang Y, Huang J (2012) Widespread impact of horizontal gene transfer on plant colonization of land. Nat Commun 3:1152