JSE Journal of Systematics and Evolution

doi: 10.1111/jse.12189

Research Article

Phylogenetic utility of *LEAFY* gene in *Cinnamomum* (Lauraceae): Gene duplication and polymerase chain reaction-mediated recombination

Jian-Feng Huang[1,3], Lang Li[1], John G. Conran[2], and Jie Li[1]*

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

[2]Australian Center for Evolutionary Biology and Biodiversity and Sprigg Geobiology Center, School of Biological Sciences, University of Adelaide, Adelaide SA 5005, Australia

[3]University of Chinese Academy of Sciences, Beijing 100049, China

*Author forcorrespondence. E-mail: jieli@xtbg.ac.cn. Tel.: 86-871-65144431. Fax: 86-871-65160916.

Received 21 October 2015; Accepted 3 December 2015; Article first published online 8 February 2016

Abstract Low copy nuclear genes show great potential to provide phylogenetic information, but their use has been hampered by several inherent adverse factors. Polymerase chain reaction (PCR)-mediated recombination ranks among these factors and occurs when high levels of similar paralogs for a low-copy nuclear gene coexist within a single PCR amplification reaction. In this study, the *LEAFY* gene was cloned and sequenced for 63 *Cinnamomum* species and two copies of the second intron were found within species of the diploid sect. *Cinnamomum*. Although these two copies are very similar, they can be distinguished easily due to a specific ca. 47-bp segment that is missing in the "short sequence" copy. The "long sequence" copy performed well in phylogenetic analysis of *Cinnamomum* and is largely consistent with phylogenetic relationships based on internal transcribed spacer region sequences. In contrast, the "short sequence" copy was problematic for phylogenetic reconstruction and PCR-mediated recombination was detected in 20 of the 38 *Cinnamomum* species with two *LEAFY* copies. Thirty-one recombinants were discriminated and the breakpoints suggested by the programs RDP3 and GARD were distributed randomly along the recombination sequences. This study shows that duplication in low-copy nuclear genes and problems associated with PCR-mediated recombination need to be given more attention in phylogenetic studies.

Key words: *Cinnamomum*, *LEAFY*, low-copy nuclear genes, PCR-mediated recombination.

The importance of using low-copy nuclear genes for phylogenetic analyses has long been recognized (Strand et al., 1997; Sang, 2002; Mort & Crawford, 2004; Whittall et al., 2006; Yuan et al., 2009; Duarte et al., 2010; Zhang et al., 2012). These genes hold great potential to improve the robustness of phylogenetic reconstruction at all taxonomic levels, especially where universal markers derived from chloroplast DNA (cpDNA) and nuclear ribosomal DNA (nrDNA) are unable to generate strong phylogenetic hypotheses (Sang, 2002). However, there are several factors that have hampered the utility of low-copy nuclear genes in phylogenetic analyses. First, compared with uniparentally inherited plastid DNA or nrDNA homogenized by concerted evolution, low-copy nuclear genes generally lack universal markers or polymerase chain reaction (PCR) primers applicable to the majority of plant groups, therefore requiring taxon-specific extra steps such as PCR primer design and cloning (Sang, 2002; Zimmer & Wen, 2013). Second, the interpretation of low-copy nuclear genes is more difficult, due to their biparental inheritance. Finally, some unpredictable events may occur; for example, *in vivo* recombination, gene conversion (de Sa & Drouin, 1996; Ford & Gottlieb, 1999; Gaut et al., 1999), or PCR-mediated recombination (Scharf et al., 1988; Bradley & Hillis, 1997; Cronn et al., 2002).

Nevertheless, despite these possible barriers, low-copy nuclear genes are an alternative option for recalcitrant taxa where plastid DNA and nrDNA show limited utility for inferring phylogenetic relationships and/or nrDNA is not completely homogenized, such as in Fagaceae (Oh & Manos, 2008), Lauraceae (Li et al., 2011), and Magnoliaceae (Nie et al., 2008). Furthermore, low-copy nuclear genes are useful for tracking relationships in lineages where speciation involves hybridization and/or subsequent alloploidy (Harris & Ingram, 1991; Small et al., 1998; Sang & Zhang, 1999).

Among the commonly used low-copy nuclear genes, *LEAFY* has been shown to be useful for phylogenetic studies, as its introns generally display a high level of sequence divergence at the species level (Hoot & Taylor, 2001; Archambault & Bruneau, 2004; Schlüter et al., 2007) and the second intron of *LEAFY* in particular has been used in a range of phylogenetic

studies (e.g., Nie et al., 2008; Tu et al., 2008; Kim et al., 2010; Li et al., 2011). *LEAFY* is a homeotic gene that regulates floral meristem induction during the early stages of reproductive ontogeny (Schultz & Haughn, 1991; Weigel, 1995; Blázquez, 1997). It is considered to have only a single copy in diploid angiosperms (Frohlich & Parker, 2000; Himi et al., 2001), although the copy number can vary in polyploids (Kelly et al., 1995; Frohlich & Parker, 2000; Wada et al., 2002).

Cinnamomum Schaeff. is a species-rich genus in the family Lauraceae, with ca. 350 species ranging from subtropical to tropical Asia, tropical America, Australia, and the Pacific Islands (Rohwer, 1993). Its members have been long recognized for their economic importance as sources of camphor, spices, medicines, and high-grade wood. Moreover, as important elements of tropical and subtropical evergreen broad-leaved forests, *Cinnamomum* species are also ecologically important. Tropical Asia has the highest species diversification in *Cinnamomum*, with a further ca. 47 species in the Neotropics (Lorea-Hernández, 1996). The Asian species are divided traditionally into two sections, sect. *Camphora* Meissn. and sect. *Cinnamomum*; all five Australian native *Cinnamomum* species belong to sect. *Cinnamomum* based on their morphology (Hyland, 1989).

Molecular phylogenetic relationships within this genus are poorly understood, due in part to a lack of applicable molecular markers at the species level. Previous studies have shown that cpDNA markers are extremely limited for phylogenetic reconstruction within Lauraceae, particularly for the Perseeae–Laureae clade, which includes *Cinnamomum* (Rohwer, 2000; Chanderbali et al., 2001; Li et al., 2004, 2011; Rohwer & Rudolph, 2005; Fijridiyanto & Murakami, 2009; Rohwer et al., 2009). Although the internal transcribed spacer (ITS) region performs better than cpDNA markers, additional phylogenetically informative markers with higher resolution are still needed. Furthermore, there is evidence of incompletely concerted evolution and pseudogenes in the ITS region in some *Cinnamomum* species (Huang et al., unpublished data), reducing its utility for phylogenetic interpretation.

Hitherto, two low-copy nuclear genes (*LEAFY* and *RPB2*) have been used for phylogenetic analyses in Lauraceae and they appear to show strong phylogenetic signal (Fijridiyanto & Murakami, 2009; Li et al., 2011). In the present study, we aim to assess the performance of the *LEAFY* second intron for phylogenetic reconstruction in *Cinnamomum*.

Material and Methods

Taxon sampling, sequencing, and alignment

A total of 63 *Cinnamomum* species were sampled (Table 1), covering nearly all of the distribution ranges. Four core Laureae species were selected as closely related outgroups based on previous studies (Chanderbali et al., 2001; Li et al., 2011). Total genomic DNA was extracted from silica gel-dried material or herbarium specimens using the Plant Geneomic DNA Kit (Tiangen Biotech Co., Beijing, China). The second intron of the *LEAFY* gene was amplified using the primers and program reported in Li et al. (2011). The amplified products were purified using the EZNA Cycle-Pure Kit (Omega Bio-Tek, Norcross, GA, USA) and then cloned using the pEASY-T3 Cloning Kit (TransGen Biotech, Beijing, China). At least five positive clones from each individual sample were sequenced and more than 10 positive clones were sequenced for some samples. The sequencing results were assembled and edited using the program Sequencher 4.5 (GeneCodes, Ann Arbor, MI, USA). The *LEAFY* dataset was aligned using Muscle 3.8.31 (Edgar, 2004) and then manually adjusted using BioEdit 7.0.9.0 (Hall, 1999). One of the identical sequences from each individual was retained for phylogenetic analyses. The corresponding ITS sequences of these 63 *Cinnamomum* species and partial *LEAFY* sequences have been used elsewhere (Huang et al., 2016). The ITS sequences of outgroups were downloaded from GenBank (Li et al., 2011); the *LEAFY* sequences originally used in this study have been deposited in GenBank (accession Nos. KU174395–KU174444, Table 1).

Recombination analyses

Based on the original alignment and phylogenetic analyses of the dataset comprised by all the different clone sequences, we identified two *LEAFY* copies in some *Cinnamomum* species. Considering the risk of PCR-mediated recombination (Scharf et al., 1988; Bradley & Hillis, 1997), two programs, RDP3 (Martin et al., 2010) and GARD (Pond et al., 2006), were applied to inspect the recombination evidence for each dataset represented by individual clone sequences, as the use of a single program like RDP3 can sometimes generate spurious results (Bruen et al., 2006). These recombinants were then removed from subsequent phylogenetic analyses. Both RDP3 and GARD can infer the position of breakpoints and RDP3 also identifies the most likely recombinants and their parental sequences.

Phylogenetic analyses

Phylogenetic analyses of the *LEAFY* and ITS datasets were carried out by maximum parsimony (MP) using the program PAUP*4.0b10 (Swofford, 2003) and Bayesian inference using the program MrBayes 3.1.2 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003). The MP analysis was implemented with the following heuristic search options: tree bisection–reconnection branch swapping, collapse of zero length branches, MulTrees on, with 1000 random taxon addition, saving 100 trees from each random sequence addition. All character states were treated as unordered and weighted equally. Bootstrap support values (BS) for internal nodes were estimated with 100 heuristic bootstrap replicates, using the same options described above. In the Bayesian inference, Modeltest 3.7 (Posada & Crandall, 1998; Posada & Buckley, 2004) was used to select the best-fit evolutionary model and gamma rate heterogeneity using the Akaike information criterion. The Markov chain Monte Carlo algorithm was run for 5 000 000 generations with one cold and three heated chains, starting from random trees and sampling one out of every 500 generations. Examination of the log likelihood values suggested that stationarity was reached in approximately 200 000 generations. Thus the first 1000 trees (500 000 generations) were discarded to make sure the burn-in period was sufficiently long and the remaining 9000 trees used to construct the 50% majority rule consensus tree with the proportion of bifurcations found in this consensus tree given as posterior probabilities (PP).

Table 1 Taxon sampling, vouchers, collecting sites, GenBank accession numbers for this study, and the breakpoint positions detected by programs GARD and RDP3

Taxon	Voucher	Locality	*LEAFY*		T/L/S/R	Position of breakpoint(s)	
			Long sequence	Short sequence		GARD	RDP3
Cinnamomum Schaeff.							
sect. *Camphora* Meissn.							
C. *bodinieri* H. Lév.	J. Li et al. 2007212 (HITBC)	Zhejiang, China	KU140251		5/–/–/–		
C. *camphora* (L.) J. Presl	X. Q. Ci et al. CXQ265 (HITBC)	Hunan, China	KU140253		6/–/–/–		
C. *chartophyllum* H. W. Li	X. Q. Ci et al. CXQ1730 (HITBC)	Yunnan, China	KU140259		7/–/–/–		
C. *glanduliferum* (Wall.) Meisn.	X. Q. Ci et al. CXQ2015 (HITBC)	Yunnan, China	KU140269		5/–/–/–		
C. *longepaniculatum* (Gamble) N. Chao ex H. W. Li	J. Li et al. CXQ452 (HITBC)	Sichuan, China	KU140283		6/–/–/–		
C. *longipetiolatum* H. W. Li	G. P. Yang 567 (HITBC)	Yunnan, China	KU140284		5/–/–/–		
C. *micranthum* (Hayata) Hayata	J. Li 2002160 (HITBC)	Guangxi, China	KU140286		6/–/–/–		
C. *mollifolium* H. W. Li	J. F. Huang, H-BN010 (HITBC)	Yunnan, China	KU140287		6/–/–/–		
C. *parthenoxylon* (Jack) Meisn.	J. F. Huang et al. 201312 (HITBC)	Bogor, Indonesia	KU140295		6/–/–/–		
C. *saxatile* H. W. Li	J. Li 20070136 (HITBC)	China	KU140306		5/–/–/–		
C. *septentrionale* Hand.-Mazz.	J. F. Huang, H-KZ05 (HITBC)	Yunnan, China	KU140307		6/–/–/–		
C. *tenuipile* Kosterm.	J. F. Huang, H-BN019 (HITBC)	Yunnan, China	KU140317		6/–/–/–		
sect. *Cinnamomum*							
C. *appelianum* Schewe	X. Q. Ci et al. CXQ204 (HITBC)	Hunan, China	KU140244		6/6/–/–		
C. *austrosinense* H. T. Chang	J. Li 2006041 (HITBC)	Guangdong, China	KU140245	KU174395	5/2/3/–		
C. *austroyunnanense* H. W. Li	X. C. Huang et al. JYL-2694 (HITBC)	Yunnan, China	KU140246	KU174396	6/3/3/–		
C. *baileyanum* (F. Muell. ex F. M. Bailey) Francis	D. Tng, CSIRO130 (CNS)	Queensland, Australia	KU140247	3L: KU174397 17L: KU174398	12/4/8/–		
C. *bejolghota* (Buch.-Ham.) Sweet	X. Q. Ci et al. CXQ105 (HITBC)	Yunnan, China	KU140249	KU174399	6/1/3/2	146, 397	152, 418
C. *burmannii* (Nees & T. Nees) Blume	X. Q. Ci et al. 200925 (HITBC)	Yunnan, China	KU140252	KU174400	6/2/1/3	438, 685	479, 677
C. *cassia* (L.) J. Presl	J. F. Huang, H-BN095 (HITBC)	Yunnan, China	KU140255	KU174401	6/2/2/2	583	598, 677
C. *celebicum* Miq.	J. F. Huang et al. 201319 (HITBC)	Bogor, Indonesia	KU140256	KU174402	6/2/1/3	105, 326, 528	91, 276, 552
C. *chaogo* B. S. Sun & H. L. Zhao	Yunda Expedition Team, YL029 (YUKU)	Yunnan, China	KU140257		8/8/–/–		
C. *chekiangense* Nakai	J. F. Huang, H-KZ30 (HITBC)	Yunnan, China	KU140260	KU174403	6/2/2/2	288, 610	276, 549
C. *cordatum* Kosterm.	S. Manickam, sm347 (KLU)	Kuala Lumpur, Malaysia	KU140261	3L: KU174404 7L: KU174405	6/2/2/2	497, 854	465, 508, 855
C. *crenulicupulum* Kosterm.	J. F. Huang, QSBG10 (HITBC)	Chiengmai, Thailand	KU140264	KU174406	7/2/3/2	492, 652	506, 620
C. *cuspidatum* Miq.	K. T. Yong, sm346 (KLU)	Kuala Lumpur, Malaysia	KU140265	KU174407	5/1/3/1	398	418
C. *daphnoides* Siebold & Zucc.	Leuenberger s.n. (HBG)	BG Halle, Germany	KU140266		12/12/–/–		
C. *doederleinii* Engl.	K. Inoue 1533 (KUN)	Ryukyu, Japan	KU140268	10L: KU174408 19L: KU174409 23L: KU174410	6/4/2/–		

Continued

Table 1 Continued

Taxon	Voucher	Locality	*LEAFY*		T/L/S/R	Position of breakpoint(s)	
			Long sequence	Short sequence		GARD	RDP3
C. grandifolium Cammerl.	J. F. Huang et al. 201314 (HITBC)	Bogor, Indonesia	KU140270	4L: KU174411 19L: KU174412	7/2/5/–		
C. heyneanum Nees	J. F. Huang, H-BN012 (HITBC)	Yunnan, China	KU140273	KU174413	5/3/2/–		
C. iners Reinw. ex Blume	X. C. Huang et al. JYL-4285 (HITBC)	Yunnan, China	KU140274	KU174414	5/2/2/1	146, 708	122, 700
C. insularimontanum Hayata	s.n. 78 (KUN)	Taiwan, China	KU140276	3L: KU174415 6L: KU174416 9L: KU174417	5/3/2/–		
C. japonicum Siebold	S. Yasuda 1351 (MO)	Honshu, Japan	KU140277	KU174418	6/3/3/–		
C. kotoense Kaneh. & Sasaki	J. F. Huang, H-KZ04 (HITBC)	Taiwan, China	KU140280	KU174419	5/3/2/–		
C. laubatii F. Muell.	C. Costion 1550 (CNS)	Queensland, Australia	KU140281	KU174420	10/4/6/–		
C. liangii C. K. Allen	X. Q. Ci et al. 20100043 (HITBC)	Zhejiang, China	KU140282	1L: KU174421 4L: KU174422	10/2/7/1	477	484
C. mairei H. Lév.	J. Li 2010265 (HITBC)	Sichuan, China	KU140285	KU174423	6/2/4/–		
C. multiflorum (Roxb.) Wight	J. F. Huang 201303 (HITBC)	Bogor, Indonesia	KU140288	KU174424	14/5/6/3	326, 520, 779	290, 522, 784
C. okinawense Hatus	K. Miyamato 52040 (QBG)	Shikoku, Japan	KU140289	KU174425	6/3/3/–		
C. oliveri F. M. Bailey	D. Tng 274 (CNS)	Queensland, Australia	KU140291	KU174426	8/3/3/2	434, 680	436, 694
C. pauciflorum Nees	X. Q. Ci et al. CXQ287 (HITBC)	Hunan, China	KU140296	KU174427	5/3/2/–		
C. pingbienense H. W. Li	X. Q. Ci et al. GBOWS955 (HITBC)	Yunnan, China	KU140297	4L: KU174429 6L: KU174428	6/2/3/1	364	298
C. pittosporoides Hand.-Mazz.	X. Q. Ci et al. CXQ1338 (HITBC)	Yunnan, China	KU140298	KU174430	6/1/5/–		
C. propinquum F. M. Bailey	D. Crayn 1156 (CNS)	Queensland, Australia	KU140301	KU174431	8/3/4/1	708	694
C. reticulatum Hayata	S. M. Liu 168 (KUN)	Taiwan, China	KU140303	3L: KU174432 14L: KU174433 15L: KU174434	7/4/3/–		
C. rhynchophyllum Miq.	J. F. Huang et al. 201323 (HITBC)	Bogor, Indonesia	KU140304	5L: KU174435 11L: KU174436	6/1/4/1	506	486
C. rigidissimum H. T. Chang	J. Li 2002127 (HITBC)	Guangxi, China	KU140305	KU174437	6/3/3/–		
C. sintoc Blume	J. F. Huang et al. 201320 (HITBC)	Bogor, Indonesia	KU140308		7/7/–/–		
C. subavenium Miq.	X. Q. Ci et al. CXQ303 (HITBC)	Zhejiang, China	KU140312	KU174438	6/3/2/1	169	180
C. tamala (Buch.-Ham.) T. Nees & Ebem.	J. F. Huang, H-BN018 (HITBC)	Yunnan, China	KU140315	KU174439	7/4/2/1	524	542
C. tenuifolium (Makino) Sugim.	K. Fumihiro 35747 (BO)	Honshu, Japan	KU140316	KU174440	8/6/2/–		
C. tonkinense (Lecomte) A. Chev.	M. Deng, DM01 (HITBC)	Bavi, Vietnam	KU140319	KU174441	6/3/2/1	796	808
C. tsangii Merr.	J. F. Huang & L. Li, H-WZS07 (HITBC)	Hainan, China	KU140324		5/5/–/–		

Continued

Table 1 Continued

Taxon	Voucher	Locality	*LEAFY*		T/L/S/R	Position of breakpoint(s)	
			Long sequence	Short sequence		GARD	RDP3
C. tsoi C. K. Allen	J. Li et al. 2010021 (HITBC)	Zhejiang, China	KU140325	KU174442	6/2/3/1	726	783
C. verum J. Presl.	J. F. Huang, H-BN013 (HITBC)	Yunnan, China	KU140326	KU174443	6/3/3/–		
C. wilsonii Gamble	X. Q. Ci et al. CXQ440 (HITBC)	Sichuan, China	KU140328	KU174444	7/2/4/1	709	743
Neotropical *Cinnamomum*							
C. costaricanum (Mez & Pittier) Kosterm.	W. D. Stevens & O. M. Montiel 29591 (MO)	Jinotega, Nicaragua	KU140263		7/–/–/–		
C. haussknechtii (Mez) Kosterm.	F. G. Lorea-Hernandez 5587 (MO)	Minas Gerais, Brazil	KU140271		5/–/–/–		
C. oleifolium (Mez) Kosterm.	F. G. Lorea-Hernandez 5582 (MO)	Minas Gerais, Brazil	KU140290		6/–/–/–		
C. padiforme (Standl. & Steyerm.) Kosterm.	W. D. Stevens et al. 30123 (MO)	Jinotega, Nicaragua	KU140292		5/–/–/–		
C. quadrangulum Kosterm.	F. G. Lorea-Hernandez 5585 (MO)	Minas Gerais, Brazil	KU140302		6/–/–/–		
C. subsessile (Meisn.) Kosterm.	H. van der Werff et al. 16822 (MO)	Amazonas, Peru	KU140313		7/–/–/–		
C. tomentulosum Kosterm.	F. G. Lorea-Hernandez 5590 (MO)	Minas Gerais, Brazil	KU140318		5/–/–/–		
C. triplinerve (Ruiz & Pav.) Kosterm.	W. D. Stevens 31499 (MO)	Chontales, Nicaragua	KU140321		7/–/–/–		
Outgroup							
Lindera erythrocarpa Makino	J. Li & L. Li 20070203 (HITBC)	Zhejiang, China	HQ697170				
Lindera megaphylla Hemsl.	L. Li 20070236 (HITBC)	Sichuan, China	HQ697173				
Litsea auriculata S. S. Chien & W. C. Cheng	J. Li & L. Li 2007095 (HITBC)	Zhejiang, China	HQ697174				
Litsea verticillata Hance	L. Li & Z. H. Wang 20070337 (HITBC)	Hainan, China	HQ697175				

–, No data available. T/L/S/R, total number of clones of the sample/number of "long sequence" clones/number of "short sequence" clones/number of recombination clones. Herbaria: BO, Herbarium Bogoriense; CNS, Australian Tropical Herbarium; HBG, Herbarium Hamburgense; HITBC, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences; KLU, Herbarium University of Malaya; KUN, Kunming Institute of Botany, Chinese Academy of Sciences; MO, Missouri Botanical Garden; QBG, Queen Sirikit Botanic Garden; YUKU, Yunnan University.

Results

Sequence characters

A total of 348 different clone sequences were obtained from the 63 sampled *Cinnamomum* species and recombinant sequences determined by both RDP3 and GARD were removed. One of several different sequences obtained from each individual was retained randomly if they all formed a clade. The final dataset was comprised of 113 ingroup and an additional four outgroup sequences downloaded from GenBank. This dataset contained 961 aligned positions and yielded 223 informative sites. The length of the aligned sequences was 98 bp for the second exon, 34 bp for the third exon, and included the entire second intron region.

Polymerase chain reaction-mediated recombination

Original alignment and phylogenetic analyses based on all the different clone sequences indicated that two *LEAFY* copies were present in 38 of the 43 sampled members of sect. *Cinnamomum* species (Table 1). However, this phenomenon did not appear in species from sect. *Camphora* or Neotropical *Cinnamomum*. Following inspection by the programs RDP3 and GARD, 31 of the 348 different clone sequences were diagnosed as recombinants, distributed in 20 of the 38 sect. *Cinnamomum* species with two copies (Table 1). Numbers of breakpoints inferred per individual dataset by RDP3 and GARD varied between one and three, but the breakpoint positions detected by RDP3 did not correspond precisely to those detected by GARD (Table 1). Apart from a single position (91) inferred by RDP3 in the second exon, all other breakpoints inferred by both RDP3 and GARD were located in the second intron. The distribution of the breakpoints along the recombination sequences seemed to be random and no recombination hotspots were observed (Table 1, Figs. S1, S2). This pattern was regarded as a key property of the recombinants from PCR-mediated recombination (Cronn et al., 2002; Lahr & Katz, 2009). Apart from an exceptional recombinant of *C. burmannii* (Nees & T. Nees) Blume, all other chimeric clones were obtained only once. Two recombination events were detected in four chimeric clones (distributed in *C. bejolghota* (Buch.-Ham.) Sweet, *C. burmannii*, *C. celebicum* Miq., and *C. iners* Reinw. ex Blume; Fig. S3).

Phylogenetic analyses

The substitution model of HKY + G for the *LEAFY* region suggested by Modeltest 3.7 was used in Bayesian inference. The consensus trees obtained from both MP (not shown) and Bayesian analysis (Fig. 1) were mostly congruent in their topologies, with three well-supported monophyletic clades recovered within *Cinnamomum*. Clade 1 (BS = 59, PP = 1.00; Fig. 1) included nearly all Asian sect. *Camphora* species and three sect. *Cinnamomum* clones (indicated with Ψ; Fig. 1), the Neotropical *Cinnamomum* species formed Clade 3 (BS = 100, PP = 1.00; Fig. 1). All the remaining ingroup samples, including two sect. *Camphora* species (*C. saxatile* H. W. Li and *C. longipetiolatum* H. W. Li), nearly all Asian sect. *Cinnamomum* sequences and the four sampled Australian species formed Clade 2 (BS = 92, PP = 1.00; Fig. 1). For the four outgroup taxa, *Lindera erythrocarpa* Makino, *Lindera megaphylla* Hemsl., and *Litsea auriculata* S. S. Chien & W. C. Cheng formed a clade, while *Litsea verticillata* Hance independently formed a branch showing a close relationship with Clade 3.

Within Clade 2, two robustly supported monophyletic sister subclades were recovered, here named the "long sequence" (BS = 100, PP = 1.00; Fig. 1) and "short sequence" clades (BS = 100, PP = 1.00; Fig. 1). They represent duplicated copies of *LEAFY*, differing by the presence of a ca. 47-bp deletion in the second intron of the "short sequence" copy (Figs. S3, S4). However, only one *LEAFY* copy was found in sect. *Camphora* and Neotropical *Cinnamomum* species and the different clone sequences of each individual of these two groups and the "long sequence" copy can form a monophyletic group. In contrast, in the "short sequence" clade, few species were represented by two clones (e.g., *C. baileyanum* (F. Muell. ex F. M. Bailey) Francis and *C. insularimontanum* Hayata), which cannot form monophyletic relationships. Two sect. *Camphora* species, *C. saxatile* and *C. longipetiolatum* (indicated with a *, Fig. 1; Li et al., 1982), were nested unexpectedly in sect. *Cinnamomum*. However, they were found to possess only one *LEAFY* copy, similar to the sect. *Camphora* species of Clade 1 and unlike most sect. *Cinnamomum* species.

Based on the phylogenetic analyses, separate *LEAFY* datasets were assembled for the "short sequence" and "long sequence" clades. For both datasets, the three aberrant sect. *Cinnamomum* sequences (indicated with Ψ; Fig. 1) nested in Clade 1 were removed. The topology of the three main clades of *Cinnamomum* in the consensus tree of the ITS dataset (cladogram C in Fig. 2) was consistent with the consensus tree of the *LEAFY* dataset without the "short sequence" clade (cladogram A in Fig. 2), but was inconsistent with consensus trees derived from both the *LEAFY* dataset without the "long sequence" clade (cladogram B in Fig. 2) and the complete *LEAFY* dataset (Fig. 1). Unlike the complete *LEAFY* dataset, the four outgroup species formed a clade in both the separate *LEAFY* and the ITS datasets, but received only weak bootstrap and moderate to strong Bayesian support (Fig. 2).

Discussion

LEAFY paralogs

The *LEAFY* gene ranks among the most frequently and successfully used low-copy nuclear markers in phylogenetic studies (Zimmer & Wen, 2013) and is generally considered to be a single copy gene in diploid angiosperms (Frohlich & Parker, 2000; Himi et al., 2001). To date, all karyologically examined *Cinnamomum* species are diploid (Mehra & Bawa, 1969; Okada & Tanaka, 1975; Sritharan et al., 1992). Although we did not use Southern hybridization and the exact number of *LEAFY* copies in the *Cinnamomum* genome is unknown, the "long sequence" and "short sequence" clades strongly indicate the presence of at least two copies in most members of sect. *Cinnamomum*. The duplication of the *LEAFY* gene appears to have occurred after the separation of sect. *Cinnamomum* from the rest of the genus, although the reason why there are at least five sect. *Cinnamomum* species with only the "long sequence" copy (indicated with **, Fig. 1) requires further investigation. The finding of more than one copy of the *LEAFY* gene in some, but not all members of *Cinnamomum* suggests that this gene may have a complex pattern of duplication.

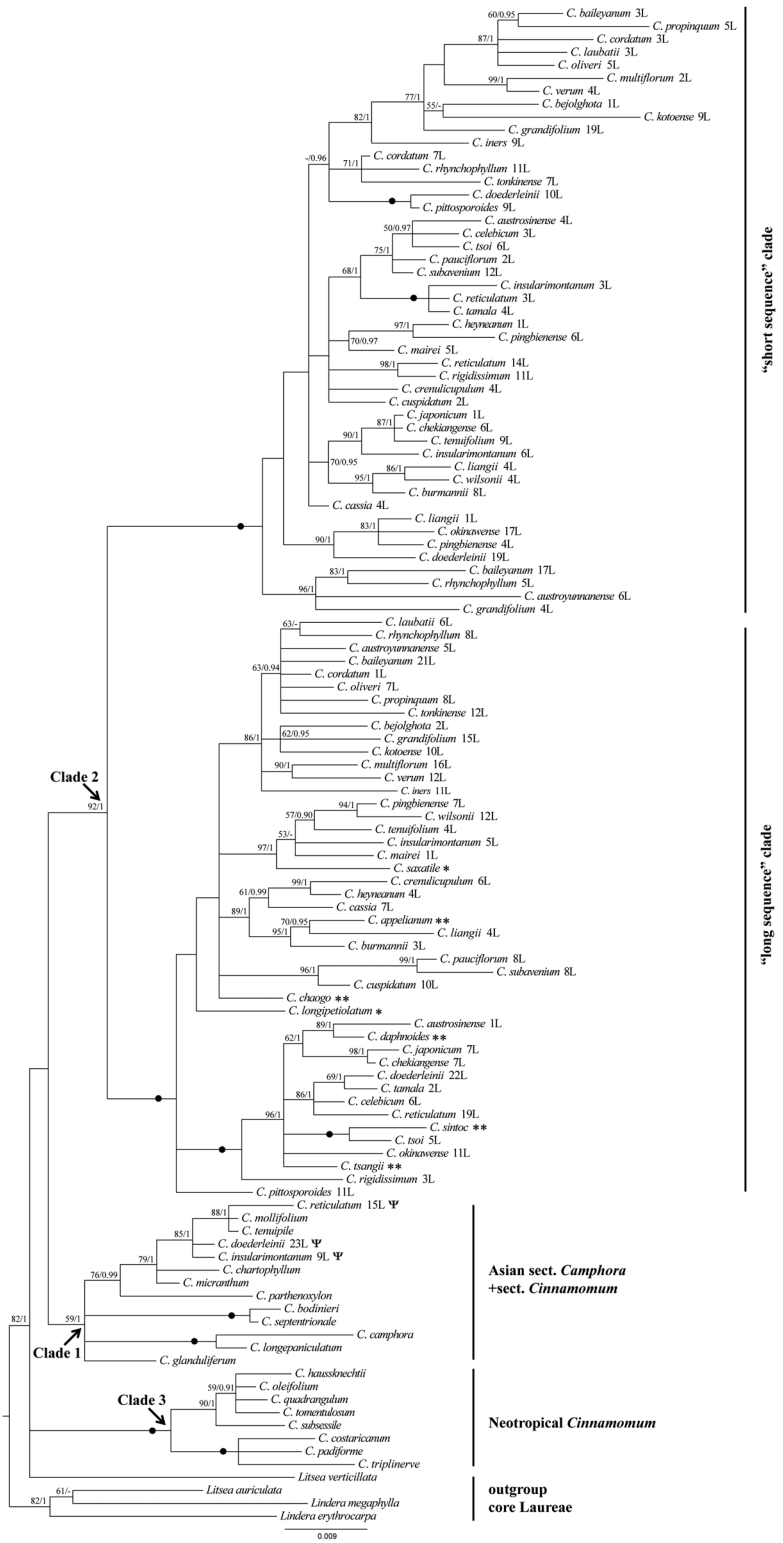

Fig. 1. Fifty percent majority rule tree obtained from Bayesian analysis of the *LEAFY* dataset. Bootstrap support values (≥50)/ Bayesian posterior probabilities (≥0.90) are shown above the branches.●, Bootstrap support value 100 and Bayesian posterior probabilities 1.00; Ψ, Three aberrant sect. *Cinnamomum* clones nested in sect. *Camphora*; *, Two sect. *Camphora* species nested in sect. *Cinnamomum*; **, Five sect. *Cinnamomum* species with only "long sequence" copy. *C.*, *Cinnamomum*. Vertical bars to the right circumscribe main clades.

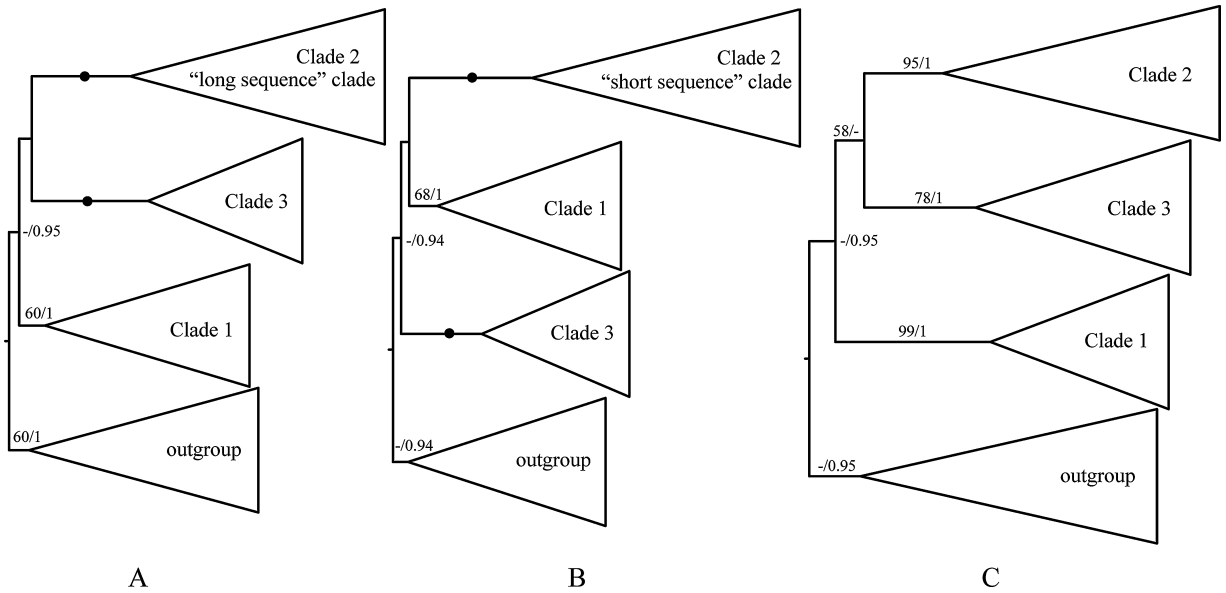

Fig. 2. Cladograms of *Cinnamomum* derived from Bayesian analysis with only the phylogenetic relationships among the main clades shown. Cladogram A is derived from the *LEAFY* dataset without the "short sequence" clade and the H81ff + G substitution model is used. Cladogram B is derived from the *LEAFY* dataset without the "long sequence" clade and the HKY + G substitution model is used. Cladogram C is derived from the internal transcribed spacer dataset and the TIM + I + G substitution model is used. Bootstrap support values (≥50)/Bayesian posterior probabilities (≥0.90) are shown above the branches.•, Bootstrap support values of 100 and Bayesian posterior probabilities 1.00.

The placement of three sect. *Cinnamomum* clones within the mainly sect. *Camphora* clade is surprising. Contamination by sect. *Camphora* DNA was ruled out, as a unique 27-bp insertion was present in two of these three sequences, but at present we cannot determine if these three sequences belong to a third *LEAFY* copy. *LEAFY* pseudogenes have been detected in some organisms (Southerton et al. 1998; Frohlich & Parker, 2000); however, as only very short parts of the exon regions were sequenced in our study, we cannot compare the entire exon for these three clones against the "long sequence" and "short sequence" copies to detect if there are frameshift mutations or stop codons to show the presence of pseudogenes. Hitherto, no diploid angiosperms have been reported expressing two functional *LEAFY* copies, so we assume that only one of the two *LEAFY* copies in *Cinnamomum* is functional, but further work is needed to test this assumption.

Polymerase chain reaction-mediated recombination

The PCR-based methods as the norm have greatly enhanced the field of molecular biology. Polymerase chain reaction is used in such studies, like population genetic analysis and phylogenetic reconstruction, mainly because the isolated sequence faithfully replicates the targeting of the specific DNA region of interest. However, this important laboratory technique is not error-free. One of the drawbacks is the phenomenon of PCR-mediated recombination resulting from *in vitro* chimera formation between amplification products from two or more templates (Scharf et al., 1988; Bradley & Hillis, 1997; Cronn et al., 2002). Chimeric PCR products can arise when two or more highly similar, non-identical target sequences are present in a single PCR amplification reaction and can be primarily attributed to the periodic formation of incompletely extended PCR products (Meyerhans et al., 1990; Shammas et al., 2001). This phenomenon has been seen most frequently in experimental templates (Meyerhans et al., 1990; Odelberg et al., 1995; Yang et al., 1996; Judo et al., 1998; Shammas et al., 2001), but has been demonstrated for natural templates from heterozygous *Adh* loci of pocket gophers (*Geomys* Rafinesque, 1817) (Bradley & Hillis, 1997), polyploid cotton (Cronn et al., 2002), and the *GapC* gene family of *Hibiscus tiliaceus* L. (Wu et al., 2007). In phylogenetic studies, PCR-mediated recombination violates the central assumption that the isolated sequence faithfully replicates the target of interest, so unrecognized, artificially produced recombinant products may therefore bias results or alter interpretations in phylogeny and evolutionary studies (Cronn et al., 2002).

The *LEAFY* "long sequence" and "short sequence" copies are highly similar, especially in the exon regions (Figs. S3, S4), and can be amplified simultaneously using the designed primer pair. A typical cycling program was employed and no measures were used to suppress the PCR-mediated recombination in the present study. Accordingly, 31 chimeric sequences were detected from 20 of the 38 sect. *Cinnamomum* species (52.6%) with two *LEAFY* copies. These chimeras used in phylogenetic reconstruction will mislead our inference for the evolutionary relationship of *Cinnamomum* species.

Identification of PCR-mediated recombination

Recombinant PCR sequences are not always easy to diagnose and in some cases may be deconstructed into parental sequences, especially when a large number of clones are examined (Cronn et al., 2002). This may be aggravated by the failed amplification of one or both parental sequences, though

generally parental sequences are more likely to be cloned and sequenced than their recombinants. In allopolyploids, the identification of recombinants is more difficult. Although no evidence of *in vivo* recombination was found in the present study, awareness of the need to discriminate between *in vivo* and PCR-mediated recombination is important. Discrimination may be necessary in some studies, such as hybrid speciation.

Compared with *in vivo* recombination, no recombination-derived hotspots are known for PCR-mediated recombinants (Cronn et al., 2002; Lahr & Katz, 2009). The test from programs RDP3 and GARD show that the breakpoint positions randomly distribute along the chimeric sequences and no breakpoint hotspot was observed (Table 1; Figs. S1, S2).

In addition, we would expect that the amplification success rate of the recombinants from *in vivo* recombination is equal to their parental sequences during a PCR cycle and those recombinants can be sequenced repeatedly like their parental sequences. However, recombinant PCR sequences mostly represent a single clone and generally cannot be sequenced twice. Multiple recombination events may occur within one recombinant PCR sequence (Cronn et al., 2002; Kelly et al., 2010). Among the 31 recombinants in present study, excluding an exceptional sequence that was obtained twice, each of the remainder was sequenced just once. Four recombination sequences showed a more complicated distribution of nucleotides indicative of two recombination events (Fig. S3).

Furthermore, as additional substitutions in PCR-mediated recombination are unlikely, it generally can be differentiated from historical recombination events by having fewer accumulated mutations (Posada et al., 2002). We inspected individual clone sequences by sight. For eight of the 31 recombinants, the two flank nucleotides separated by the breakpoints are correspondingly identical with their inferred parental sequences and no accumulated mutations were observed (Fig. S4). Within 16 recombinants, one flank is identical with the inferred parental sequences and the other flank shows one to three accumulated mutations (Figs. S3, S4). Seven recombinants show one to three nucleotides variation in each flank compared with the inferred parental sequences. However, the inferred parents may not be true and the high error rate of *Taq* polymerase used in this study may play a role in generating such mutations.

Control measures such as lengthening the extension times (Judo et al., 1998), reducing the number of amplification cycles (Odelberg et al., 1995), adding PCR additives such as betaine and dimethysulfoxide in PCR systems (Shammas et al., 2001), using proofreading process-enhanced polymerases, and reducing the initial template concentration (Lahr & Katz, 2009) have all been shown to be effective in decreasing the frequency of PCR-mediated recombination. However, these measures will not influence the amplification of *in vivo* recombination sequences and therefore can be used as an alternative method for differentiation from PCR-mediated recombination. Originally, as we did not expect to find *LEAFY* paralogs in *Cinnamomum*, no preventive methods were used to reduce the possible impact of PCR-mediated recombination in the current study. Nonetheless, based on the above-mentioned evidence, we consider that all of the chimeric sequences are derived from PCR-mediated recombination.

Phylogenetic utility of the *LEAFY* gene

Although the second intron of *LEAFY* has been used in many phylogenetic studies (e.g., Nie et al., 2008; Tu et al., 2008; Kim et al., 2010; Li et al., 2011), the presence of paralogs in some plant lineages may cause problems. However, given that the "long sequence" and "short sequence" copies can be identified easily, the second intron of *LEAFY* still represents a promising source of phylogenetic information, especially when using the "long sequence" copy.

For example, in the dataset that included 63 *Cinnamomum* species with only the "long sequence" copy, 301 of the 796 aligned second intron nucleotide positions (37.8%) were variable, whereas only 17 nucleotide positions (12.9%) were variable in the 132 aligned exon regions. This compares well with the equivalent ITS dataset where 182 of the 659 aligned nucleotide positions (27.6%) are variable. In contrast, the "short sequence" copy in several species was represented by two clones per individual that did not cluster together and this short copy was also absent in several sect. *Cinnamomum* species. The conflict between the ITS dataset (cladogram C in Fig. 2, Fig. S5), the *LEAFY* dataset with the "long sequence" copy excluded (cladogram B in Fig. 2), and the complete *LEAFY* dataset (Fig. 1) also suggest that the "short sequence" copy may be more problematic for phylogenetic inference of species relationships in *Cinnamomum*. Similarly, the placement of three sect. *Cinnamomum* sequences within sect. *Camphora* indicates that more research is needed into the precise nature of these paralogs, in particular, whether these three sequences may be pseudogenes.

This study suggests strongly that the genus *Cinnamomum* consists of three clades: sect. *Camphora*, sect. *Cinnamomum*, and the Neotropical species corresponding to our Clades 1, 2 and 3, respectively, except for the two aberrant sect. *Camphora* species (*C. saxatile* and *C. longipetiolatum*, indicated with * in Fig. 1) nested in sect. *Cinnamomum*. These three clades also received moderate to strong support in the analyses of the ITS dataset (Figs. 2, S5). The consensus tree topologies obtained from the *LEAFY* dataset excluding the "short sequence" copy and ITS dataset are congruent for the major clades and only minor variation in the relationships of a few terminal nodes was observed, possibly caused by insufficient phylogenetic signal in the data. Although the ITS region performs well relative to cpDNA (Chanderbali et al., 2001; Li et al., 2004; Rohwer et al., 2009; Li et al., 2011) and matches the *LEAFY* region for phylogenetic reconstruction within the family Lauraceae, the low sequencing success rate (Lu et al., 2013) and incompletely concerted evolution and pseudogenes (Huang et al., unpublished data) reduce its utility for phylogenetic interpretation. Hence, screening applicable low-copy nuclear gene makers, such as the *LEAFY* gene, is important for inferring evolutionary relationships in 'difficult' plant groups, including genus *Cinnamomum*.

Despite several adverse factors, low-copy nuclear genes will still be a significant phylogenetically informative source and are frequently used in phylogenetic reconstruction in nest years, particularly for plants, like *Cinnamomum*, where cpDNA and/or ITS do not work well. However, the high levels of genetic redundancy in eukaryotic genomes should make recombination artifacts common. Some classical low-copy nuclear genes markers, such as *GBSSI* (Evans et al., 2000;

Small, 2004), *RPB2* (Oxelman & Bremer, 2000), and *LEAFY* (Archambault & Bruneau, 2004; the present study), were first thought to be single copy genes, but have been duplicated in some lineages and may be found in many others. Therefore, PCR-mediated recombination may appear unexpectedly during the amplification of low-copy nuclear genes when other paralogs of these gene markers are present. In addition, as Cronn et al. (2002) suggested, PCR-mediated recombination may be problematic for highly reiterated sequences such as rDNA. Therefore, the phenomenon of PCR-mediated recombination and the possible issues stemming from its occurrence should be given more attention by evolutionary biologists.

Acknowledgements

We would like to thank the staff of the herbaria and botanical gardens and collectors in Table 1 for collection and permission to sample materials for molecular analysis. We thank Baogui Li, Mengmeng Lu, Yee Wen Low, Wong Khoon Meng, Monthn Norsaengsri, Ratchada Pongsattayapipa, Prachaya Srisanga, Manickam Sugumaran, Suyanee Vessabutr, Pingyuan Wang, Yong Xu, and Qing Yin for their assistance in sample collection. We are grateful for support from the Missouri Botanical Garden DNA Bank and James C. Solomon, Alyse Kuhlman, and Eric Feltz for providing American samples. We also thank Wumei Xu for his assistance in the Shapiro–Wilk test. This work was supported by the National Natural Science Foundation of China (Grant Nos. 31370245 and 31200167).

References

Archambault A, Bruneau A. 2004. Phylogenetic utility of the *LEAFY*/*FLORICAULA* gene in the Caesalpinioideae (Leguminosae): Gene duplication and a novel insertion. *Systematic Botany* 29: 609–626.

Blázquez MA. 1997. Illuminating flowers: *CONSTANS* induces *LEAFY* expression. *BioEssays* 19: 277–279.

Bradley RD, Hillis DM. 1997. Recombinant DNA sequences generated by PCR amplification. *Molecular Biology and Evolution* 14: 592–593.

Bruen TC, Philippe H, Bryant D. 2006. A simple and robust statistical test for detecting the presence of recombination. *Genetics* 172: 2665–2681.

Chanderbali AS, van der Werff H, Renner SS. 2001. Phylogeny and historical biogeography of Lauraceae: Evidence from the chloroplast and nuclear genomes. *Annals of the Missouri Botanical Garden* 88: 104–134.

Cronn R, Cedroni M, Haselkorn T, Grover C, Wendel JF. 2002. PCR-mediated recombination in amplification products derived from polyploid cotton. *Theoretical and Applied Genetics* 104: 482–489.

de Sa MM, Drouin G. 1996. Phylogeny and substitution rates of angiosperm actin genes. *Molecular Biology and Evolution* 13: 1198–1212.

Duarte JM, Wall PK, Edger PP, Landherr LL, Ma H, Pires JC, Leebens-Mack J. 2010. Identification of shared single copy nuclear genes in *Arabidopsis, Populus, Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evolutionary Biology* 10: 61.

Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797.

Evans RC, Alice LA, Campbell CS, Kellogg EA, Dickinson T. 2000. The Granule-Bound Starch Synthase (*GBSSI*) gene in the Rosaceae: Multiple loci and phylogenetic utility. *Molecular Phylogenetics and Evolution* 17: 388–400.

Fijridiyanto IA, Murakami N. 2009. Phylogeny of *Litsea* and related genera (Laureae-Lauraceae) based on analysis of *rpb2* gene sequences. *Journal of Plant Research* 122: 283–298.

Ford VS, Gottlieb LD. 1999. Molecular characterization of *PgiC* in a tetraploid plant and its diploid relatives. *Evolution* 53: 1060–1067.

Frohlich MW, Parker DS. 2000. The mostly male theory of flower evolutionary origins: From genes to fossils. *Systematic Botany* 25: 155–170.

Gaut BS, Peek AS, Morton BR, Clegg MT. 1999. Patterns of genetic diversification within the *Adh* gene family in the grasses (Poaceae). *Molecular Biology and Evolution* 16: 1086–1097.

Hall TA. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nuclear Acids Symbol Series* 41: 95–98.

Harris SA, Ingram R. 1991. Chloroplast DNA and biosystematics: The effect of intraspecific diversity and plastid transmission. *Taxon* 40: 393–412.

Himi S, Sano R, Nishiyama T, Tanahashi T, Kato M, Ueda K, Hasebe M. 2001. Evolution of *MADS*-box gene induction by *FLO/LFY* genes. *Journal of Molecular Evolution* 53: 387–393.

Hoot SB, Taylor WC. 2001. The utility of nuclear ITS, a *LEAFY* homolog intron, and chloroplast *atp*B-*rbc*L spacer region data in phylogenetic analyses and species delimitation in *Isoetes*. *American Fern Journal* 91: 166–177.

Huang JF, Li L, van der Werff H, Li HW, Rohwer JG, Crayn DM, Meng HH, van der Merwe M, Conran JG, Li J. 2016. Origins and evolution of cinnamon and camphor: A phylogenetic and historical biogeographical analysis of the *Cinnamomum* group (Lauraceae). *Molecular Phylogenetics and Evolution* 96: 33–44.

Huelsenbeck JP, Ronquist F. 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755.

Hyland BPM. 1989. A revision of Lauraceae in Australia (excluding *Cassytha*). *Australian Systematic Botany* 2: 138–162.

Judo MSB, Wedel AB, Wilson C. 1998. Stimulation and suppression of PCR-mediated recombination. *Nucleic Acids Research* 26: 1819–1825.

Kelly AJ, Bonnlander MB, Meeks-Wagner DR. 1995. NFL, the tobacco homolog of *FLORICAULA* and *LEAFY*, is transcriptionally expressed in both vegetative and floral meristems. *Plant Cell* 7: 225–234.

Kelly LJ, Leitch AR, Clarkson JJ, Hunter RB, Knapp S, Chase MW. 2010. Intragenic recombination events and evidence for hybrid speciation in *Nicotiana* (Solanaceae). *Molecular Biology and Evolution* 27: 781–799.

Kim C, Shin H, Chang YT, Choi HK. 2010. Speciation pathway of *Isoëtes* (Isoëtaceae) in East Asia inferred from molecular phylogenetic relationships. *American Journal of Botany* 97: 958–969.

Lahr DJ, Katz LA. 2009. Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. *Biotechniques* 47: 857–866.

Li HW, Pai PY, Lee SK, Wei FN, Wei YT, Yang YC, Huang PH, Cui HB, Xia ZD, Li JL. 1982. Lauraceae. In: *Flora Reipublicae Popularis Sinicae*. Beijing: Science Press. 31: 1–463.

Li J, Christophel DC, Conran JG, Li HW. 2004. Phylogenetic relationships within the 'core' Laureae (*Litsea* complex, Lauraceae) inferred from sequences of the chloroplast gene *mat*K and

nuclear ribosomal DNA ITS regions. *Plant Systematics and Evolution* 246: 19–34.

Li L, Li J, Rohwer JG, van der Werff H, Wang ZH, Li HW. 2011. Molecular phylogenetic analysis of the *Persea* group (Lauraceae) and its biogeographic implications on the evolution of tropical and subtropical amphi-Pacific disjunctions. *American Journal of Botany* 98: 1520–1536.

Lorea-Hernández FG. 1996. *A systematic revision of the Neotropical species of* Cinnamomum *Schaeffer (Lauraceae)*. Ph.D. Dissertation. St. Louis: University of Missouri.

Lu MM, Ci XQ, Yang GP, Li J. 2013. DNA barcoding of subtropical forest trees—a study from Ailao Mountains Nature Reserve, Yunnan, China. *Plant Diversity and Resources* 35: 733–741.

Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P. 2010. RDP3: A flexible and fast computer program for analyzing recombination. *Bioinformatics* 26: 2462–2463.

Mehra PN, Bawa KS. 1969. Chromosomal evolution in tropical hardwoods. *Evolution* 23: 466–481.

Meyerhans A, Vartanian JP, Wain-Hobson S. 1990. DNA recombination during PCR. *Nucleic Acids Research* 18: 1687–1691.

Mort ME, Crawford DJ. 2004. The continuing search: Low-copy nuclear sequences for lower-level plant molecular phylogenetic studies. *Taxon* 53: 257–261.

Nie ZL, Wen J, Azuma H, Qiu YL, Sun H, Meng Y, Sun WB, Zimmer EA. 2008. Phylogenetic and biogeographic complexity of Magnoliaceae in the Northern Hemisphere inferred from three nuclear data sets. *Molecular Phylogenetics and Evolution* 48: 1027–1040.

Odelberg SJ, Weiss RB, Hata A, White R. 1995. Template-switching during DNA synthesis by *Thermus aquaticus* DNA polymerase I. *Nucleic Acids Research* 23: 2049–2057.

Oh SH, Manos PS. 2008. Molecular phylogenetics and cupule evolution in Fagaceae as inferred from nuclear *CRABS CLAW* sequences. *Taxon* 57: 434–451.

Okada H, Tanaka R. 1975. Karyological studies in some species of Lauraceae. *Taxon* 24: 271–280.

Oxelman B, Bremer B. 2000. Discovery of paralogous nuclear gene sequences coding for the second largest subunit of RNA polymerase II (*RPB2*) and their phylogenetic utility in Gentianales of the Asterids. *Molecular Biology and Evolution* 17: 1131–1145.

Pond SLK, Posada D, Gravenor MB, Woelk CH, Frost SDW. 2006. GARD: A genetic algorithm for recombination detection. *Bioinformatics* 22: 3096–3098.

Posada D, Buckley TR. 2004. Model selection and model averaging in phylogenetics: Advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology* 53: 793–808.

Posada D, Crandall KA. 1998. Modeltest: Testing the model of DNA substitution. *Bioinformatics* 14: 817–818.

Posada D, Crandall KA, Holmes EC. 2002. Recombination in evolutionary genomics. *Annual Review of Genetics* 36: 75–97.

Rohwer JG. 1993. Lauraceae. In: Kubitzki K, Rohwer JG, Bittrich V eds. *The families and genera of vascular plants. Flowering plants – Dicotyledons*, vol. 2. Berlin: Springer-Verlag. 366–391.

Rohwer JG. 2000. Toward a phylogenetic classification of the Lauraceae: Evidence from *mat*K. *Systematic Botany* 25: 60–71.

Rohwer JG, Li J, Rudolph B, Schmidt SA, van der Werff H, Li HW. 2009. Is *Persea* (Lauraceae) monophyletic? Evidence from nuclear ribosomal ITS sequences. *Taxon* 58: 1153–1167.

Rohwer JG, Rudolph B. 2005. Jumping genera: The phylogenetic positions of *Cassytha*, *Hypodaphnis*, and *Neocinnamomum* (Lauraceae) based on different analyses of *trn*K intron sequences. *Annals of the Missouri Botanical Garden* 92: 153–178.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.

Sang T. 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology* 37: 121–147.

Sang T, Zhang D. 1999. Reconstructing hybrid speciation using sequences of low-copy nuclear genes: Hybrid origins of five *Paeonia* species based on *Adh* gene phylogenies. *Systematic Botany* 24: 148–163.

Scharf SJ, Friedmann A, Brautbar C, Szafer F, Steinman L, Horn G, Gyllensten U, Erlich HA. 1988. HLA class II allelic variation and susceptibility to pemphigus vulgaris. *Proceedings of the National Academy of Sciences USA* 85: 3504–3508.

Schlüter PM, Kohl G, Stuessy TF, Paulus HF. 2007. A screen of low-copy nuclear genes reveals the *LFY* gene as phylogenetically informative in closely related species of orchids (*Ophrys*). *Taxon* 56: 493–504.

Schultz EA, Haughn GW. 1991. *LEAFY*, a homeotic gene that regulates inflorescence development in *Arabidopsis*. *Plant Cell* 3: 771–781.

Shammas FV, Heikkilä R, Osland A. 2001. Fluorescence-based method for measuring and determining the mechanisms of recombination in quantitative PCR. *Clinica Chimica Acta* 304: 19–28.

Small RL. 2004. Phylogeny of *Hibiscus* sect. *Muenchhusia* (Malvaceae) based on chloroplast *rpL16* and *ndhF*, and nuclear ITS and *GBSSI* sequences. *Systematic Botany* 29: 385–392.

Small RL, Ryburn JA, Cronn RC, Seelanan T, Wendel JF. 1998. The tortoise and the hare: Choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. *American Journal of Botany* 85: 1301–1315.

Southerton SG, Strauss SH, Olive MR, Harcourt RL, Decroocq V, Zhu X, Liewellyn DG, Peacock WJ, Dennis ES. 1998. *Eucalyptus* has a functional equivalent of the *Arabidopsis* floral meristem identity gene *LEAFY*. *Plant Molecular Biology* 37: 897–910.

Sritharan R, Jacob VJ, Balasubramaniam S, Bavappa KVA. 1992. Palynological and cytological studies of the genus *Cinnamomum*. In: Levy A, Bernáth J, Craker LE eds. *WOCMAP I – Medicinal and Aromatic Plants Conference*. The Hague: International Society for Horticultural Science. 107–114.

Strand AE, Leebens-Mack J, Milligan BG. 1997. Nuclear DNA-based markers for plant evolutionary biology. *Molecular Ecology* 6: 113–118.

Swofford DL. 2003. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0b10. Sunderland: Sinauer Associates.

Tu T, Dillon MO, Sun H, Wen J. 2008. Phylogeny of *Nolana* (Solanaceae) of the Atacama and Peruvian deserts inferred from sequences of four plastid markers and the nuclear *LEAFY* second intron. *Molecular Phylogenetics and Evolution* 49: 561–573.

Wada M, Cao QF, Kotoda N, Soejima JI, Masuda T. 2002. Apple has two orthologues of *FLORICAULA/LEAFY* involved in flowering. *Plant Molecular Biology* 49: 567–577.

Weigel D. 1995. The genetics of flower development: From floral induction to ovule morphogenesis. *Annual Review of Genetics* 29: 19–39.

Whittall JB, Medina-Marino A, Zimmer EA, Hodges SA. 2006. Generating single-copy nuclear gene data for a recent adaptive radiation. *Molecular Phylogenetics and Evolution* 39: 124–134.

Wu L, Tang T, Zhou R, Shi S. 2007. PCR-mediated recombination of the amplification products of the *Hibiscus tiliaceus* cytosolic

glyceraldehyde-3-phosphate dehydrogenase gene. *Journal of Biochemistry and Molecular Biology* 40: 172–179.

Yang YL, Wang G, Dorman K, Kaplan AH. 1996. Long polymerase chain reaction amplification of heterogeneous HIV type 1 templates produces recombination at a relatively high frequency. *AIDS Research and Human Retroviruses* 12: 303–306.

Yuan YW, Liu C, Marx HE, Olmstead RG. 2009. The pentatricopeptide repeat (*PPR*) gene family, a tremendous resource for plant phylogenetic studies. *New Phytologist* 182: 272–283.

Zhang N, Zeng L, Shan H, Ma H. 2012. Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. *New Phytologist* 195: 923–937.

Zimmer EA, Wen J. 2013. Reprint of: Using nuclear gene data for plant phylogenetics: Progress and prospects. *Molecular Phylogenetics and Evolution* 662: 539–550.

Supplementary Material

The following supplementary material is available online for this article at http://onlinelibrary.wiley.com/doi/10.1111/jse.12189/suppinfo:

Fig. S1. Distribution of breakpoints detected by the program GARD along the 961-bp aligned length of the sequenced fragment. The distribution of breakpoints along the sequence does not deviate significantly ($P > 0.05$) from the expected normal distribution, as measured by the Shapiro–Wilk test.

Fig. S2. Distribution of breakpoints detected by the program RDP3 along the 961-bp aligned length of the sequenced fragment. The distribution of breakpoints along the sequence does not deviate significantly ($P > 0.05$) from the expected normal distribution, as measured by the Shapiro–Wilk test.

Fig. S3. Intraspecific base differences among the six clones of *Cinnamomum bejolghota* with the spaces generated during dataset alignment included. Clone 2L is a *LEAFY* "long sequence" copy and clones 1L, 6L, and 10L are *LEAFY* "short sequence" copies. Clones 3L and 7L are recombinants. The red rectangles A, B, C, and D indicate the inferred breakpoints detected by programs GARD and RDP3. The left-side nucleotides split by breakpoints (397 and 418) of clone 7L show a base variation compared with the inferred parental sequence (clone 2L); the right side is identical with the inferred parental sequences (clones 1L and 10L). Clone 3L was detected with two recombination events. The starting part before breakpoints 146 or 152 is identical with inferred parental sequence (clone 2L) and the ending part after breakpoints 397 or 418 shows a base variation compared with the inferred parental sequence (clone 2L). The middle part is identical with clone 1L.

Fig. S4. Intraspecific base differences among the six clones of *Cinnamomum cassia* (L.) J. Presl with the spaces generated during dataset alignment included. Clones 5L and 7L are identical and belong to the *LEAFY* "long sequence" copy. Clones 2L and 4L belong to the *LEAFY* "short sequence" copy and clones 3L and 8L are recombinants. The red rectangles A, B, and C indicate the inferred breakpoints detected by programs GARD and RDP3. The left-side nucleotides split by breakpoints (583 and 598) of clone 3L are identical with the inferred parental sequences (clones 5L and 7L); the right side shows a base variation compared with the inferred parental sequence (clone 2L). Both sides split by the breakpoint (677) of clone 8L are identical with the inferred parental sequences (clones 5L, 7L and 2L).

Fig. S5. Fifty percent majority rule tree obtained from Bayesian analysis of the internal transcribed spacer dataset. Bootstrap support values ($\geq$50)/Bayesian posterior probabilities ($\geq$0.90) are shown above the branches. Vertical bars to the right circumscribe main clades. *, Two sect. *Camphora* species nested in sect. *Cinnamomum*.