Chia-Chen Wu, Fang-Hua Chu, Cheng-Kuen Ho, Chia-Hsuan Sung and Shu-Hwa Chang*

Comparative analysis of the complete chloroplast genomic sequence and chemical components of *Cinnamomum micranthum* and *Cinnamomum kanehirae*

DOI 10.1515/hf-2016-0133
Received August 25, 2016; accepted November 17, 2016; previously published online January 18, 2017

Abstract: *Cinnamomum micranthum* (Hayata) Hayata is a tree species that is often confused with *Cinnamomum kanehirae* Hayata, which is an endemic species in Taiwan and the sole natural host of the valuable medicinal fungus *Antrodia cinnamomea*. However, the two species are highly similar in morphology and difficult to distinguish based on traditional vegetative structures or wood anatomical methods. The aim of the present study was to identify of these two species by DNA and chemical analysis. The complete chloroplast genome of *C. micranthum* has been determined and compared with that of *C. kanehirae*. The leaf methanol extracts of these species were also analyzed by ultra-performance liquid chromatography (UPLC). The length of the *C. micranthum* chloroplast genome was 152 675 bp, i.e. 25 bp shorter than that of *C. kanehirae*. In phylogenetic analysis, *C. micranthum* was more closely related to *C. kanehirae* than other six Lauracea species. Six validated insertion/deletions of bases in the DNA (InDels) are suitable for differentiation of the two species. Moreover, high amounts of linalool and sesamin were present in the leaves of *C. kanehirae*, but not in *C. micranthum*. Overall, this study provided new insight for distinguishing the two species on the basis of the chemical composition of the leaves and the InDel markers.

***Corresponding author: Shu-Hwa Chang,** Silviculture Division, Taiwan Forestry Research Institute, Taipei, Taiwan, Tel.: +886-2-23039978-2003, e-mail: shchangt@tfri.gov.tw
Chia-Chen Wu: School of Forest Resource and Conservation, National Taiwan University, Taipei, Taiwan; and Silviculture Division, Taiwan Forestry Research Institute, Taipei, Taiwan
Fang-Hua Chu: School of Forest Resource and Conservation, National Taiwan University, Taipei, Taiwan
Cheng-Kuen Ho: Silviculture Division, Taiwan Forestry Research Institute, Taipei, Taiwan
Chia-Hsuan Sung: Planning and Information Division, Fisheries Research Institute, Keelung, Taiwan

Keywords: chemotaxonomy, chloroplast genome, *Cinnamomum kanehirae* Hayata, *Cinnamomum micranthum* (Hayata) Hayata, identification of similar wood species, illegal logging, insertion and deletion (InDel), Lauraceae, next-generation sequencing, phylogenetic relationship, ultra-performance liquid chromatography (UPLC)

Introduction

Cinnamomum is a genus in the Lauraceae family that is composed of about 250 species mainly distributed in tropical and subtropical Asia, Australia, and the Pacific islands (Wuu-kuang 2011; Cheng et al. 2015). *Cinnamomum kanehirae* is an endemic tree species and is one of the five most precious broadleaf wood species in Taiwan. Its wood is in high demand as it is the original host of the parasitizing fungus *Antrodia cinnamomea*, which is very expensive (US$10 000 per kilogram) and known to have anti-oxidative, anti-inflammatory, and anti- carcinogenic effects (Liao et al. 2010; Lin et al. 2011; Tzeng and Geethangili 2011). The health-promotion functions of the fungus are specifically connected to *C. kanehirae* as the growing substrate (Chang and Chou 1995, 2004; Lin et al. 2011) and produce ergostane-type triterpenoids and benzolics (Shen et al. 2004; Lin et al. 2011). If *A. cinnamomea* is cultivated on other wood species, different metabolites are present. The chemical extract of *C. kanehirae* leaf also has anti-microbial activities and inhibitive effects on the hepatoma cell line (Cheng et al. 2015; Liu et al. 2015). This tree is difficult to find in the wilderness because of its unrestrained illegal logging (Lin et al. 1997; Hung et al. 2014). Thus, *C. kanehirae* is an endangered wood species, which also plays an important economic and ecological role in Taiwan (Liao et al. 2010; Cheng et al. 2015).

Cinnamomum kanehirae and *C. micranthum* have different fruits, but otherwise they are difficult to differentiate due to their similar morphologies. Both species grow in similar habitats, at 200–2000 m elevation throughout the island of Taiwan (Lin 1993; Liu et al. 2000; Huang 2003). Their wood

anatomical features are also very similar (Cai 2013). However, their essential oils contain deviating major components, i.e. they could be differentiated by chemotaxonomy (Fujita 1952, 1960). Moreover, their allozyme analysis reveals very different allele patterns (Lin 1993; Lin et al. 1997). The use of DNA markers for the identification of similar wood species was intensively discussed in the last decade, and obviously, this approach is also suited for tracing the origin of trees from illegal logging or for the identification of endangered species (Degen and Fladung 2007; Michael et al. 2012; Jiao et al. 2014, 2015; Sandak et al. 2015; Yu et al. 2016).

The DNA markers and chemical profiles of *C. micranthum* and *C. kanehirae* have not been subjected to a comparative analysis. The goal of this study was to fill this gap and to provide useful molecular and chemical methods for identifying the two species in focus. The intention was to determine the complete chloroplast (cp) genome sequences of these species and to extract their leaf material with methanol and to compare the corresponding data under consideration with the data in "GenBank accession No. KR014245" (Wu et al. 2016). A phylogenetic tree should be constructed under consideration of the species in Magnoliidae including Laurales, Magnoliales, and Canellales (outgroup) in order to validate the relationship to *C. micranthum* cp genome sequence. Finally, the insertion/deletion of bases in the cp DNA (InDel markers) should be developed to identify the wood and leaf samples of these two species.

Materials and methods

Plant material and DNA extraction: During winter (October–January) in Taiwan, 23 fresh leaf samples of *C. micranthum* were collected for chemical analyses from Taipei, Ilan, and Natou, while 23 *C. kanehirae* samples were collected from 13 sites throughout the Taiwan island, including sampling areas covered by previous studies (Lin 1993; Liao et al. 2010). Wood materials (with height over 3 m) were collected from 12 individuals from Fushan Botanical Garden in Ilan and Liouguei Research Center in Kaohsiung city of Taiwan. The diameters of wood discs collected were over 60 mm. The *C. micranthum* wood samples were fresh or air-dried for 2 years; the *C. kanehirae* wood samples were air-dried for 2 or 4 years.

The DNA extraction of leaf samples followed the CTAB method (Doyle and Doyle 1990). The extraction of sapwood DNA followed the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) protocol according to Jiao et al. (2014). The 2-µg DNA sample was subjected to quality checks before next-generation sequencing (NGS) using a Qubit fluorometer with a Quant-iT dsDNA BR assay kit (Thermo Fisher Scientific, Waltham, MA, USA) and agarose gel electrophoresis assay.

Next-generation sequencing and chloroplast genome assembly: Genomic DNAs were sequenced with the Illumina Miseq system platform (Illumina, San Diego, CA, USA) and a MiSeq Reagent Kit V3 (600 cycles) at National Yang-Ming University, Taiwan. The paired-end libraries were prepared with Illumina TruSeq DNA polymerase chain reaction (PCR)-free sample preparation (550 bp insert size). The adapter sequences and low-quality score of the raw reads were trimmed by the CLC quality trimming tool with default parameters, and all the trimmed reads were used for *de novo* assembly with the CLC genome assembler (Genomics Workbench 8.0.1, CLC, Rarhus, Denmark). The assembled contigs were then submitted to BLASTn (cut-off e-value = 1e-5) with the reference cp genome (*C. kanehirae*, KR01425), (Wu et al. 2016). The contigs, which have high similarity to the *C. kanehirae* cp DNA, were manually assembled into a circular genomic sequence. Validations of the overlapping regions between contigs were done by PCR and all of the primers and PCR conditions are described in Supplementary Table S1. BWA v0.6.2 (Li and Durbin 2009) and IGV v2.3.25 (Robinson et al. 2011) were also consulted for mapping all raw reads onto the final assembly and visually inspected the genome coverage levels.

Genome annotation: Protein coding, ribosomal (r)RNA, and transfer (t)RNA genes of the cp genome were predicted with DOGMA (Wyman et al. 2004) and CpGAVAS (Liu et al. 2012). The start and stop codons of all predicted genes were manually corrected. The *C. kanehirae* cp genome (KR014245) served as a reference to adjust the gene annotation. Eventually, a complete *C. micranthum* cp genome was obtained and deposited in the NCBI database (KT833081). The cp genome circle map was drawn with OGDRAW (Lohse et al. 2007).

Comparative analysis within the chloroplast genomic sequence: A WebSat online analysis (Martins et al. 2009) was used to find simple sequence repeat (SSR) regions, with the repeat units being 12 for mono-nucleotides, six for di-nucleotides, five for tri-ncleotides, and three for tetra-, penta-, and hexa-nucleotides. A repetitive structure analysis was conducted with REPuter (Kurtz et al. 2001), (minimum size of repeat event was 20 bp). InDel events were calculated with the DNaSP 5.0 software (Librado and Rozas 2009). Long InDel regions were found with MEGA 6.0 by alignment (Tamura et al. 2013).

Phylogenetic tree construction: Sixteen complete cp genome sequences related to *C. micranthum* served for constructing the phylogenetic tree, including all available complete cp genomes in Lauraceas. The complete cp sequences were aligned with MAFFT (Katoh and Standley 2013) with default setting. A maximumlikelihood phylogeny was constructed using MEGA 6.0 (Tamura et al. 2013) with the GTR+I+G model. Bootstrap-values were calculated with 1000 replicates. Eight Magnoliales, Magnoliaceae species: *Magnolia tripetala* (KJ408574), *Magnolia liliiflora* (JN867576), *Magnolia denudate* (NC018357), *Magnolia yunnanensis* (KF753638), *Magnolia kwangsiensis* (NC015892), *Magnolia grandiflora* (NC020318), *Magnolia officinalis* (NC020316), and *Liriodendron tulipifera* (DQ899947), and eight Laurales, Lauraceae species: *C. kanehirae* (KR014245), *C. micranthum* (KT833081), *Machilus yunnanensis* (KT348516), *Machilus balansae* (KT348517), *Phoebe omeiensis* (KX437772), *Phoebe sheareri* (KX43773), *Persea americana* (KX43771), and *Calycanthus fertilis* (AJ42841) formed the basis for the phylogenetic tree construction. The *Drimys granadensis* (DQ887676) (Canellales, Winteraceae) served as the outgroup.

PCR amplification and sequencing for validating the InDel molecular markers: For InDel validation, primer sets were designed with the Primer 3.0 software (Untergasser et al. 2012). The PCR contained

10 ng of genomic DNA and was performed with a Taq polymerase kit (GenetBio, Daejeon, Korea) following the manual in a total volume of 50 µl. The thermocycler Veriti (Applied Biosystems, Foster city, CA, USA) was used for all PCRs with the following cycling parameters: 94°C for 5 min; 32 cycles of 94°C for 30 s, the annealing temperature for 30 s (see Supplementary Table S1), and 72°C for 1 min; and a final extension at 72°C for 10 min. The details of the PCR primer sets are shown in Table 2 and in Supplementary Table S1. Forty-six leaf and 12 wood samples were used for InDel primers' validation. PCR products were analyzed by 2–3% agarose gel electrophoresis (Agarose SFR™, AMRESCO, Solon, OH, USA) and detected by ethidium bromide under UV light for visualization.

Solvent extraction and the ultra-performance liquid chromatography analysis: Fresh leaves of *C. micranthum* and *C. kanehirae* trees were sampled in the same winter season. They were cut into pieces, and 1 g of fresh weight was sampled with three replications. Each sample was put into a 50-ml centrifugal tube, 10-ml MeOH was added, the mixture was ground to a fine powder in a homogenizer, and then rinsed with 10-ml MeOH. These tubes were cooled in an ice bath and ultrasonicated for 60 min. The supernatant was collected in a centrifuge at 2500 rpm. The residue was re-extracted with 15-ml MeOH, and a 2nd supernatant was obtained with the same procedure. Clear supernatants were combined and passed through a 0.2-µm filter (Xiboshi, Hydrophilic PTFE, Tianjin, China). The UPLC analyses were carried out with Acquity UPLC Waters (Waters, Milford, MA, USA) equipped with a C-18 column (150 x 2.1 mm, 1.8-µm particle size; Acquity UPLC HSS T3, Waters). The mobile phase for elution was distilled-deionized water (DDW): acetonitrile (70 : 30) for 23 min, DDW graded down to 30% for 25 min, increased up to 70% for 5 min, and maintained for 5 min at a flow rate of 0.3 ml min^{-1}, followed by UV detection at 200 nm. A standard curve was obtained based on authentic terpinen-4-ol (Acros Organic, Geel, Belgium), linalool (Acros Organic), sesamin (Sigma-Aldrich, St. Louis, MO, USA), and safrole standards (Acros Organic). Concentrations of extracts were expressed as % b.o. dry wt. For dry-weight measurements, residues were dried at 60°C for 48 h after extraction.

Table 1: Summary of the *C. micranthum* and *C. kanehirae* cp genome analysis.

Parameter	*C. micranthum*	*C. kanehirae*
NGS raw reads (no.)	12 491 350	16 233 870
Raw sequencing bases (no.)	3 634 256 229	4 886 394 870
Reads after trimmed (no.)	12 196 250	16 233 870
Clean bases after trimmed (no.)	1 806 905 010	4 833 754 899
Reads for cp *de novo* assembly (no.)	484 117	589 056
Mean of coverage (no.)	425	1314
Total cp genome size (bp)	152 675	152 700
Large single copy (LSC) (bp)	93 662	93 643
Inverted repeat (IR) region (bp)	20 069	20 106
Small single copy (SSC) (bp)	18 875	18 845
GC content (%)	39.13	39.13
LSC (%)	37.94	37.93
IR (%)	44.41	44.39
SSC (%)	33.83	33.83
Total no. of different genes	112	112
Different protein-coding genes	78	78
Different tRNA genes	30	30
Different rRNA genes	4	4
No. of different duplicated genes	15	15
No. of different genes with introns	15	15

Results and discussion

De novo assembly and gene organization

Raw data obtained from NGS technology of a total 3 634 256 229 bp (12 491 350 reads) were from *C. micranthum* sequencing. After trimming, 12 196 250 clean reads with 1 806 905 010 bp were obtained and used for *de novo* assembly (Table 1). Three long contigs which covered the entire cp genome sequence of *C. kanehirae* (KR014245) were used for BLASTn and showed high similarity (99.61–99.9%) to the *C. kanehirae* large single copy (LSC), small single copy (SSC), and inverted repeat (IR) region. It can be inferred that these three long contigs are the LSC, IR, and SSC, respectively. The two IR sequences were reversed and complementary; the assembly software could not combine these three contigs as one contig. This result is similar to the assembly of the Tartary buckwheat (*Fagopyrum tataricum*) cp genome based on NGS (Cho et al. 2015). The three long contigs revealed about 10–15-bp overlap between each of their flanking regions and then were manually assembled as a single circular cp genome. The Sanger sequencing of overlapping regions (LSC-IRa, IRa-SSC, SSC-IRb, and IRb-LSC) was 100% identical with the NGS-based assembled cp genome, and these sequences formed the basis for drawing a circle map (Figure S1). The final assembly resulted in an average of 425-fold coverage of pair-end reads with a mapping quality of 60. The complete *C. micranthum* cp genome was 152 675 bp in length (GenBank no. KT833081), which was 25 bp shorter than the *C. kanehirae* cp genome. Both cp genomes consisted of one LSC, one SSC, and two IRs. The respective lengths of the LSCs in *C. kanehirae* and *C. micranthum* were 93 642 and 93 662 bp. The SSC of *C. kanehirae* was 18 845 bp and that of *C. micranthum* was 18 875 bp. The total lengths of the two IRs in *C. kanehirae* and *C. micranthum* were 40 212 and 40 138 bp, respectively. Two genes, *ndhF* and *trnH*, are located at the IR/LSC and SSC/IR junction areas and have different lengths of expansion and contraction. IR contraction and expansion at the junction of the SSC or LSC lead to cp size variations in different plants and contribute to the formation of pseudogene (Lin et al. 2012). The *ycf1* and *ycf2* span the junctions of the IRa/SSC and LSC/IRa, respectively, showed incomplete duplication and caused the contraction of the IR regions. This result

is consistent with other Lauraceae cp genomes (Song et al. 2015, 2016). The overall GC content and gene contents of the *C. micranthum* cp genome are listed in Table 1. The overall GC content also falls in the ranges of vascular plants (38 ~ 39%) (Palmer and Stein 1986). The value 39.1% (Table 1) is almost the same as that in other plants in the Lauraceae (Song et al. 2016). The gene contents between *C. micranthum* and *C. kanehirae* were the same (Table 1). Eight of 112 genes (*trnV-UAC*, *psaA*, *rps2*, *atpF*, *trnK-UUU*, *ycf1*, *ycf2*, and *ndhB*) had different sequence orientations from those of *C. kanehirae*. Among these annotated genes, 97 genes were a single copy and 15 genes were duplicated. Among the 78 protein coding genes in the *C. micranthum* cp genome, 63 genes were located in the LSC region, 11 genes in the SSC region, three genes in the IRa, and five genes in the IRb. Ten protein-coding genes and five tRNA genes contained introns in *C. micranthum*, and among these, 10 protein-coding genes and eight genes (*atpF*, *petB*, *ndhA*, *ndhB*, *rpoC1*, *rps12*, *rps16*, and *rpl2*) had one intron, and two genes (*ycf3* and *clpP*) had two introns.

Distribution of SSRs and repetitive structures

The cp SSR regions may be variable and typically non-recombination, uniparentally inherited and effectively haploid, and have been used for population genetic studies in many plant species (Navascues and Emerson 2005; Desiderio et al. 2012). Within the two cp genomes of *C. micranthum* and *C. kanehirae*, the SSRs were analyzed. In total, 43 SSR loci accounted for 595 bp of the total sequence in *C. micranthum* (0.39%). In the *C. kanehirae* cp genome, 39 SSR loci were found which accounted for 520 bp of the total sequence (0.34%). In total, the 43 SSR loci of *C. micranthum* cp genome can be subdivided into 24 mono-nucleotides, five di-nucleotides, three tri-nucleotides, eight tetra-nucleotides, two penta-nucleotides, and one hexa-nucleotide. Among the Lauraceae species, the number of SSR showed little variation and almost all were located at inter-genic spacer (IGS) regions as has been seen in other plants (Do et al. 2013; Ma et al. 2013; Zhang et al. 2014; Song et al. 2015, 2016). In the *C. micranthum* cp genome, 43 SSR primer sets were designed (Table S1), and these SSR loci can be used for further population genetic research in the *Cinnamomum* genus.

In addition, 41 repetitive structures were found in *C. micranthum* containing three complement matches, eight reverse matches, 13 forward matches, and 17 palindromic matches. Compared to the repetitive structures in *C. kanehirae*, 30 of 41 repeat events (80.5%) were located close to the same location. Other repeat events (eight in *C. micranthum* and seven in *C. kanehirae*) revealed different locations between the two genomes. Most of the repeat events occurred in IGS in both species. This is in accordance with findings in other plants (Do et al. 2013; Ma et al. 2013; Su et al. 2014; Zhang et al. 2014; Cho et al. 2015). In comparison with two other Lauraceae species, *Machilus balansae* and *M. yunnanensis*, the number of repeats (with lengths of > 20 bp) varied from 39 to 41 (Song et al. 2015). About 61% of *C. micranthum* sequence repeats belonged to reverse or palindromic repeats. The palindromic repeats may form secondary structures (hairpin structure) and contribute duplicated sequence, and then induce InDel sequences (Johnson et al. 2016).

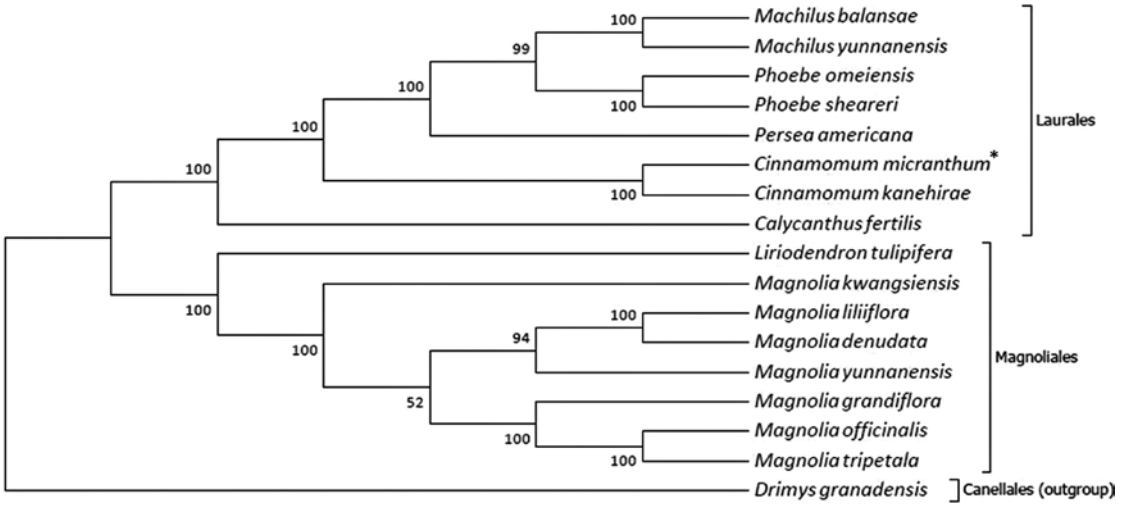

Figure 1: Maximum-likelihood phylogeny of complete cp genomes in the Magnoliales, Laurales and Canellales. Bootstrap supports are shown above the branches. The *C. micranthum* genome is denoted by an asterisk.

Thus, four of the six validated InDels in this study were located around the palindromic sequence region. The IRs were the largest palindromic structure in the cp genome, and larger IRs can stabilize the structure of the cp genome and reduce the inversion probability of the sequence (Ma et al. 2013).

Phylogenetic tree analysis of *C. micranthum* and other Lauraceae species

In the present study, all available complete cp genomes were selected in Lauraceae and relative Magnoliaceae species to determine the phylogenetic relationship of *C. micranthum*. The phylogenetic tree (Figure 1) revealed two major clades, Magnoliales and Laurales. Within the Laurales, the Calycanthaceae formed the first branch, followed by the Lauraceae. In the clade of the Lauraceae, genus *Cinnamomum* was as a sister group to the other Lauraceae species. *Persea americana* is closest to *Machilus* and *Phoebe*. Phylogenetic tree is in accordance with angiosperm phylogenic group classification (Chanderbali et al. 2001; Bremer et al. 2009; Renner, 2011; Song et al. 2016). This phylogenetic tree is also consistent with APG III and provides the basic framework for studying Lauraceae phylogenetics (Bremer et al. 2009) and also supports the relationship of *C. micranthum* cp genome being close to *C. kanehirae* in *Cinnamomum*.

InDel predictions and identification of *C. micranthum* and *C. kanehirae*

In total, 72 InDel events were detected between the *C. micranthum* and *C. kanehirae* cp genome sequences. Six InDel length of > 10 bp had been validated for distinguishing between *C. micranthum* and *C. kanehirae* (Table 2). Three InDels were located at the SSC, one in the LSC and two at IRs. Most of the InDels were found within the IGS area. InDels 2, 3, 4, and 6 were also found at repetitive structure regions. InDels 2 and 4 were palindromic matching structures. InDels 3 and 6 were insertion sequences of *C. kanehirae* with a forward matching structure. InDels 3 and 6 were palindromic sequences with each other at the *ycf1* gene, because they were located at IRs. In this study, we successfully extracted DNA from the sapwood and utilized as InDel biomarkers to authenticate wood and leaf of *C. micranthum* and *C. kanehirae*. Six InDels show identical amplicon sizes and clear polymorphism in all wood and leaf samples between two species. Six InDel markers can be successfully applied to identify wood and leaf samples.

Table 2: List of InDel markers (over 10 bp) and locations in *C. micranthum* and *C. kanehirae* cp genome.

InDel no.	Forward primer (5′–3′)	Reverse primer (5′–3′)	Expected size (bp)		InDel size	Position	Location of *C. micranthum*	Type[a]
			C. micranthum	*C. kanehirae*				
1	ATTAAATCAGACGCGAATTTG	TGTTTCAGTTTTGGCGAAAAAG	316	287	29	LSC	IGS (*rps16-trnQ-UUG*)	Deletion
2	AGGTACAGCTTGAGCAATGAA	TGTTCCACGATATGGTCCGT	240	222	18	SSC	*ycf1*	Deletion
3	TGCCTCATTTCATAGGGAAC	CCATAATAGAGTCCGACCAC	334	353	19	IRa	IGS (*trnN-GUU-ycf1*)	Insertion
4	ACTGTGTGATTTTTAAAAGCGT	GGAACATCGTCCCGTATATT	155	165	10	SSC	IGS (*ndhG-ccsA*)	Insertion
5	TTGGTGTTGCAAGTTTCTTTG	GAATTCATAAACCGAATAG	282	300	18	SSC	IGS (*ndhF-ycf1*)	Insertion
6	ATCATCACAAACCTCCCTTT	TCCATGCCTCATTTCATAGG	254	273	19	IRb	IGS (*ycf1-trnN-GUU*)	Insertion

[a]Indicates that *C. kanehirae* was used as reference.

The results of agarose gel electrophoresis showed six wood collections of *C. micranthum* and *C. kanehirae* (Figure 2). *Cinnamomum kanehirae* wood samples that were air-dried for 4 years were used in the validation of our data: it revealed that the amount of wood DNA extraction was enough for PCR amplification. This might be because cp DNA was a high copy number in the cells based on sapwood. Many studies reported that the less potential inhibitors of DNA isolation and PCR amplification are present in the sapwood (Rachmayanti et al. 2006; Tnah et al. 2012; Yu et al. 2016). There are many more living parenchyma cells in the sapwood than in the heartwood (Jiao et al. 2014; Yu et al. 2016). Thus, the six InDel primers are useful for wood differentiation between *C. micranthum* and *C. kanehirae* and can be helpful in forensic investigations in the court. Compared to morphological characteristics, molecular markers are more universal, which are not affected by environmental and developmental factors (Zhao et al. 2015). The features of *C. micranthur* and *C. kanehirae* are so similar that the adulteration of these two species happened in the seedling and wood trade. These six useful validated cp InDel markers can be used to identify *C. micranthum* and *C. kanehirae* and detect the adulteration problem.

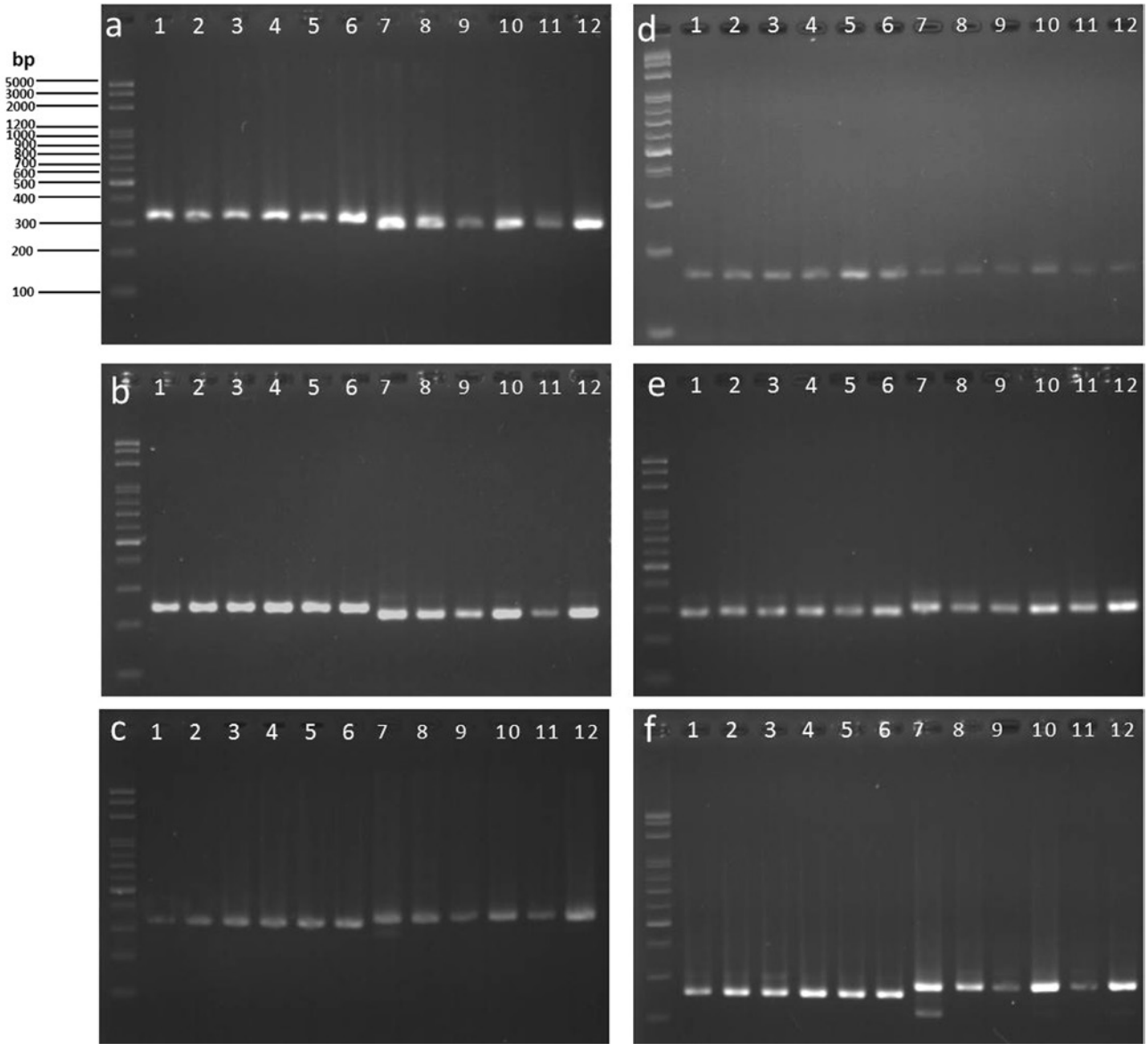

Figure 2: Confirmation of insertion/deletion (InDel) markers between *C. micranthum* and *C. kanehirae* wood by PCR amplification. a–f: PCR amplifications with InDel nos. 1–6. The 1–6 indicate *C. micranthum* wood collections; the 7–12 indicate *C. kanehirae* wood collections. DNA maker ladder: 100-bp DNA ladder (Cat# GTL101, GeneTeks BioSicence, Inc., Taipei, Taiwan.). The detail DNA band sizes are labeled next to the figure.

Table 3: The content (ppm) of *C. micranthum* and *C. kanehirae* leaf methanol extracts by UPLC analysis.

Species tree no.	Locations[a]	Sources of leaves[b]	Terpine-4-ol (ppm)	Linalool (ppm)	Sesamin (ppm)	Safrole (ppm)
C. micranthum						
No. 1–23		Wild tree	Non-detected	Non-detected	Non-detected	Non-detected
C. kanehirae						
No. 1–9	SE	Plantation	Non-detected	4061–26877	1771–9366	Non-detected
No. 10–11	SE	Wild tree	0–151	6121–6946	3120–5243	Non-detected
No. 12–17	SW	Plantation	Non-detected	415–41026	5138–12462	Non-detected
No. 18–19	SW	Wild tree	Non-detected	2970–3139	5047–10088	Non-detected
No. 20–21	NW	Wild tree	Non-detected	2663–3064	12881–15082	0–7
No. 22	CW	Wild tree	Non-detected	1286–1435	5981–6132	Non-detected
No. 23	Other	Plantation	Non-detected	3257–3368	3928–4405	113–164

[a]SE, Southern east; SW, southern west; NW, northern west; CW, central west of Taiwan.
[b]Planation had 10-year-old clonal trees planted at the same area in north eastern Taiwan; wild tree was collected from natural stands.

Differences in chemical constituents of *C. micranthum* and *C. kanehirae* leaves

Chemotaxonomical studies are also very helpful to investigate the relationships at intergenus, interspecies, and intraspecies levels (Angioni et al. 2004; de Sena Filho et al. 2012; Rasheed et al. 2012). The results of the chemical profiles further suggested that *C. micranthum* is different from *C. kanehirae*. The *C. micranthum* leaves do not contain terpinen-4-ol, linalool, safrole, and sesamin, but sesamin and linalool are abundant in the *C. kanehirae* leaves. In the investigated *C. kanehirae* samples, the sesamin content varied from 1771 to 15 082 ppm with a mean of 12 870 ppm; the linalool content varied from 1286 to 41 026 ppm with a mean of 6696 ppm. The safrole and terpinen-4-ol contents were scant, and some contents were detected in specimens of *C. kanehirae* (Table 3). Previous studies mentioned that the major component of *C. micranthum* leaf essential oil is decylaldehyde and a trace component is linalool, while the major components in the *C. kanehirae* essential oil include linalool and terpinen-4-ol (Fujita, 1952, 1960, 1967). *C. kanehirae* methanol extraction also contained sesamin in the previous report (Hsieh et al. 2005). Our study provides the comparative chemical profiles of methanol extracts between these two species. Terpinen-4-ol was not found in the extract, perhaps because methanol was used for extraction. Compared with a previous study (Cheng et al. 2015), the plant samples investigated here were collected from throughout the islands of Taiwan. Thus, it is reasonable that variable linalool concentrations occurred in the UPLC analysis and it might be further classified into five different chemotypes (Cheng et al. 2015). The chemical profiles showed obvious discrimination between *C. micranthum* and *C. kanehirae*. Thus, it is confirmed that the natural chemical characters are different between these two species. Our study provides the first report comparing the chemical constituents of *C. micranthum* and *C. kanehirae* leaves based on methanol extraction. These chemotaxonomical data are useful and can be an important supplement to those of genetic identification.

Conclusions

The first complete cp genome analysis of *C. micranthum* was reported and it provided a cp genome comparison between *C. micranthum* and *C. kanehirae*. Six InDel markers were identified for distinguishing the endangered *C. kanehirae* from *C. micranthum*. The chemical profiles of leaf extracts provided additional help in differentiating between the two *Cinnammomum* species in focus.

Supplementary materials

Supplementary Table S1: List of primers; Supplementary Figure S1: Circle map of the *Cinnamomum micranthum* cp genome.

Acknowledgments: We deeply appreciate Dr. Huei-Jiun Su for manuscript discussion and phylogenetic tree reconstruction. We thank the Ministry of Science and Technology, Taiwan, who supported this work (grant no. MOST 102-2313-B-054-003-MY3). We also thank the National Yang-Ming University VYM Genome Research Center, who provided the assistance for sequencing service.

References

Angioni, A., Barra, A., Cereti, E., Barile, D., Coisson, J.D., Arlorio, M., Dessi, S., Coroneo, V., Cabras, P. (2004) Chemical composition, plant genetic differences, antimicrobial and antifungal activity investigation of the essential oil of *Rosmarinus officinalis* L. J. Agric. Food Chem. 52:3530–3535.

Bremer, B., Bremer, K., Chase, M.W., Fay, M.F., Reveal, J.L., Bailey, L.H., Soltis, D.E., Soltis, P.S., Stevens, P.F. (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Bot. J. Linn. Soc. 161:105–121.

Cai, J.R. (2013) DNA extraction and sequencing of *Cinnamomum kanehirae* wood. National Chiayi University Master Thesis, Chiayi, Taiwan, 2013.

Chanderbali, A.S., van der Werff, H., Renner, S.S. (2001) Phylogeny and historical biogeography of Lauraceae: evidence from the chloroplast and nuclear genomes. Ann. Missouri Bot. Gard. 88:104–134.

Chang, T.T., Chou, W.N. (1995) *Antrodia cinnamomea* sp. nov. on *Cinnamomum kanehirai* in Taiwan. Mycol. Res. 99:756–758.

Chang, T.T, Chou, W.N. (2004) *Antrodia cinnamomea* reconsidered and *A. salmonea* sp. nov. on *Cunninghamia konishii* in Taiwan. Bot. Bull. Acad. Sin. 45:347–352.

Cheng, S.S., Lin, C.Y., Yang, C.K., Chen, Y.J., Chung, M.J., Chang, S.T. (2015) Chemical polymorphism and composition of leaf essential oils of *Cinnamomum kanehirae* using gas chromatography/mass spectrometry, Cluster Analysis, and Principal Component Analysis. J. Wood Chem. Technol. 35:207–219.

Cho, K.S., Yun, B.K., Yoon, Y.H., Hong, S.Y., Mekapogu, M., Kim, K.H., Yang, T.J. (2015) Complete chloroplast genome sequence of tartary buckwheat (*Fagopyrum tataricum*) and comparative analysis with common buckwheat (*F. esculentum*). PLoS One 10:e0125332.

Degen, B., Fladung, M. (2007) Use of DNA-markers for tracing illegal logging. In: Proceedings of the International Workshop "Fingerprinting Methods for the Identification of Timber Origins", Bonn, Germany. pp. 6–14.

de Sena Filho, J.G., Rabbani, A.R.C., dos Santos Silva, T.R., da Silva, A.V.C., Souza, I.A., Santos, M.J.B.A., de Jesus, J.R., Nogueira, P.C. de L., Duringer, J.M. (2012) Chemical and molecular characterization of fifteen species from the *Lantana* (Verbenaceae) genus. Biochem. Syst. Ecol. 45:130–137.

Desiderio, F., Bitocchi, E., Bellucci, E., Rau, D., Rodriguez, M., Attene, G., Papa, R., Nanni, L. (2012) Chloroplast microsatellite diversity in *Phaseolus vulgaris*. Front. Plant Sci. 3:1–15.

Do, H.D.K., Kim, J.S., Kim, J.H. (2013) Comparative genomics of four Liliales families inferred from the complete chloroplast genome sequence of *Veratrum patulum* O. Loes. (Melanthiaceae). Gene 530:229–235.

Doyle, J., Doyle, J. (1990) Isolation of plant DNA from fresh tissue. Focus (Madison). 12:13–15.

Fujita, Y. (1952) *Cinnamomur camphora* Sieb. and its allied species: their inter-relations considered from the view-points of species characterics, chemical constituents, geographical distributions and Evolution. Bot. Mag. Tokoyo 65:245–250.

Fujita, Y. (1960) Classification of the plants viewed from the constituents of essential oils (I) *Cinnamomum micranthum* HAY. and *C. kanahirai* HAY. Acta Phytotax Geobot 8:178–179.

Fujita, Y. (1967) Classification and phylogeny of genus *Cinnamomum* viewed from the constituents of essential oil. Bot. Mag. Tokoyo 80:261–271.

Hsieh, T.J., Lu, L.H., Su, C.C. (2005) NMR spectroscopic, mass spectroscopic, X-ray crystallographic, and theoretical studies of molecular mechanics of natural products: farformolide B and sesamin. Biophys. Chem. 114:13–20.

Huang, T.C. (2003) Flora of Taiwan, Second. ed. Editorial Committee of the Flora of Taiwan, Taipei, Taiwan, 2003.

Hung, K.H., Lin, C.H., Chuan, S.H., Chung, C.Y., Ping, J.L. (2014) Development, characterization and cross-species amplification of new microsatellite primers from an endemic species *Cinnamomum kanehirae* (Lauraceae) in Taiwan. Conserv. Genet. Resour. 6:911–913.

Jiao, L., Yin, Y., Cheng, Y., Jiang, X. (2014) DNA barcoding for identification of the endangered species *Aquilaria sinensis*: comparison of data from heated or aged wood samples. Holzforschung 68:487–494.

Jiao, L., Liu, X., Jiang, X., Yin, Y. (2015) Extraction and amplification of DNA from aged and archaeological *Populus euphratica* wood for species identification. Holzforschung 69:925–931.

Johnson, M.T.J., Wright, S.I., Pellizzer, T., Sobanski, J., Bock, R. (2016) Spontaneous chloroplast mutants mostly occur by replication slippage and show a biased pattern in the plastome of Oenothera. Plant Cell 28:911–929.

Katoh, K., Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–780.

Kurtz, S., Choudhuri, J. V, Ohlebusch, E., Schleiermacher, C., Stoye, J., Giegerich, R. (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 29:4633–4642.

Li, H., Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

Liao, P.C., Kuo, D.C., Lin, C.C., Ho, K.C., Lin, T.P., Hwang, S.Y. (2010) Historical spatial range expansion and a very recent bottleneck of *Cinnamomum kanehirae* Hay. (Lauraceae) in Taiwan inferred from nuclear genes. BMC Evol. Biol. 10:124.

Librado, P., Rozas, J. (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452.

Lin, T.P. (1993) *Cinnamomum kanehirae* Hay. and *Cinnamomum micranthum* (Hay.) Hay. Bull. Taiwan. Res. Inst. 8:11–20.

Lin, T.P., Cheng, Y.P., Huang, S.G. (1997) Allozyme Variation in Four Geographic Areas of *Cinnamomum kanehlrae*. J. Hered. 88:433–438.

Lin, T.Y., Chen, C.Y., Chien, S.C., Hsiao, W.W., Chu, F.H., Li, W.H., Lin, C.C., Shaw, J.F., Wang, S.Y. (2011) Metabolite profiles for *Antrodia cinnamomea* fruiting bodies harvested at different culture ages and from different wood substrates. J. Agric. Food Chem. 59:7626–7635.

Lin, C.P., Wu, C.S., Huang, Y.Y., Chaw, S.M. (2012) The complete chloroplast genome of *Ginkgo biloba* reveals the mechanism of inverted repeat contraction. Genome Biol. Evol. 4:374–381.

Liu, H.Y., Yang, Y.P., Lu, S.Y., Shih, B.L. (2000) Manual of Taiwan Vascular Plants, Taipei, Taiwan, 2000.

Liu, C., Shi, L., Zhu, Y., Chen, H., Zhang, J., Lin, X., Guan, X. (2012) CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely

sequenced chloroplast genome sequences. BMC Genomics 13:715.

Liu, Y., Chen, K.H., Leu, Y., Way, T., Wang, L., Chen, Y., Liu, Y.M. (2015) Ethanol extracts of *Cinnamomum kanehirai* Hayata leaves induce apoptosis in human hepatoma cell through caspase-3 cascade. Onco. Targets. Ther. 8:99–109.

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

Ma, J., Yang, B., Zhu, W., Sun, L., Tian, J., Wang, X. (2013) The complete chloroplast genome sequence of *Mahonia bealei* (Berberidaceae) reveals a significant expansion of the inverted repeat and phylogenetic relationship with other angiosperms. Gene 528:120–131.

Martins, W.S., Lucas, D.S.C, de Souza Neves, K.F, Bertioli D.J. (2009) Bioinformation WebSat – a web software for microsatellite marker development. Bioinformation 3:282–283.

Michael H.A., Schröder, H., Wischnewski, N., Degen, B., Magel, E., Fladung, M. (2012) Development of DNA-based methods to identify CITES-protected timber species: a case study in the Meliaceae family. Holzforschung 66:97–104.

Navascues, M., Emerson, B.C. (2005) Chloroplast microsatellites: measures of genetic diversity and the effect of homoplasy. Mol. Ecol. 14:1333–1341.

Palmer, J.D., Stein, D.B. (1986) Conservation of chloroplast genome structure among vascular plants. Curr. Genet. 10:823–833.

Rachmayanti, Y., Leinemann, L., Gailing, O., Finkeldey, R. (2006) Extraction, amplification and characterization of wood DNA from dipterocarpaceae. Plant Mol. Biol. Rep. 24:45–55.

Rasheed, N.M.A., Nagaiah, K., Goud, P.R., Sharma, V.U.M. (2012) Chemical marker compounds and their essential role in quality control of herbal medicines. Ann. Phytomedicine 1:1–8.

Renner, S.S. (2011) Laurales. In: eLS. John Wiley & Sons, Ltd: Chichester.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P. (2011) Integrative Genomics Viewer. Nat. Biotechnol. 29:24–26.

Sandak, J., Sandak, A., Cantini, C., Autino, A. (2015) Differences in wood properties of *Picea abies* L. Karst. in relation to site of provenance and population genetics. Holzforschung 69:385–397.

Shen, Y.C., Chou, C.J., Wang, Y.H., Chen, C.F., Chou, Y.C., Lu, M.K. (2004) Anti-inflammatory activity of the extracts from mycelia of *Antrodia camphorata* cultured with water-soluble fractions from five different *Cinnamomum* species. FEMS Microbiol. Lett. 231:137–143.

Song, Y., Dong, W., Liu, B., Xu, C., Yao, X., Gao, J., Corlett, R.T. (2015) Comparative analysis of complete chloroplast genome sequences of two tropical trees *Machilus yunnanensis* and *Machilus balansae* in the family Lauraceae. Front. Plant Sci. 6:1–8.

Song, Y., Yao, X., Tan, Y., Gan, Y., Corlett, R.T. (2016) Complete chloroplast genome sequence of the avocado: gene organization, comparative analysis, and phylogenetic relationships with other Lauraceae. Can. J. For. Res. 46:1293–1301.

Su, H.J., Hogenhout, S.A., Al-Sadi, A.M., Kuo, C.H. (2014) Complete chloroplast genome sequence of omani lime (*Citrus aurantiifolia*) and comparative analysis within the Rosids. PLoS One 9:e113049.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. (2013) MEGA6 : molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30:2725–2729.

Tnah, L.H., Lee, S.L., Ng, K.K.S., Bhassu, S., Othman, R.Y. (2012) DNA extraction from dry wood of *Neobalanocarpus heimii* (Dipterocarpaceae) for forensic DNA profiling and timber tracking. Wood. Sci. Technol. 46:813–825.

Tzeng, Y.M., Geethangili, M. (2011) Review of pharmacological effects of *Antrodia camphorata* and its bioactive compounds. Evidence-based Complement. Altern. Med. 2011:1–17.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S.G. (2012) Primer3 – new capabilities and interfaces. Nucleic Acids Res. 40:1–12.

Wu, C.C., Ho, C.K., Chang, S.H. (2016) The complete chloroplast genome of *Cinnamomum kanehirae* Hayata (Lauraceae). Mitochondrial DNA part A 27:2681–2682.

Wuu-kuang, S. (2011) Taxonomic revision of *Cinnamomum* (Lauraceae) in Borneo. Blumea 56:241–264.

Wyman, S.K., Jansen, R.K., Boore, J.L. (2004) Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20:3252–3255.

Yu, M., Liu, K., Zhou, L., Zhao, L., Liu, S. (2016) Testing three proposed DNA barcodes for the wood identification of *Dalbergia odorifera*. Holzforschung 70:127–136.

Zhang, Y., Ma, J., Yang, B., Li, R., Zhu, W., Sun, L., Tian, J., Zhang, L. (2014) The complete chloroplast genome sequence of *Taxus chinensis var. mairei* (Taxaceae): loss of an inverted repeat region and comparative analysis with related species. Gene 540:201–209.

Zhao, Y., Yin, J., Guo, H., Zhang, Y., Xiao, W., Sun, C., Wu, J., Qu, X., Yu, J., Wang, X., Xiao, J. (2015) The complete chloroplast genome provides insight into the evolution and polymorphism of *Panax ginseng*. Front. Plant Sci. 5:1–12.

Supplemental Material: The online version of this article (DOI: 10.1515/hf-2016-0133) offers supplementary material, available to authorized users.