MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2010) **10**, 594–605 doi: 10.1111/j.1755-0998.2009.02815.x

DNA BARCODING

Species identification of *Alnus* (Betulaceae) using nrDNA and cpDNA genetic markers

BAO-QING REN,*† XIAO-GUO XIANG*† and ZHI-DUAN CHEN*
**State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Xiangshan, Beijing 100093, China, †Graduate School of the Chinese Academy of Sciences, Beijing 100039, China*

Abstract

One nuclear and three chloroplast DNA regions (ITS, *rbcL*, *matK* and *trnH-psbA*) were used to identify the species of *Alnus* (Betulaceae). The results showed that 23 out of all 26 *Alnus* species in the world, represented by 131 samples, had their own specific molecular character states, especially for three morphologically confused species (*Alnus formosana*, *Alnus japonica* and *Alnus maritima*). The discriminating power of the four markers at the species level was 10% (*rbcL*), 31.25% (*matK*), 63.6% (*trnH-psbA*) and 76.9% (ITS). For ITS, the mean value of genetic distance between species was more than 10 times the intraspecific distance (0.009%), and 13 species had unique character states that differentiated them from other species of *Alnus*. The *trnH-psbA* region had higher mean values of genetic distance between and within species (2.1% and 0.68% respectively) than any other region tested. Using the *trnH-psbA* region, 13 species are distinguished from 22 species, and seven species have a single diagnostic site. The combination of two regions, ITS and *trnH-psbA*, is the best choice for DNA identification of *Alnus* species, as an improvement and supplement for morphologically based taxonomy. This study illustrates the potential for certain DNA regions to be used as novel internet biological information carrier through combining DNA sequences with existing morphological character and suggests a relatively reliable and open taxonomic system based on the linked DNA and morphological data.

Keywords: *Alnus*, DNA barcoding, ITS, molecular identification, morphological taxonomy, *trnH-psbA*

Received 27 July 2009; revision received 27 October 2009; accepted 10 November 2009

Introduction

Alnus Mill. (Betulaceae), an anemophilous woody genus, is distributed throughout the Northern Hemisphere. *Alnus* species are characterized by their strobilus-like woody infructescences with persistent scales and their symbiotic relationship with the nitrogen-fixing actinomycete *Frankia*, which induces formation of root nodules (Benson & Silvester 1993). Phylogenetic and biogeographical studies of *Alnus* using morphological and molecular data (e.g. Bousquet *et al.* 1992; Chen *et al.* 1999; Chen & Li 2004) support its monophyly and sister relationship with *Betula* L. There are 29–35 species of *Alnus* in the world, with 9 species in the New World, 4–5 in Europe and 18–23 in Asia (Murai 1964; Furlow 1979; Chen 1994; Govaerts & Frodin 1998). However, taxonomy of *Alnus* is difficult, particularly for several species pairs or complexes, including *Alnus incana* (L.) Moench ssp. *incana* and *Alnus glutinosa* (L.) Gaertn., *Alnus trabeculosa* Hand.-Mazz. and *Alnus japonica* (Thunb.) Steud., *Alnus formosana* (Burkill) Makino and *A. japonica*.

In the past three decades, molecular systematics has become a widely accepted and adopted approach to reconstruct phylogeny. Based on molecular techniques, DNA barcoding was proposed as a new biological tool to attain accurate, rapid and automatable species identification without morphological knowledge by using short and standardized gene or DNA regions that can be amplified easily by polymerase chain reaction (PCR) (Hebert *et al.* 2003). Combining DNA sequences with existing morphological characters accelerates the rate of

Correspondence: Zhi-Duan Chen, Fax: 861062590843;
E-mail: zhiduan@ibcas.ac.cn

classification and identification for global biological species (Smith *et al.* 2005; Will *et al.* 2005; DeSalle 2006; Hajibabaei *et al.* 2007).

Most of the previous barcode studies in plants were carried out on a large scale to find universal and consistent makers for angiosperms or land plants (e.g. Chase *et al.* 2005, 2007; Kress *et al.* 2005; Cowan *et al.* 2006; Newmaster *et al.* 2006; Presting 2006; Kress & Erickson 2007; Sass *et al.* 2007; Erickson *et al.* 2008; Fazekas *et al.* 2008; Lahaye *et al.* 2008; Devey *et al.* 2009; Ford *et al.* 2009). On the other hand, some authors used one or several candidate markers to test their appropriateness through dense sampling in a single family or genus, such as Hymenophyllaceae (Nitta 2008), *Compsoneura* Warb. (Newmaster *et al.* 2008), *Heracleum* L. (Logacheva *et al.* 2008), *Aspalathus* L. (Edwards *et al.* 2008), *Acacia* Mill. (Newmaster & Ragupathy 2009), *Carex* L. (Starr *et al.* 2009) and *Crocus* L. (Seberg & Petersen 2009). DNA barcoding, albeit controversial (Will *et al.* 2005), has provided an alternative potential means to help identify species in plant taxa.

In this study, we use four DNA regions (*rbcL*, *matK*, *trnH-psbA* and ITS) to propose a DNA barcoding protocol and database for differentiating species of *Alnus*, which not only contributes to taxonomy of *Alnus* but also provides a benchmark data for biological and ecological studies of *Alnus*. We address the following issues: (i) whether there are appropriate markers that can be used to identify *Alnus* species from the whole genus or not and (ii) how to utilize molecular data as a rapid and accurate convenient tool to complement morphological taxonomy.

Materials and methods

Materials

Multiple samples of each species recognized in the taxonomic revision of Furlow (1979) for new world species and our unpublished data for Eurasian species were included in this study to cover both morphological and geographical range of each taxon. In total, we sampled 131 individuals representing all 26 species of *Alnus* (see Appendix S1, Supporting Information). Three species of *Betula* were used as outgroups (Bousquet *et al.* 1992; Chen *et al.* 1999).

DNA extraction, amplification and sequencing

Total DNAs were isolated from silica gel-dried leaves, bud material or herbarium specimens (Appendix S1) following the protocol of Bousquet *et al.* (1990). Amplification of DNA regions was performed using PCR. Primer sequences for amplification and sequencing were presented in Appendix S2. PCR cycling conditions that used by Kress *et al.* (2005) and Sass *et al.* (2007). PCR products were sequenced directly using BigDye Terminator Cycle Sequencing Ready Reaction Kit and an ABI 3730 DNA Sequencer (Applied Biosystems). The sequences were first aligned using ClustalX (Thompson *et al.* 1997) software and then manually adjusted in BioEdit v.7 (Hall 1999). GenBank Accession nos of newly determined sequence are FJ825380–FJ825433, FJ844483–FJ844605 and GU112746–GU112750 (Appendix S1).

Data analyses

Pairwise K2P (Kimura 2-parameter) distances for all four DNA regions were calculated in MEGA 3.1 (Kumar *et al.* 2004) to evaluate intraspecific and interspecific divergence in *Alnus*. Indels were coded with the simple indel coding method of Simmons & Ochoterena (2000). Three tree-based methods were used to exhibit the molecular identification results and test the monophyly of species. Neighbour joining (NJ) and maximum parsimony (MP), maximum likelihood (ML) and Bayesian inference (BI) were performed in PAUP v.4.0b10 (Swofford 2002), PhyML v. 2.4.4 (Guindon & Gascuel 2003) and MrBayes v. 3.1.2 (Huelsenbeck & Ronquist 2001) respectively. Additionally, the sequence character-based method (Rach *et al.* 2008) was used with DnaSP (Rozas *et al.* 2003), and the information from each site was treated as a character to distinguish the taxa from each other.

Results

The evaluation of DNA markers

We obtained 24 *rbcL* sequences from 20 different alder species, 21 *matK* from 16 species, 90 ITS from 26 species and 70 *trnH-psbA* from 22 species. The total number of new sequences generated in this study was 173 (Appendix S1). With regard to universality of primer and success of sequence amplification, the proportion at each of the four regions was more than 95% (Table 1). The *rbcL* matrix had 1357 bp and no indels; the distribution of seven informative sites and 19 variable sites was dispersive and sparse across the matrix (after alignment using ClustalX and adjustment in BioEdit). For *matK* matrix, aligned sequence length was 679 bp; the distribution of 14 informative sites and 46 variable sites was dispersive and sparse across the matrix, without included indels. For the ITS matrix, aligned sequence length was 529 bp; the distribution of 37 informative sites and 51 variable sites was intensive and dense across the matrix, and there were three indels 1–10 bp long. For *trnH-psbA* matrix, aligned sequence length was 450 bp; the distribution of 28 informative sites and 45 variable sites was intensive and dense across the matrix, and there were seven indels

Table 1 The evaluation of four DNA markers

DNA region	*rbcL*	*matK*	ITS	*trnH-psbA*
Universal ability to primer	Yes	Yes	Yes	Yes
Percentage PCR success	100	100	100	100
Percentage sequencing success	100	95	95	100
Aligned sequence length (bp)	1357	679	529	450
Indels length (bp)	0	0	3 (1–10)	7 (1–58)
No. information sites/variable sites	7/19	14/46	37/51	28/45
Distribution of variable sites	Di & S	Di & S	I & D	I & D
No. samples species (individuals)	20 (24)	16 (21)	26 (90)	22 (70)
Interspecific distance mean (range), %	0.18 (0–0.5)	0.93 (0–1.95)	1.5 (0–5.9)	2.1 (0–6.79)
Intraspecific distance mean (range), %	—	—	0.009 (0–0.4)	0.68 (0–2.15)
Ability to discriminate	2/20	5/16	20/26	14/22
%	10	31.25	76.9	63.6

Di, dispersive; S, sparse; I, intensive; D, dense.

1–58 bp long. The distribution of congeneric species distance from three markers is shown in Fig. 1. The mean sequence divergences in *Alnus* were 0.18% (*rbcL*), 0.93% (*matK*) and 1.5% (ITS) respectively. The distribution of interspecific and intraspecific distance is shown in Fig. 2. For ITS, the mean value of genetic distance between species was more than 10 times the intraspecific distance (0.009%). The *trnH-psbA* region generated higher mean values of genetic distance between and within species (2.1% and 0.68% respectively). The discriminating power of the four markers at the species level was 10% (*rbcL*), 31.25% (*matK*), 63.6% (*trnH-psbA*) and 76.9% (ITS). Therefore, the two-locus combination of *matK* and *rbcL* suggested by the consortium for the barcode of life (Hollingsworth *et al.* 2009) is insufficient to discriminate the genus *Alnus* at the species level because of their lower discriminating power. By contrast, combination of ITS and *trnH-psbA* can discriminate alder species in the world efficiently and should be considered as a useful supplementary barcode.

ITS data

In the ITS data set, 13 species have unique character states that differentiates them from other species of *Alnus*, and 13 monophyletic groups with higher support values are obtained (Table 2; Fig. 3). For example, *Alnus firma* Sieb. & Zucc. and the *Alnus acuminata* group (*A. acuminata* H. B. K. and *Alnus jorullensis* H. B. K.) each have two unique diagnostic sites (Position 139: C or 576: T could act as the diagnostic site for *A. firma*; position 436: C or 438: G for *A. acuminata* group). Other species with unique character states included *Alnus viridis* (Villar) DC. (position 192: T), *Alnus japonica* (only positions 445: C and 479: G could distinguish it) and *Alnus incana* ssp. *hirsuta* (Spach) A. Löve & D. Löve (positions 135: G and 140: C). And there are four species that share character states such as *Alnus cremastogyne* Burkill and *Alnus ferdinandi-coburgii* C. K. Schneid. (position 502: C), *Alnus oblongifolia* Torrey and *Alnus rhombifolia* Nutt. (position 514: T).

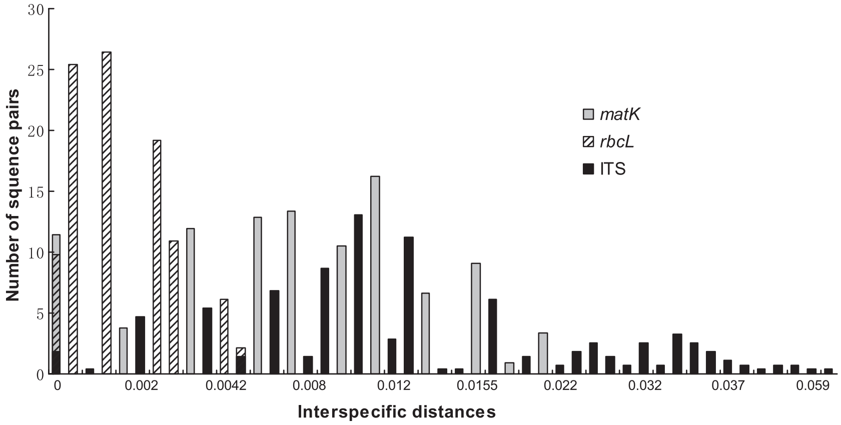


Fig. 1 Relative distribution of interspecific distances between congeneric species from three DNA regions.

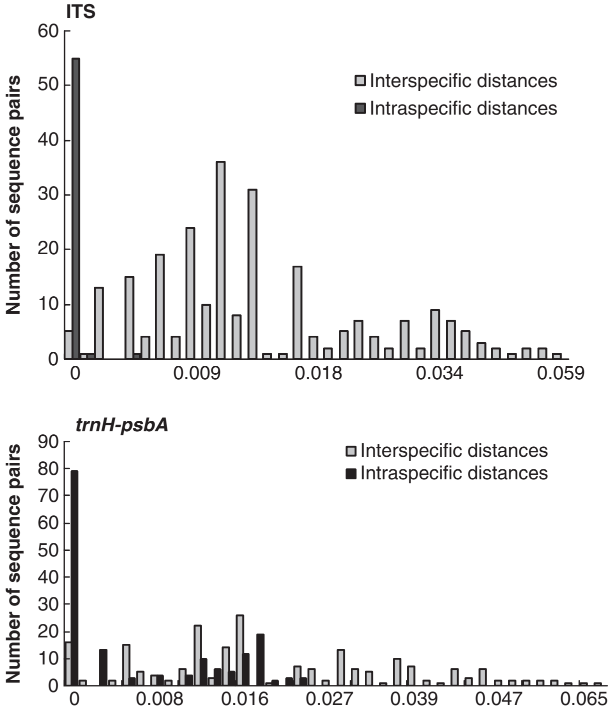


Fig. 2 Relative distribution of interspecific and intraspecific distances from ITS and *trnH-psbA* respectively.

TrnH-psbA data

Seven species of *Alnus* have unique *trnH-psbA* character states, and 10 monophyletic groups with higher support values are obtained (Table 3; Fig. 4). Both *Alnus orientalis* Decne. and *Alnus pendula* Matsum. have *trnH-psbA* sequences with three different diagnostic sites (including indel position 128: -, 244: C or 439: A for *A. orientalis* Decne.; three indel positions, 411, 431, 439 respectively, for *A. pendula*). *Alnus cordata* (Lois.) Duby displays the unique character state with C in site 165 and A in site 411, and *Alnus nepalensis* D. Don displays T in position 196. The combination of G in site 132 and A in position 444 differentiated *Alnus serrulata* Willd. from other species of *Alnus*. There are two pairs of taxa that share a single unique character state: *A. cremastogyne* and *A. ferdinandi-coburgii* (position 431: T), *Alnus glutinosa* and *A. incana* ssp. *incana* (26 bp long indel beginning from the position 165). *Alnus incana* ssp. *hirsuta* and *A. japonica* are divided into two groups, only one of which is resolved by higher monophyletic support values.

ITS and trnH-psbA combined

The result based on combined DNA regions (ITS and *trnH-psbA*) and two methods (tree-based and

Table 2 Character-based DNA database for *Alnus* species from ITS region. Character states (nucleotides) at 22 selected positions (ranging from position 126–650) are shown; abbreviations of taxa are according to Appendix S3; the number of individuals analysed per species is given in brackets. Taxa with bold style have unique DNA character state by specific single diagnostic site; taxa with italic style share specific DNA character state for each other; the rest taxa have their unique DNA character state by combining more than two sites. The grey cells show important diagnostic character sites; '—' means the indel site

	Position: 126–650																				
Taxa (*n*)	132	135	139	140	177	192	209	255	432	436	438	445	464	471	479	502	514	533	551	576	597	616
Afi (7)	G	A	C	C	T	G	C	A	C	—	A	C	T	A	—	T	C	T	A	C	T	T
Aac/Ajo (2/1)	G	A	T	C	C	G	C	G	C	C	G	C	T	A	—	T	C	T	A	T	C	C
Av (8)	G	A	T	C	T	T	C	G	C	—	A	C	T	A	—	T	C	T	A	T	T	T
Ani (3)	G	A	T	C	C	G	C	G	T	—	A	C	C	A	—	T	C	T	A	T	T	T
Amar (2)	G	A	T	C	C	G	C	G	C	—	A	C	C	A	—	T	C	T	G	T	T	T
Ased (2)	G	A	T	C	C	G	C	G	C	—	A	T	C	G	G	T	C	T	A	T	T	T
Afa (2)	G	A	T	C	C	G	C	G	C	—	A	C	C	—	—	T	C	T	A	T	T	T
Amat (2)	G	G	T	T	T	G	C	G	C	—	A	C	T	A	—	T	C	T	A	T	T	T
Aru (3)	G	A	T	C	T	G	T	G	C	—	A	C	T	A	—	T	C	T	A	T	T	T
Aisi (3)	A	A	T	C	T	G	C	G	C	—	A	C	T	A	—	T	C	T	A	T	C	T
Aino (2)	G	A	T	C	T	G	C	G	C	—	A	C	T	A	—	T	C	T	A	T	G	T
Ane (4)	G	A	T	C	C	G	C	G	C	—	A	C	T	A	—	T	A	T	A	T	T	T
Afo (4)	G	A	T	C	C	G	C	G	C	—	A	C	T	A	—	T	C	C	A	T	T	T
Aor (2)	G	A	T	C	C	G	C	G	C	—	A	C	T	A	—	T	C	T	A	T	T	A
Acr/Afe (3/2)	G	A	T	C	C	G	C	G	C	—	A	C	T	A	—	C	C	T	A	T	T	T
Aob/Arh (1/1)	G	A	T	C	C	G	C	G	C	—	A	C	T	A	—	T	T	T	A	T	C	C
Ap (2)	G	A	T	C	T	G	C	A	C	—	A	C	T	G	—	T	C	T	A	T	T	T
Aj (5)	G	A	T	C	C	G	C	G	C	—	A	C	C	G	G	T	C	T	A	T	T	T
Aish (7)	G	G	T	C	T	G	C	G	C	—	A	C	T	A	—	T	C	T	A	T	C	T
Ase (2)	G	A	T	C	C	G	C	G	C	—	A	C	C	G	—	T	C	T	A	T	T	T
Ag (4)	G	A	T	C	C	G	C	G	C	—	A	C	T	A	—	T	C	T	A	T	C	T

character-based) is shown in Fig. 5. In total, 23 species could be identified. There were 13 taxa distinguished by using either ITS or *trnH-psbA* data, including the *A. cremastogyne* group, *A. nepalensis*, *Alnus nitida* (Spach) Endl., etc. Five taxa could be identified only by ITS data (except for *A. oblongifolia* group and *A. acuminata* group for which no data from chloroplast genome were available) including *Alnus matsumurae* Callier and *Alnus inokumai* Murai & Kusaka, and three taxa were discriminated only by *trnH-psbA* data, namely *A. incana* ssp. *tenuifolia* (Nutt.) Breitung, *A. incana* ssp. *rugosa* (DuRoi) Clausen and *A. cordata*. Two species, *Alnus subcordata* C. A. Meyer and *Alnus trabeculosa*, could be discriminated only when combining the two DNA regions from different genomes.

Discussion

Several DNA barcoding markers have been used in woody and herbaceous plant taxa with different levels of taxon sampling and various identification success rates (Edwards *et al.* 2008; Lahaye *et al.* 2008; Logacheva *et al.* 2008; Newmaster *et al.* 2008; Nitta 2008; Newmaster & Ragupathy 2009; Starr *et al.* 2009), whereas standard barcoding protocols have been pursued for land plants (Chase *et al.* 2007). A successful barcoding project requires comprehensive species sampling and should facilitate high rates of distinguishing species. The barcoding database for *Alnus* represents such a project. Our data sets include all 26 species in the world and the combination of ITS and *trnH-psbA* produces a high rate of correct identification. The mean value of the genetic distance for ITS and *trnH-psbA* is markedly higher between than within species (Table 1); and they show a higher resolving power based on their sequence matrix analyses than the results from the *rbcL* and *matK* matrices. Although ITS has sometimes been treated as an unsuitable marker because of the possible impact of incomplete concerted evolution (Alvarez & Wendel 2003), our results indicate that in *Alnus* the problem may not play an important role. By contrast, the ITS region is very useful in our study because of its shortness and few indels, allowing relatively easy alignment and reliable discrimination. The ITS data differentiate 76.9% (20/26) of the species within *Alnus*.

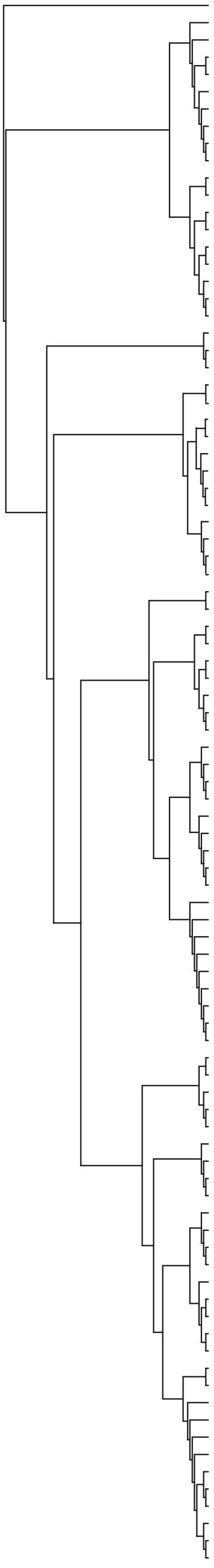
Studies on licorice (Kondo *et al.* 2007), *Compsoneura* Warb. (Newmaster *et al.* 2008), orchid (Lahaye *et al.* 2008) and filmy ferns (Nitta 2008), have shown that *trnH-psbA* may be a promising marker for DNA barcoding. With the inclusion of indel information, five species of *Alnus* have unique diagnostic DNA character states. For instance, the 47-bp-long indel is unique for *Alnus pendula*; and the 26-bp-long indel is shared by *Alnus incana* ssp. *incana* and *Alnus glutinosa*. Therefore, *trnH-psbA* is also an informative molecular marker for differentiating *Alnus* species.

Sequences of ITS and *trnH-psbA* can complement each other and the combination of them can improve the ability to discriminate at the species level (Fig. 5). For example, *Alnus trabeculosa* and *Alnus cordata* share one DNA character state in the ITS sequence matrix, but in the chloroplast gene *trnH-psbA* sequence matrix, *A. cordata* has its unique DNA character state, which offsets the deficiency from only ITS data. Conversely, *A. incana* ssp. *incana* and *A. glutinosa* share the same character state in the *trnH-psbA* data (Fig. 4), but differ from each other in the ITS sequences (T and C in the position 177 respectively; see Table 2). *Alnus japonica* and *A. incana* ssp. *hirsuta* are divided into two groups respectively, because of higher intraspecific divergence in the *trnH-psbA* matrix, which is unfortunate for DNA barcoding (Fig. 4). Fortunately, this puzzle is overcome with a fixed diagnostic state and consistent morphological characters in the ITS matrix (Fig. 3).

In addition, for the tree-based method, the disagreement in topology between trees generated with ITS data and trees generated with *trnH-psbA* data offers information that can be used to distinguish *Alnus* taxa. *Alnus incana* is divided into four different subspecies, *A. incana* ssp. *incana*, *A. incana* ssp. *hirsuta*, *A. incana* ssp. *tenuifolia* and *A. incana* ssp. *rugosa*, according to the morphological character and geographical distribution information. They are not distinguished from each other in the ITS matrix because of lower sequence divergence, except for *A. incana* ssp. *incana*. On the contrary, *trnH-psbA* data reflect the distribution relationship of alder species to some extent. The taxa distributed in North America are differentiated due to their unique character state and specific location on topology, such as *A. incana* ssp. *tenuifolia* and *A. incana* ssp. *rugosa*. The same condition has also occurred for *A. trabeculosa* and *Alnus subcordata*.

Having both nuclear and chloroplast DNA markers may be advantageous in discerning hybrid species due to their different pattern of inheritance. Within *Alnus*, *Alnus mayrii* Callier has been known as a hybrid species between *A. japonica* and *A. incana* ssp. *hirsuta* (Spach) A. Löve & D. Löve (Murai 1964). It is grouped with

Fig. 3 Neighbour-joining tree based on the ITS sequence matrix for 26 alder species; every individual is shown with the order of GenBank Accession no., DNA number and the name before and after taxonomic revision. The rest of the columns are character-based diagnostic site information and support values (bootstrap or Bayesian posterior probabilities, in percentage) with different tree-based methods respectively. The frames with shading indicate some error during sampling or labelling, and the condition of shared DNA character states is shown.



Access No.	DNA No.	Name before	Name after	Diagnostic site	Values (MP/BI/ML)
AB343907		*B. apoiensis*	*B apoiensis*		
AY352315	1910	*A. mandschurica*	*A. viridis*	**192 : T**	
AY352325	1250	*A. sinuata*			
AB243877		*A. maximowiczii*			
AY352309	1784	*A. fruticosa*			
AJ251608		*A. sinuata*			
AY352316		*A. maximowiczii*			
AJ251681		*A. crispa*			
AY352329	660	*A. viridis*			
FJ825380	F13-4	*A. viridis*			
AY352317	416	*A. pendula*	*A. pendula*	**471 : G**	**80/100/99**
AJ251682		*A. pendula*			
FJ825385	F131-2	*A. firma*	*A. firma*	**139 : C**	**75/100/72**
FJ825386	F131-4	*A. firma*			
FJ825384	F127	*A. sieboldiana*			
GU112746	1787	*A. sieboldiana*			
FJ825383	F131-3	*A. firma*			
FJ825381	F131-1	*A. firma*			
FJ825382	F131-6	*A. firma*			
AJ251167		*A. nitida*	**?**		
AJ251678		*A. formosana*			
AJ251679		*A. maritima*			
FJ825428	F15	*A. serrulatoides*	*A. serrulatoides*	**445 : T**	**69/99/91**
GU112749	F15-1	*A. serrulatoides*			
FJ825433	F16	*A. fauriei*	*A. fauriei*	**471 : -**	**62/100/94**
GU112750	F16-1	*A. fauriei*			
FJ825432	F126-1	*A. japonica*	*A. japonica*	**479 : G**	
FJ825431	F135	*A. mayrii*			
FJ825429	F125-2	*A. japonica*			
FJ825430	F125-9	*A. japonica*			
FJ825426	F125-1	*A. japonica*			
FJ825425	F125-13	*A. japonica*			
FJ825427	F137	*A. serrulata*	*A. serrulata*		
AY352322	1789	*A. serrulata*			
FJ825423	1249	*A. maritima*	*A. maritima*	**551 : G**	**95/100/98**
FJ825424	1249-2	*A. maritima*			
FJ825422	F121	*A. subcordata*	*A. subcordata*		
AJ251664		*A. subcordata*			
AY352320		*A. orientalis*	*A. orientalis*	**661 : A**	**86/100/100**
FJ825421	F120	*A. orientalis*			
AJ783638		*A. nitida*	*A. nitida*	**432 : T**	**87/100/98**
FJ825419	F19-4	*A. nitida*			
FJ825420	F19-1	*A. nitida*			
AY352318		*A. nepalensis*	*A. nepalensis*	**514 : A**	**58/99/61**
FJ011767		*A. nepalensis*			
FJ825418	688	*A. nepalensis*			
AJ251677		*A. nepalensis*			
FJ825417	F128-3	*A. ferdinandi-coburgii*	*A. ferdinandi-coburgii*	**502 : C**	**65/94/70**
FJ825416	F128-2	*A. ferdinandi-coburgii*			
FJ825415	337	*A. cremastogyne*	*A. cremastogyne*		
FJ825413	F129-7	*A. cremastogyne*			
FJ825414	690	*A. cremastogyne*			
AJ251663		*A. cordata*	*A. cordata*		
FJ825409	F11-3	*A. cordata*			
FJ825410	F11-1	*A. cordata*			
FJ825408	F126-7	*A. trabeculosa*	*A. trabeculosa*		
FJ825412	F126-3	*A. trabeculosa*			
FJ825407	1252	*A. formosana*	*A. formosana*	**533 : C**	**64/96/60**
FJ825406	F133	*A. formosana*			
FJ825405	F133-2	*A. formosana*			
GU112748	F18	*A. henryi*			
AY352319		*A. oblongifolia*	*A. oblongifolia*	**436 : C**	**82/96/98**
AJ251669		*A. rhombifolia*	*A. rhombifolia*		
AJ251673		*A. acuminata*	*A. acuminata*	**516 : T**	**89/100/99**
AF432066		*A. acuminata*			
AJ251672		*A. jorullensis*	*A. jorullensis*		
FJ825404	94	*A. sibirica*	*A. glutinosa*		
FJ825403	144	*A. japonica*			
FJ825399	F17-1	*A. glutinosa*			
AJ251662		*A. glutinosa*			
FJ825393	F122-2	*A. hirsuta*	*A. incana* ssp. *hirsuta*	**135 : G**	
FJ825392	F122-1	*A. hirsuta*			
FJ825390	F123	*A. hirsuta*			
FJ825391	F122-3	*A. hirsuta*			
FJ825394	F122-10	*A. hirsuta*			
FJ825395	F122-12	*A. hirsuta*			
FJ825396	F122-6	*A. hirsuta*			
FJ825397	F134	*A. matsumurae*	*A. matsumurae*	**140 : T**	**84/100/93**
GU112747	F134-1	*A. matsumurae*			
FJ825398	F132-2	*A. inokumai*	*A. inokumai*	**597 : G**	**61/99/72**
AJ251671		*A. inokumai*			
AY352327	F136	*A. tenuifolia*	*A. incana* ssp. *tenuifolia*		
AJ251666		*A. tenuifolia*			
AY352313		*A. rugosa*	*A. incana* ssp. *rugosa*		
FJ825389	F122-13	*A. hirsuta*			
AY352312		*A. incana*	*A. incana* ssp. *incana*	**132: A**	**57/96/59**
FJ825401	F14-3	*A. incana*			
AJ251665		*A. incana*			
FJ825402	F12-3	*A. rubra*	*A. rubra*	**209: T**	**64/95/62**
AJ251668		*A. rubra*			
AY352321		*A. rubra*			

Table 3 Character-based DNA database for *Alnus* species from *trnH-psbA* region. Character states (nucleotides) at 16 selected positions (ranging from position 95–485) are shown; taxa abbreviations are according to Appendix S3; numbers of individuals analysed per species are given in brackets. Taxa with bold style have unique DNA character state by specific single diagnostic site; taxa with italic style share specific DNA character state for each other; the rest taxa have their unique DNA character state by combining more than two sites. The grey cells show important diagnostic character sites; '—' means the indel site

	Position: 95–485															
Taxa (*n*)	95	128	131	132	139	165	196	213	244	357	378	411	431	439	444	461
Ap (4)	G	T	A	T	—	T	C	G	A	A	G	—	—	—	G	A
Aor (2)	G	—	—	—	—	T	C	G	C	A	T	C	G	A	G	G
Ac (3)	G	T	A	T	T	C	C	G	A	A	T	A	G	G	G	A
Aj (4)	G	T	A	T	—	T	C	G	A	—	T	C	G	G	G	A
Ani (3)	T	T	A	T	C	T	C	G	A	A	T	C	G	G	G	A
Ane (3)	G	T	A	T	—	T	T	G	A	A	T	C	G	G	G	A
Aru (2)	G	T	A	—	—	T	C	C	A	A	T	C	G	G	G	A
Ag/Aisi (3/3)	G	T	A	T	T	—	C	G	A	A	T	C	G	G	T	A
Acr/Afe (5/6)	G	T	T	T	C	T	C	G	A	A	T	C	T	G	G	A
Ase (2)	G	A	A	G	T	T	C	G	A	A	A	C	G	G	A	A
Amar (1)	G	A	A	G	T	T	C	G	A	A	T	C	G	G	G	A

A. japonica in the ITS matrix (Fig. 3), but with *A. incana* ssp. *hirsuta* in the *trnH-psbA* database (Fig. 4).

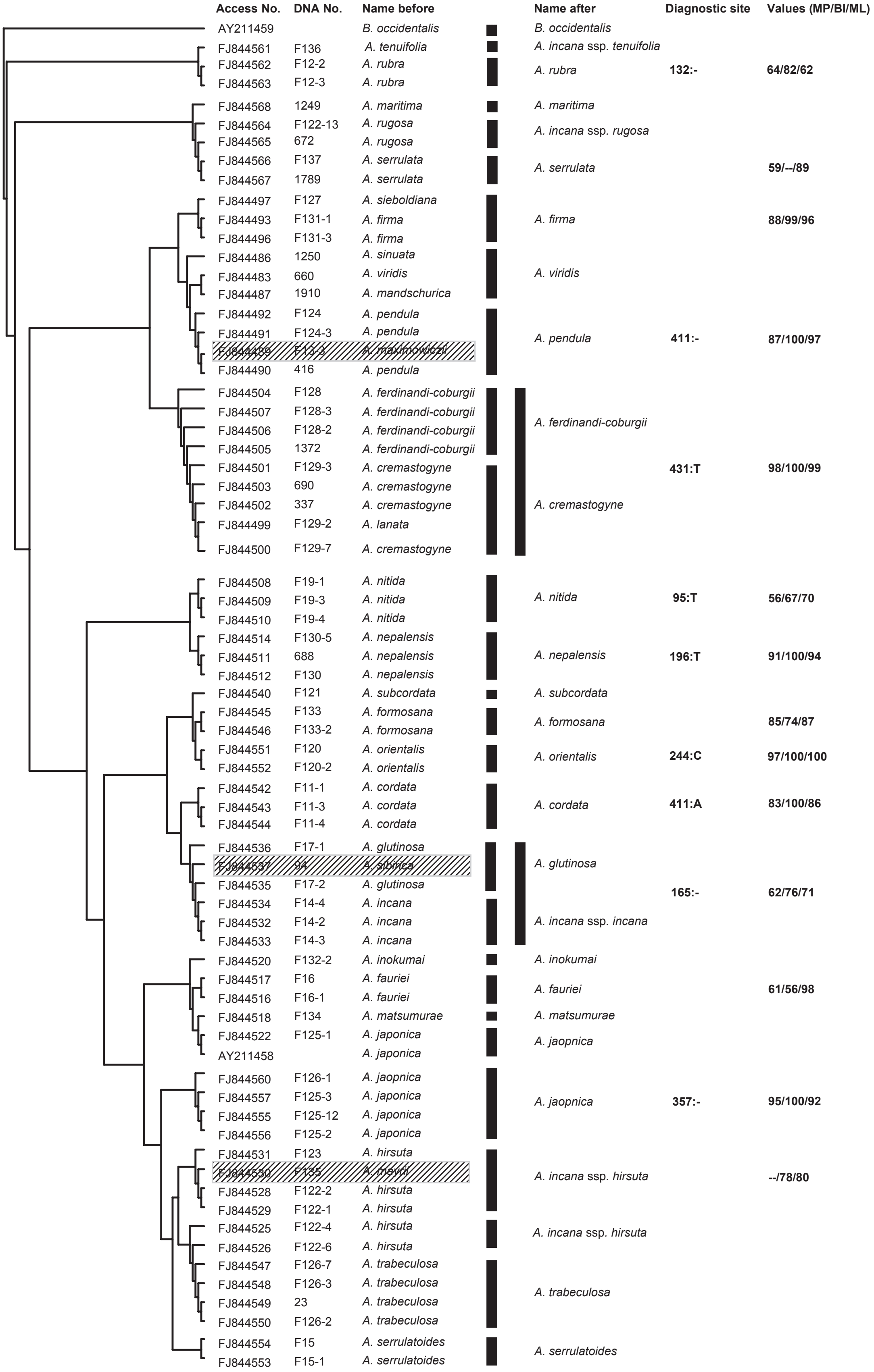
This study has identified unique DNA character combinations for most (23/26) species of alder (Fig. 5). Two species, *A. subcordata* and *A. trabeculosa*, could be discriminated only when combining the two ITS and *trnH-psbA* regions. Two species pairs (*Alnus acuminanta* and *Alnus jorullensis*; *Alnus oblongifolia* and *Alnus rhombifolia*) in Central and South America and one species complex (*Alnus viridis*) need further study. This should be carried out by sampling more individuals from different populations throughout their areas of distribution.

Molecular data can complement morphologically based taxonomy

Our results in *Alnus* show that species with distinctive morphology have specific DNA character states. This can be seen in *Alnus nepalensis* with unique diagnostic female inflorescences and obvious stipules, in *Alnus rubra* with regular sawtooth and leaf shape. This demonstrates consistency between molecular data and morphology. However, it is hazardous to have phylogenetic analysis and DNA identification database without the foundation of taxonomic revision (Kristiansen *et al.* 2005; Newmaster *et al.* 2008). Our study further confirmed this situation by raising some questions as a result of previous studies. Navarro *et al.* (2003) first recovered a clade using ITS data that consisted of three species that flower in autumn (*Alnus nitida*, *Alnus formosana* and *Alnus maritima* Muhl. ex Nutt.), an unusual condition in *Alnus*. This clade was consistent with the previously described subgenus *Clethropsis* (Furlow 1979). Later, phylogenetic analysis by Chen & Li (2004) used the same ITS sequences of these three species and arrived at similar conclusion as above. In this study, we sampled more than two individuals for each of the three species and found that they were not monophyletic but scattered in different clades (Fig. 3). The results of previous studies may have been the result of contamination or misidentification, and the voucher of materials that was used to extract total DNA by Navarro *et al.* (2003) should be examined. This sort of problem was avoided in this study through sampling more than two individuals for each species, such that when constructing a DNA identification database, the accuracy of each sequence was verified against other conspecifics and the range of variation within a species was included as much as possible. Additionally, we found that within species, certain positions displayed two or three different character states, further emphasizing the need for extensive sampling at the population level.

As shown in Fig. 3, the incorrectly labelled samples were found in our previous total DNA bank after sequencing and alignment. *Alnus japonica* (144) and *Alnus*

Fig. 4 Neighbour-joining tree based on the *trnH-psbA* sequences matrix for 22 alder species. Every individual is shown with the order of GenBank Accession no., DNA number and the name before and after taxonomic revision. The rest columns are character-based diagnostic site information and support values (bootstrap or Bayesian posterior probabilities, in percentage) with three different tree-based methods respectively. The frames with shading indicate some error during sampling or labelling, and the condition of sharing DNA character state is shown.


Access No. DNA No. Name before Name after Diagnostic site Values (MP/BI/ML)
AY211459 B. occidentalis B. occidentalis
FJ844561 F136 A. tenuifolia A. incana ssp. tenuifolia
FJ844562 F12-2 A. rubra A. rubra 132:- 64/82/62
FJ844563 F12-3 A. rubra
FJ844568 1249 A. maritima A. maritima
FJ844564 F122-13 A. rugosa A. incana ssp. rugosa
FJ844565 672 A. rugosa
FJ844566 F137 A. serrulata A. serrulata 59/--/89
FJ844567 1789 A. serrulata
FJ844497 F127 A. sieboldiana
FJ844493 F131-1 A. firma A. firma 88/99/96
FJ844496 F131-3 A. firma
FJ844486 1250 A. sinuata
FJ844483 660 A. viridis A. viridis
FJ844487 1910 A. mandschurica
FJ844492 F124 A. pendula
FJ844491 F124-3 A. pendula
FJ844489 F13-3 A. maximowiczii A. pendula 411:- 87/100/97
FJ844490 416 A. pendula
FJ844504 F128 A. ferdinandi-coburgii
FJ844507 F128-3 A. ferdinandi-coburgii A. ferdinandi-coburgii
FJ844506 F128-2 A. ferdinandi-coburgii
FJ844505 1372 A. ferdinandi-coburgii
FJ844501 F129-3 A. cremastogyne 431:T 98/100/99
FJ844503 690 A. cremastogyne
FJ844502 337 A. cremastogyne A. cremastogyne
FJ844499 F129-2 A. lanata
FJ844500 F129-7 A. cremastogyne
FJ844508 F19-1 A. nitida
FJ844509 F19-3 A. nitida A. nitida 95:T 56/67/70
FJ844510 F19-4 A. nitida
FJ844514 F130-5 A. nepalensis
FJ844511 688 A. nepalensis A. nepalensis 196:T 91/100/94
FJ844512 F130 A. nepalensis
FJ844540 F121 A. subcordata A. subcordata
FJ844545 F133 A. formosana A. formosana 85/74/87
FJ844546 F133-2 A. formosana
FJ844551 F120 A. orientalis A. orientalis 244:C 97/100/100
FJ844552 F120-2 A. orientalis
FJ844542 F11-1 A. cordata
FJ844543 F11-3 A. cordata A. cordata 411:A 83/100/86
FJ844544 F11-4 A. cordata
FJ844536 F17-1 A. glutinosa
FJ844537 94 A. sibirica A. glutinosa
FJ844535 F17-2 A. glutinosa 165:- 62/76/71
FJ844534 F14-4 A. incana
FJ844532 F14-2 A. incana A. incana ssp. incana
FJ844533 F14-3 A. incana
FJ844520 F132-2 A. inokumai A. inokumai
FJ844517 F16 A. fauriei A. fauriei 61/56/98
FJ844516 F16-1 A. fauriei
FJ844518 F134 A. matsumurae A. matsumurae
FJ844522 F125-1 A. japonica A. jaopnica
AY211458 A. japonica
FJ844560 F126-1 A. jaopnica
FJ844557 F125-3 A. japonica A. jaopnica 357:- 95/100/92
FJ844555 F125-12 A. japonica
FJ844556 F125-2 A. japonica
FJ844531 F123 A. hirsuta
FJ844530 F135 A. mayrii A. incana ssp. hirsuta --/78/80
FJ844528 F122-2 A. hirsuta
FJ844529 F122-1 A. hirsuta
FJ844525 F122-4 A. hirsuta A. incana ssp. hirsuta
FJ844526 F122-6 A. hirsuta
FJ844547 F126-7 A. trabeculosa
FJ844548 F126-3 A. trabeculosa
FJ844549 23 A. trabeculosa A. trabeculosa
FJ844550 F126-2 A. trabeculosa
FJ844554 F15 A. serrulatoides A. serrulatoides
FJ844553 F15-1 A. serrulatoides

sibirica (94) were in fact samples of *A. glutinosa*. *Alnus pendula* should be the correct name of seed sample labelled *A. viridis* (*maximowiczii*) (F13-3) in Fig. 4; these seed samples from Japan were confused because of similar morphological characteristics of the two species. Our study further confirmed that building up a credible DNA identification database indeed required proper sample collection, and it was important to correct mistakes that accumulate during sampling and experimentation.

Three species (*A. japonica*, *A. formosana* and *A. maritima*) are considered difficult to distinguish from each other using morphological characters alone. This taxonomic puzzle is resolved with the addition of DNA sequence data that offers unique character state at the species level. This solution also applies to species pairs such as *A. incana* ssp. *incana* and *A. glutinosa*, and *A. japonica* and *A. trabeculosa*. Therefore, DNA barcoding can complement and reinforce classical morphologically based taxonomy to some extent. On the contrary, *Alnus cremastogyne* and *Alnus ferdinandi-coburgii* shared one DNA character state, which differed from the result of classical taxonomy in establishing the species. Further study should be carried out to understand this phenomenon to construct more reliable taxonomic system. In addition, fewer species or groups were confirmed by tree-based monophyly testing than by character-based method (Fig. 5); this indicated that insufficient information for resolving phylogenetic relationships was sometimes enough to be used to distinguish the alder species.

In Fig. 6, we use a double helix to show the relationship between the traditional taxonomy and modern

Species	ITS		*trnH-psbA*
A. cremastogyne *A. ferdinandi-coburgii*	■ ●		■ ●
A. nepalensis	■ ●		■ ●
A. nitida	■ ●		■ ●
A. orientalis	■ ●		■ ●
A. firma	■ ●		□ ●
A. rubra	■ ●		■ ●
A. pendula	□ ●		■ ●
A. formosana	■ ●		□ ●
A. maritima	■ ●		□
A. japonica	■		□
A. viridis	■		□
A. incana ssp. *hirsuta*	■		□
A. serrulata	□		□
A. oblongifolia *A. rhombifolia*	■ ●		
A. acuminata *A. jorullensis*	■ ●		
A. incana ssp. *incana*	■ ●		
A. inokumai	■ ●		
A. matsumurae	■		
A. serrulatoides	■		
A. fauriei	■		
A. glutinosa	□		■ ●
A. cordata			■ ●
A. incana ssp. *rugosa*			□
A. incana ssp. *tenuifolia*			□
A. subcordata		▨	
A. trabeculosa		▨	

■ Single diagnostic site □ Combining sites ▨ Combining genomes ● Support values >50%

Fig. 5 The result of combining two markers, ITS and *trnH-psbA*, from two different genomes. The species with specific character states using character-based method is shown by three different squares; black ones indicate that a single diagnostic site is used; white ones mean that combining diagnostic sites are used; shaded ones mean that unique DNA character state is obtained through combined site information from two different markers. The circle indicates the support values of each clade that are higher than 50% with different tree-based methods.

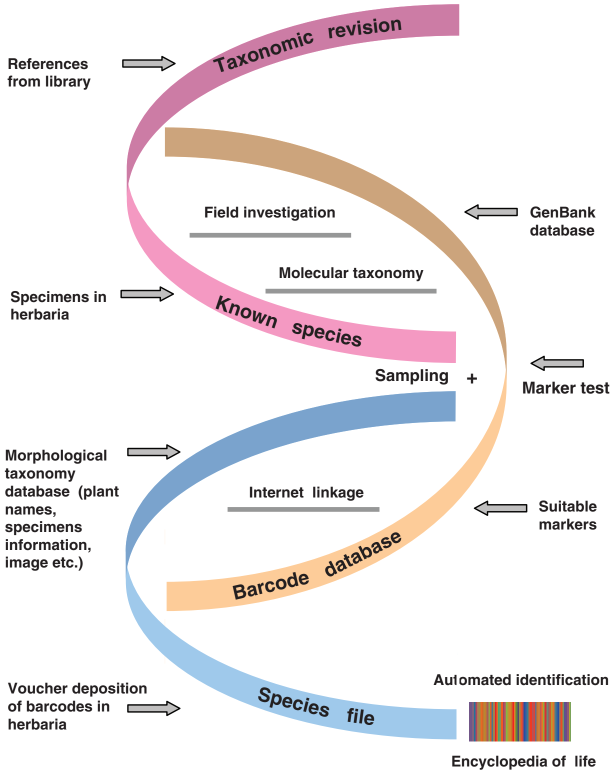


Fig. 6 A double helix to show the relationship between classical taxonomy and molecular database. It is divided into three parts with different colours. Pink denotes the process of taxonomic revision; yellow indicates the workflow of the construction of the barcode database and blue indicates the integration of information from bioinformatics platform after combining classical taxonomy and DNA barcoding data, utilizing the advantages of internet techniques and management systems from large herbaria.

molecular identification. It is divided into three parts, including the process of taxonomic revision, the workflow of the construction of the DNA barcode database and the integration of information from the bioinformatics platform.

First, the taxonomic revision establishes the primary number of species in a genus by checking references from libraries, scrutinizing specimens from herbaria, quantitative morphological analyses and field investigation. This forms the basis for recognizing discreet species within a genus, even though there are still some confusing species. Second, the DNA barcode database will be more credible and valuable when the samples are collected to cover morphological and geographical and ecological variability, and the primary number of species may be revised according to the molecular data. Then automated identification of species can be realized by combining classical taxonomy and DNA barcoding data, utilizing the advantages of internet technique and management system from large herbaria.

With the bioinformatics platform, the DNA data become the carrier of biological species information through the internet, forming a dynamic and relative reliable and open system. The linked information includes type specimen details, images, correct name, morphological description, geographical distribution map, molecular database and economical and medical use. With regard to more future collaboration all over the world, original materials like the protologue of species name that was difficult to access for taxa, scattered in different continents could be saved together as a species file. Using this management model (Fig. 6), such file will enhance the power of herbaria and enable them to offer more complete information conveniently to the public.

However, the appearance of cryptic species or species sharing the same DNA character state is the evidence of the conflicts between morphological and molecular taxonomy (Lahaye *et al*. 2008; Newmaster & Ragupathy 2009). Under this condition, fully confident decisions will only be possible after making further taxonomic revision based on multiple data of the taxa, such as ecological, morphological and additional genetic data (Savolainen *et al*. 2005; Haase *et al*. 2007).

Conclusion

Sequences of nuclear ITS and chloroplast *trnH-psbA* can successfully differentiate 23 out of 26 *Alnus* species in the world. The DNA barcoding protocol lays a foundation for ecological and biological studies of *Alnus*, an important tree genus in temperate forests of the Northern Hemisphere.

The development of rapid and accurate species identification tools is a growing field in biology today and will be important in the future. Combining DNA sequences with existing morphological characters allows DNA regions to become a novel internet biological information carrier, and a relatively reliable and open taxonomic system will be completed and amended by adding related information and constant expert supervision. Building up a comprehensive and accurate integrated information database (encyclopaedia of life, linking all kinds of information) should be the goal pursued by modern taxonomist.

Acknowledgements

The authors thank De-Yuan Hong, Jian-Hua Fan, Jian-Hua Li, Jun Liu, Shinsuke Sato and the following institutions: Arnold Arboretum, Harvard University, Royal Botanic Gardens, Kew, Royal Botanic Gardens, Edinburgh, Leiden Garden and University of Zurich, Switzerland for their help in the field work or providing plant tissue for this study; Wei Wang and Rui-Qi Li for their laboratory assistance. Special thanks to David Dilcher (University of Florida), Nathan Jud (University of Maryland) and Jian-Hua Li (The Arnold Arboretum, Harvard University) for their careful revision of an early version of the manuscript. This research work was supported by the Research Fund from the Chinese Academy of Sciences (KSCX2-YW-R-136), the Large-scale Scientific Facilities of the Chinese Academy of Sciences (2009-LSF-GBOWS-01), National Basic Research Program of China (973 Program no. 2007CBI411600) and National Natural Science Foundation of China (30740088 and 30370094).

References

Alvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution*, **29**, 417–434.

Benson DR, Silvester WB (1993) Biology of *Frankia* strains, actinomycete symbionts of actinorhizal plants. *Microbiological Reviews*, **57**, 293–319.

Bousquet J, Simon L, Lalonde M (1990) DNA amplification from vegetative and sexual tissues of trees using polymerase chain-reaction. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, **20**, 254–257.

Bousquet J, Strauss SH, Li P (1992) Complete congruence between morphological and *rbcL*-based molecular phylogenies in birches and related species (Betulaceae). *Molecular Biology and Evolution*, **9**, 1076–1088.

Chase MW, Salamin N, Wilkinson M *et al.* (2005) Land plants and DNA barcodes: short-term and long-term goals. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, **360**, 1889–1895.

Chase MW, Cowan RS, Hollingsworth PM *et al.* (2007) A proposal for a standardised protocol to barcode all land plants. *Taxon*, **56**, 295–299.

Chen ZD (1994) Phylogeny and phytogeography of Betulaceae. *Acta Phytotaxonomic Sinica*, **32**, 1–32, 101–153.

Chen ZD, Li JH (2004) Phylogenetics and biogeography of *Alnus* (Betulaceae) inferred from sequences of nuclear ribosomal

DNA its region. *International Journal of Plant Sciences*, **165**, 325–335.

Chen ZD, Manchester SR, Sun HY (1999) Phylogeny and evolution of the Betulaceae as inferred from DNA sequences, morphology, and paleobotany. *American Journal of Botany*, **86**, 1168–1181.

Cowan RS, Chase MW, Kress WJ, Savolainen V (2006) 300,000 species to identify: problems, progress, and prospects in DNA barcoding of land plants. *Taxon*, **55**, 611–616.

DeSalle R (2006) Species discovery versus species identification in DNA barcoding efforts: response to Rubinoff. *Conservation Biology*, **20**, 1545–1547.

Devey DS, Chase MW, Clarkson JJ (2009) A stuttering start to plant DNA barcoding: microsatellites present a previously overlooked problem in non-coding plastid regions. *Taxon*, **58**, 7–15.

Edwards D, Horn A, Taylor D, Savolain V, Hawkins JA (2008) DNA barcoding of a large genus, *Aspalathus* L. (Fabaceae). *Taxon*, **57**, 1317–1327.

Erickson DL, Spouge J, Resch A, Weigt LA, Kress WJ (2008) DNA barcoding in land plants: developing standards to quantify and maximize success. *Taxon*, **57**, 1304–1316.

Fazekas AJ, Burgess KS, Kesanakurti PR *et al.* (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *Public Library of Science, ONE*, **3**, e2802.

Ford CS, Ayres KL, Toomey N *et al.* (2009) Selection of candidate coding DNA barcoding regions for use on land plants. *Botanical Journal of the Linnean Society*, **159**, 1–11.

Furlow JJ (1979) The systematics of American species of *Alnus* (Betulaceae). *Rhodora*, **81**, 1–121, 151–248.

Govaerts R, Frodin DG (1998) *World Checklist and Bibliography of Fagales*. Royal Botanical Garden, Kew, pp. 17–35.

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696–704.

Haase M, Wilke T, Mildner P (2007) Identifying species of *Bythinella* (Caenogastropoda: Rissooidea): a plea for an integrative approach. *Zootaxa*, **1563**, 1–16.

Hajibabaei M, Singer GAC, Hebert PDN, Hickey DA (2007) DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics*, **23**, 167–172.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **270**, 313–321.

Hollingsworth PM, Forrest LL, Spouge JL *et al.* (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 12794–12797.

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

Kondo K, Shiba M, Yamaji H *et al.* (2007) Species identification of licorice using nrDNA and cpDNA genetic markers. *Biological & Pharmaceutical Bulletin*, **30**, 1497–1502.

Kress WJ, Erickson DL (2007) A two-locus global DNA barcode for land plant: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *Public Library of Science, ONE*, **2**, e508.

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 8369–8374.

Kristiansen KA, Cilieborg M, Drabkova L *et al.* (2005) DNA taxonomy—the riddle of Oxychloe (Juncaceae). *Systematic Botany*, **30**, 284–289.

Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics*, **5**, 150–163.

Lahaye R, Van der Bank M, Bogarin D *et al.* (2008) DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 2923–2928.

Logacheva MD, Valiejo-Roman CM, Pimenov MG (2008) ITS phylogeny of West Asian *Heracleum* species and related taxa of Umbelliferae-Tordylieae W.D.J. Koch, with notes on evolution of their *psbA-trnH* sequences. *Plant Systematics and Evolution*, **270**, 139–157.

Murai S (1964) Phytotaxonomical and geobotanical studies on genus *Alnus* in Japan III taxonomy of whole world species and distribution of each sect. *Bulletin Government Forest Experimental Station of Japan*, **171**, 1–107.

Navarro E, Bousquet J, Moiroud A *et al.* (2003) Molecular phylogeny of *Alnus* (Betulaceae), inferred from nuclear ribosomal DNA ITS sequences. *Plant and Soil*, **254**, 207–217.

Newmaster SG, Ragupathy S (2009) Testing plant barcoding in a sister species complex of pantropical *Acacia* (Mimosoideae, Fabaceae). *Molecular Ecology Resources*, **9**, 172–180.

Newmaster SG, Fazekas AJ, Ragupathy S (2006) DNA barcoding in land plants: evaluation of rbcL in a multigene tiered approach. *Canadian Journal of Botany-Revue Canadienne De Botanique*, **84**, 335–341.

Newmaster SG, Fazekas AJ, Steeves RAD, Janovec J (2008) Testing candidate plant barcode regions in the Myristicaceae. *Molecular Ecology Resources*, **8**, 480–490.

Nitta JH (2008) Exploring the utility of three plastid loci for biocoding the filmy ferns (Hymenophyllaceae) of Moorea. *Taxon*, **57**, 725–736.

Presting GG (2006) Identification of conserved regions in the plastid genome: implications for DNA barcoding and biological function. *Canadian Journal of Botany-Revue Canadienne De Botanique*, **84**, 1434–1443.

Rach J, DeSalle R, Sarkar IN, Schierwater B, Hadrys H (2008) Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **275**, 237–247.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497.

Sass C, Littlle DP, Stevenson DW, Specht CD (2007) DNA barcoding in the Cycadales: testing the potential of proposed barcoding marker for species identification of Cycads. *Public Library of Science, ONE*, **11**, e1154.

Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R (2005) Towards writing the encyclopedia of life: an introduction to DNA barcoding. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, **360**, 1805–1811.

Seberg O, Petersen G (2009) How many loci does it take to DNA barcode a Crocus? *Public Library of Science, ONE*, **4**, e4598.

Simmons MP, Ochoterena H (2000) Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology*, **49**, 369–381.

Smith MA, Fisher BL, Hebert PDN (2005) DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, **360**, 1825–1834.

Starr JR, Naczi RFC, Chouinard BN (2009) Plant DNA barcodes and species resolution in sedges (*Carex*, Cyperaceae). *Molecular Ecology Resources*, **9**, 151–163.

Swofford DL (2002) *PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods)*, version 4.0b10. Sinauer, Sunderland.

Thompson JD, Gibson TJ, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequences alignment aided by quality analysis tools. *Nucleic Acids Research*, **24**, 4876–4882.

Will KW, Mishler BD, Wheeler QD (2005) The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology*, **54**, 844–851.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Voucher information and GenBank Accession no. for samplings used in this study

Appendix S2 Primers used in this study

Appendix S3 The full and abbreviation name of predefined species on *Alnus*

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.