MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2013) **13**, 21–31 doi: 10.1111/1755-0998.12020

Efficient distinction of invasive aquatic plant species from non-invasive related species using DNA barcoding

R. GHAHRAMANZADEH,*† G. ESSELINK,* L. P. KODDE,* H. DUISTERMAAT,‡ J. L. C. H. VAN VALKENBURG,§ S. H. MARASHI,¶ M. J. M. SMULDERS* and C. C. M. VAN DE WIEL*

Wageningen UR Plant Breeding, PO Box 16, Wageningen, NL-6700 AA, The Netherlands, †College of Agriculture, University of Mohaghegh Ardabili, PO Box 179, Ardabil, Iran, ‡NCB Naturalis, Section Nationaal Herbarium Nederland, PO Box 9514, Leiden, NL-2300 RA, The Netherlands, §National Plant Protection Organization (NPPO), PO Box 9102, Wageningen, NL-6700 HC, The Netherlands, ¶College of Agriculture, Ferdowsi University of Mashhad, PO Box 91775-1163, Mashhad, Iran

Abstract

Biological invasions are regarded as threats to global biodiversity. Among invasive aliens, a number of plant species belonging to the genera *Myriophyllum*, *Ludwigia* and *Cabomba*, and to the Hydrocharitaceae family pose a particular ecological threat to water bodies. Therefore, one would try to prevent them from entering a country. However, many related species are commercially traded, and distinguishing invasive from non-invasive species based on morphology alone is often difficult for plants in a vegetative stage. In this regard, DNA barcoding could become a good alternative. In this study, 242 samples belonging to 26 species from 10 genera of aquatic plants were assessed using the chloroplast loci *trnH-psbA*, *matK* and *rbcL*. Despite testing a large number of primer sets and several PCR protocols, the *matK* locus could not be amplified or sequenced reliably and therefore was left out of the analysis. Using the other two loci, eight invasive species could be distinguished from their respective related species, a ninth one failed to produce sequences of sufficient quality. Based on the criteria of universal application, high sequence divergence and level of species discrimination, the *trnH-psbA* noncoding spacer was the best performing barcode in the aquatic plant species studied. Thus, DNA barcoding may be helpful with enforcing a ban on trade of such invasive species, such as is already in place in the Netherlands. This will become even more so once DNA barcoding would be turned into machinery routinely operable by a nonspecialist in botany and molecular genetics.

Keywords: *Cabomba*, Hydrocharitaceae, *Ludwigia*, *matK*, *Myriophyllum*, *rbcL*, *trnH-psbA*

Received 25 May 2012; revision received 8 August 2012; accepted 16 August 2012

Introduction

Invasion by exotic species represents one of the greatest threats to biodiversity worldwide and is considered a major component of global change (Mack *et al.* 2000; Mooney & Hobbs 2000). In addition to affecting ecosystems and contributing to the local extinction of native species, invasive exotic species can also cause socio-economic damage (Pimentel *et al.* 2005). The introduction of exotic species has increased dramatically in frequency and extent in recent decades (McNeely *et al.* 2001), partly as a result of increased transport and trade. Many aliens have entered new areas through commerce, either purposely, for example, as garden or aquarium plant, or by accident as stowaway or weed. Species that cause inconvenience through strong increase in occurrence are called invasive species. The inconvenience can be economic damage (e.g. noxious weeds such as *Cyperus esculentus* in agriculture, or obstruction of waterways through rampant growth of *Hydrocotyle ranunculoides*) or problems in relation to health (e.g. allergic reactions when in skin contact with *Heracleum mantegazzianum* or inhaling pollen from *Ambrosia artemisiifolia*) or damage to ecosystems (e.g. biodiversity loss along waterways by the vegetation being overgrown by *Fallopia japonica*).

During the last few decennia, the flora of the Netherlands experienced a considerable increase in exotic plant species, particularly in some groups of aquatic plants, including *Myriophyllum* spp., *Cabomba* spp., *Ludwigia* spp. and some genera from the Hydrocharitaceae family. The submerged aquatic plant genus, *Myriophyllum*, is among the most species rich of the aquatic plants, and these water milfoils have a worldwide distribution. The genus is well known for its invasive species, such as *Myriophyllum aquaticum*, *Myriophyllum heterophyllum* and *Myriophyllum spicatum* (Moody & Les

Correspondence: Clemens C.M. van de Wiel,
Fax: +31-317-483457; E-mail: clemens.vandewiel@wur.nl.

2010), the latter species, however, being a harmless native in the Netherlands. *Ludwigia peploides* and *L. grandiflora* both originate from South America and can now be found in Europe. *Ludwigia* species are morphologically very similar and are difficult to differentiate in the absence of flowers. Their populations' rapid and extensive development can block waterways, reduce biodiversity and degrade water quality (Dandelot *et al.* 2008). The non-native invasive plant species *Lagarosiphon major* (Hydrocharitaceae) is a submerged aquatic macrophyte that poses a significant threat to water bodies in Europe (Baars *et al.* 2010). *Cabomba caroliniana* is a fully submerged aquatic plant, originally a native of the Americas, it was introduced into other countries as an aquarium plant. The genus *Cabomba* is currently recognized as having five species that are difficult to distinguish from each other. It has now become a pest plant in countries in which it has entered open waters (Mackey & Swarbrick 1997).

As it is better to prevent than to cure, obviously, one would try to prevent from entering the country those species that have an increased likelihood to cause problems. Therefore, an important issue is how to recognize known or suspected invasive plant species during border inspections. As plant material offered for inspection may vary from spores or seed, seedlings and vegetative parts to complete and sterile or fruiting plants, identification based on morphology alone will not always be possible. For those situations where material cannot be identified morphologically, a promising method would be to develop DNA barcodes enabling species identification of any plant part or developmental stage from which DNA can be successfully extracted. DNA barcoding is an aid to taxonomic identification that uses a short, standard DNA region that is universally present in the target lineages and has sufficient sequence variation for species discrimination (Hebert *et al.* 2003; Savolainen *et al.* 2005), but not so much that it is also variable within species. Most recently, the Plant Working Group of the Consortium for the Barcode of Life recommended a two-locus combination of *rbcL* and *matK* as a universal plant barcode (Hollingsworth *et al.* 2009). For this study on aquatic invasive plant groups in the Netherlands, we tested therefore these two sequences. As we previously had good results with another frequently used locus, *trnH-psbA*, in another aquatic invasive species, *Hydrocotyle ranunculoides* (Van de Wiel *et al.* 2009), we also tested this sequence as a barcode to distinguish invasive plants from non-invasive foreign and native species. For this study, we used a panel of representative related species from all over the world, including those found in trade and those native to the Netherlands.

Materials and methods

Plant materials

In this study, 242 samples belonging to 26 species from 10 genera of four families: Onagraceae, Haloragaceae, Hydrocharitaceae and Cabombaceae were collected from several sites in the Netherlands and obtained from other sources or herbaria. An overview of accessions, their species assignments and origins is presented in Table S1 and Fig. S1 (Supporting information). For all newly collected material, voucher specimens have been stored at the National Herbarium Nederland NHN in Leiden (L), the Netherlands. The fresh leaf material was immediately dried and stored on silica gel in bags.

DNA analysis

Extractions were performed on leaf tissue from all samples using the CTAB-mini DNA Extraction protocol. Dry leaf material was disrupted in individual lysing tubes with a bead mill. DNA extraction was conducted following the protocol from Doyle & Doyle (1990). For obtaining high-quality DNA from old dried samples, we used the QIAamp DNA Stool Mini Kit (from QIAGEN) following the manufacturer's instructions. Short fragments of specific regions of plastid DNA (*rbcL*, *trnH-psbA* and *matK*) sequences were amplified from the dried leaf extracts. Universal primers for the three DNA loci used are listed in Table S2 (Supporting information). *rbcL* and *trnH-psbA* primers were used according to Kress *et al.* (2005). We conducted PCR amplification in the first instance using a standard (non-hot-start) DNA polymerase using approximately 20 ng of genomic DNA as a template in a 20-μL reaction mixture (2 μL 10× reaction buffer Dream Taq™ (Fermentas, Lithuania), 2 pM of each dNTP, 4 pM of each primer and 0.2 U Dream Taq™ DNA polymerase). We used the following PCR protocol: one cycle for 5 min at 94 °C, 35 cycles of 30 s at 94 °C, 30 s at 55 °C, 60 s at 72 °C, followed by 10 min at 72 °C. For samples that did not amplify in this way, we used a hot-start DNA polymerase (Amplitaq-Gold™ DNA polymerase from Applied Biosystems), combined with a PCR protocol at a lower stringency (50 °C annealing temperatures and 40 cycles), following the manufacturer's instructions. For *matK*, we tested 41 primer combinations (Table S2, Supporting information) and five amplification protocols (Table S3, Supporting information) and an M13 tail on the primer sets (Table S4, Supporting information), in an attempt to improve amplification and sequencing success. With samples performing poorly in sequencing, PCR products were additionally purified before sequencing using illustra™ Sephadex Columns (from GE Healthcare) according to the manufacturer's

instruction. Direct sequencing of (diluted) PCR was performed using BigDye Terminator v3.1 on a 3100 sequencer (Applied Biosystems). Sequences were aligned by ClustalW (Thompson *et al.* 1994); with *trnH-psbA*, alignments were subsequently checked for the occurrence of ambiguities caused by the presence of indels and edited where necessary (sequence alignments available as online Supporting Information). Genetic distances were computed using MEGA 4.0 according to the Kimura 2-parameter (K2P) model (Tamura *et al.* 2007), to check for a 'barcoding gap' between intraspecific and interspecific variation of the species (Meyer & Paulay 2005; Meier *et al.* 2008; Chen *et al.* 2010). Based on the sequence alignments, dendrograms were constructed by neighbour-joining, using K2P and pairwise deletion of indels in the program MEGA 4.0 (Tamura *et al.* 2007).

Results

PCR amplification and sequencing were successful with the *trnH-psbA* spacer and the *rbcL* locus in most of the samples (Table 1). With the *matK* locus, there were more difficulties in amplification and in sequencing as well. We used 41 primer combinations and at least five amplification protocols (summarized in Tables S2 and S3, respectively, Supporting information) to improve amplification success, which ranged from 58% (*Ludwigia* spp.) to 84% (*Cabomba* spp.) using Phusion Taq polymerase (Table S5, Supporting information). Despite the use of an M13 tail (Table S4, Supporting information), we were able to sequence only 59% of all amplified samples, ranging from 25% in *Myriophyllum* spp. to 93% in *Ludwigia* spp. As *matK* thus did not prove sufficiently reliable for our barcoding purposes, we discuss below in detail only the results of *rbcL* and *trnH-psbA* per plant group.

Table 1 Summary of the proportion of individuals successfully amplified and sequenced, respectively, from two plastid loci in aquatic plant groups assessed in this study

Locus	*rbcL*				*trnH-psbA*			
Plant group	P.A.	P.S.O.	L.(bp)	V.S.	P.A.	P.S.O.	L.(bp)	V.S.
Myriophyllum	90	98	553	29	97	94	352	181
Ludwigia	86	84	553	14†	94	77	534	144†
Hydrocharitaceae	72	93	553	17	87	85	321	101
Cabomba	87	70	553	2	94	66	246	24

P.A, percentage amplification; P.S.O, percentage sequences obtained; L, consensus sequence length; V.S, number of variable sites.
†Excluding outgroups *Fuchsia* and *Circaea*.

Myriophyllum spp

In the *Myriophyllum* genus, we investigated 71 accessions labelled with 12 different species names. *trnH-psbA* and *rbcL* were amplified in 97% and 90%, respectively, of samples. We were able to sequence 94% and 98% of the amplified samples for *trnH-psbA* and *rbcL*, respectively (Table 1). To assess their barcoding effectiveness, we compared the maximum intraspecific distance with the minimum interspecific distance found for each species. The *trnH-psbA* sequence (Fig. S2a, Supporting information) shows more variation in SNPs than *rbcL* (Fig. S2b, Supporting information), which is reflected in higher interspecific distances. In most cases, intraspecific distance was zero, and generally, minimum interspecific distances were larger than corresponding intraspecific distances. Two exceptions are found in *Myriophyllum*, which is due to the occurrence of two aberrant samples, *M. robustum* H20 and *M. spicatum* 3410, respectively. As can be seen in the dendrograms based on neighbour-joining (NJ) analysis of the *trnH-psbA* and *rbcL* sequences alignments in Fig. 1a,b, respectively, all other samples cluster into clearly distinguishable species, with the only exception that *M. robustum* samples could not be distinguished from *M. aquaticum* samples on the basis of their *rbcL* sequence (Fig. 1a,b). One of the invasive species, *M. heterophyllum*, could be distinguished both from the foreign non-invasive species in the same cluster, *M. tuberculatum* and *M. simulans*, and from the native species, *M. alterniflorum, M. spicatum and M. verticillatum*, in the other cluster. In this other cluster, the second invasive species, *M. aquaticum*, could also be distinguished from the native species in this cluster (Fig. 1a,b).

The *rbcL* and *trnH-psbA* sequences additionally proved helpful in identification of samples that showed difficulties on the basis of morphology (e.g. a number of samples of *M. heterophyllum*: 419, 421, 423, B6 and B9) or that carried taxonomically problematic names used in trade. Thus, four samples labelled '*propinquum*' or '*propinum*' either clustered with *M. heterophyllum* (3266 and 3267) or with *M. simulans* (3478 and 3480) and seven samples labelled '*scabratum*' (actually a synonym of *M. pinnatum*) either clustered with *M. heterophyllum* (3373 and 3485) or with an unknown species (provisionally labelled as sp. 1) most likely originating from SE Asia (3306, 3334, 3369, 3440 and 3477). Proper identification of the latter species is the subject of ongoing research. Only with the *trnH-psbA* sequence, *M. robustum* could be separated from *M. aquaticum* and the other species in the cluster. Morphological distinction of *M. robustum* also proved difficult, samples from trade came under *M. 'brasiliense'* (3298, 3342 and 3472) or

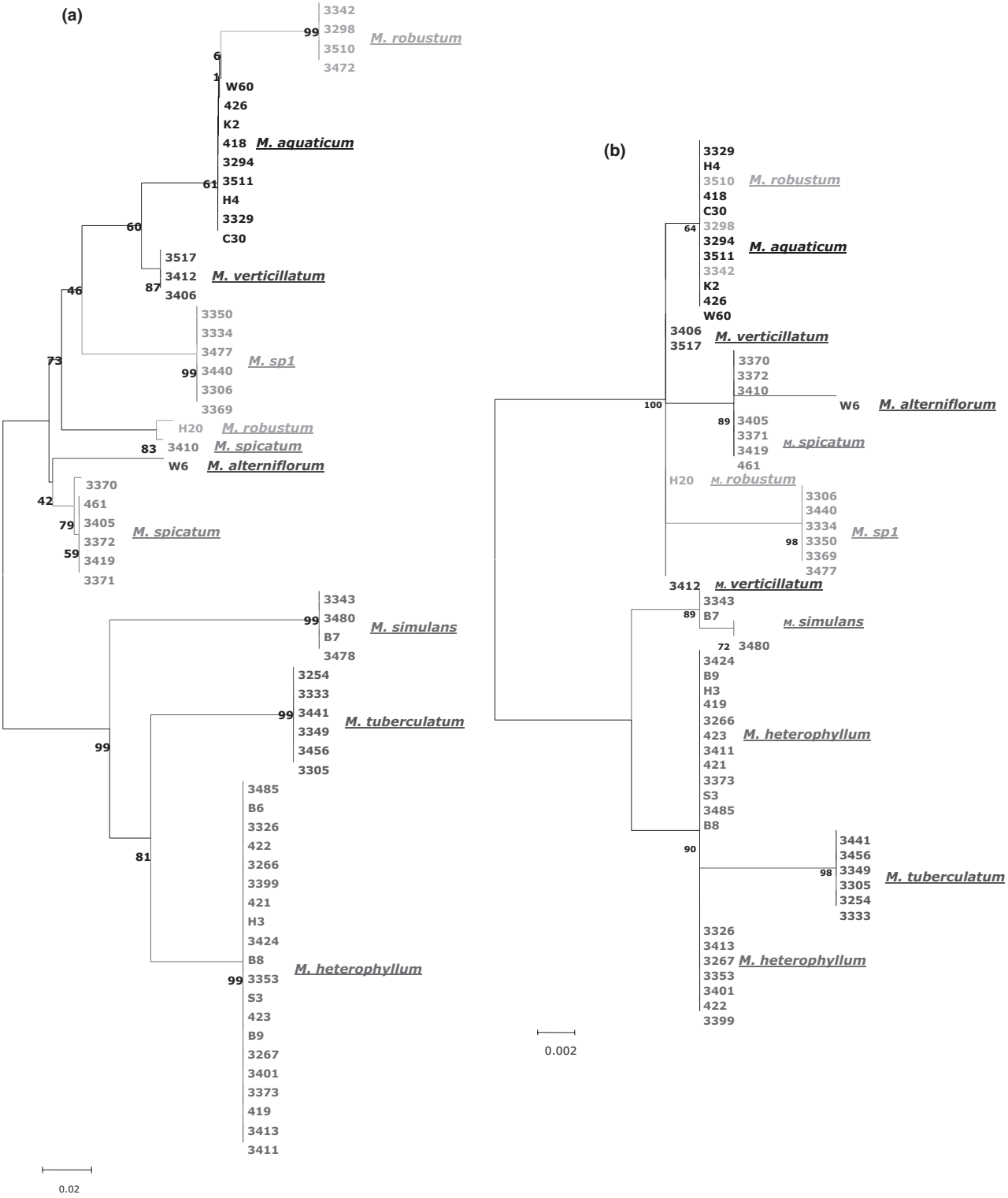

Fig. 1 Neighbour-joining trees (unrooted) of *Myriophyllum* species based on ClustalW alignment of plastid sequences *trnH-psbA* (a) and *rbcL* (b) using MEGA version4.

unidentified (3510). There also remained two problematic identifications, the samples 3410 and H20 already mentioned above: 3410 from the Netherlands was tentatively identified as *M. spicatum* and was confirmed as such by *rbcL*. However, for *trnH-psbA*, it showed a sequence most similar to the one sample that came in as

M. robustum, H20 from New Zealand, its native area. In turn, H20 had an *rbcL* sequence identical to the ones of *M. verticillatum*. Results are summarized in Table 2, which shows that about half of the samples needed renaming on the basis of the DNA barcoding results in combination with morphological re-assessment.

Ludwigia spp

In our study, we assessed 13 species from this genus and two species from other genera *(Circaea* and *Fuchsia*) as outgroup (Table S1, Supporting information). *trnH-psbA* and *rbcL* were amplified in 94% and 86%, respectively, of the samples in *Ludwigia* spp. (Table 1), and we were able to sequence 77% and 84% of amplified samples for *trnH-psbA* and *rbcL*, respectively. The comparison of the maximum intraspecific and minimum interspecific genetic distances for each species did not include the outgroups *Fuchsia* and *Circaea*. In most species, there was a gap between maximum intraspecific distance and minimum interspecific distance. Exceptions were *L. repens* and *L. palustris*, which are part of a hybridizing species complex that is impossible to identify at the vegetative stage, and the one sample of *L. adscendens*. The dendrograms of Fig. 2a (*trnH-psbA*) and 2b (*rbcL*) show that the one sample of *L. repens* clustered in between the variable *L. palustris* samples and that *L. adscendens* could not be discriminated from *L. grandiflora*. Both invasive species, *L. grandiflora* and *L. peploides*, clustered closely together but could still be distinguished from each other. The DNA barcoding results corroborated recently developed morphological markers for the vegetative stage of the two closely related invasive species (shapes of stipules/bracteoles, http://www.q-bank.eu/Plants/, Van Valkenburg 2011). The *L. adscendens* sample could be separated from *L. grandiflora* only by a two-bp inversion that occurred inside an indel in the *trnH-psbA* sequence, so that the calculated interspecific distance from *L. grandiflora* was actually zero, and thus, the small distinction was not visible in the dendrogram of Fig. 2a. In addition, contrary to *trnH-psbA*, *rbcL* could not distinguish between the one sample of *L. octovalvis* and one of the two samples of *L. sedioides* (cf. Fig. 2a,b).

Table 2 Identification of *Myriophyllum* samples with the aid of DNA barcoding

Trade name/ morphological identification	Re-assessment after barcoding	Number of samples confirmed	Number of samples re-identified
alterniflorum	*alterniflorum*	1	
aquaticum	*aquaticum*	8	
brasiliense	*robustum*		3
gigantea/cf. *aquaticum*	sp. 1		1
heterophyllum	*heterophyllum*	11	
mattogrossense	*tuberculatum*		6
propinquum	*heterophyllum*		2
propinquum/propinum	*simulans*		2
robustum	*robustum*†	1	
scabratum	*heterophyllum*		2
scabratum	sp. 1		5
simulans	*simulans*	2	
spicatum	*spicatum*	5	
verticillatum	*verticillatum*	3	
cf. *verticillatum*	*spicatum*		1
Unidentified	*aquaticum*		1
Unidentified	*heterophyllum*		5
Unidentified	*robustum*		1
Unidentified	*spicatum*†		1
Total		31	30

†Sequences different from other *robustum* and *spicatum* samples, respectively, see further under Discussion.

Hydrocharitaceae

In this study, nine species from four genera (*Elodea* spp., *Egeria* spp., *Hydrilla verticillata* and *Lagarosiphon* spp.) from the Hydrocharitaceae family were investigated. *trnH-psbA* and *rbcL* were amplified in 87% and 72%, respectively, of the samples (Table 1). We were able to sequence 85% and 93% of the amplified samples for *trnH-psbA* and *rbcL*, respectively. None of the samples of the invasive species *Hydrilla verticillata* produced unequivocal sequences for both loci. In this group of species, intraspecific distances were mostly zero and there usually was a gap between maximum intraspecific and minimum interspecific distance, except for the two species of the *Elodea* genus in *rbcL* (Fig. S2a,b, Supporting information). This is visualized by the dendrograms of Fig. 3a,b: the invasive *Elodea nuttallii* could be separated from the nowadays non-invasive neophyte *E. canadensis* by *trnH-psbA*, but not by *rbcL*. Invasive *Egeria densa* could be distinguished from non-invasive *E. najas* by both loci. The invasive *Lagarosiphon major* could in principle also be well distinguished from the non-invasive foreign congener, *L. muscoides*. However, two samples identified as *L. muscoides* (627 and 630) showed sequences identical to *L. major* in both *trnH-psbA* and *rbcL* (Fig. 3a,b).

Cabomba

We investigated three different *Cabomba* species, two truly tropical species (*C. furcata* and *C. aquatica*), the first of which recently was recognized as an invasive in Peninsular Malaysia (Siti-Munirah & Chew 2010), and one well-known invasive also occurring at higher latitudes (*Cabomba caroliniana)*. *trnH-psbA* and *rbcL* were amplified in 94% and 87%, respectively, of the samples

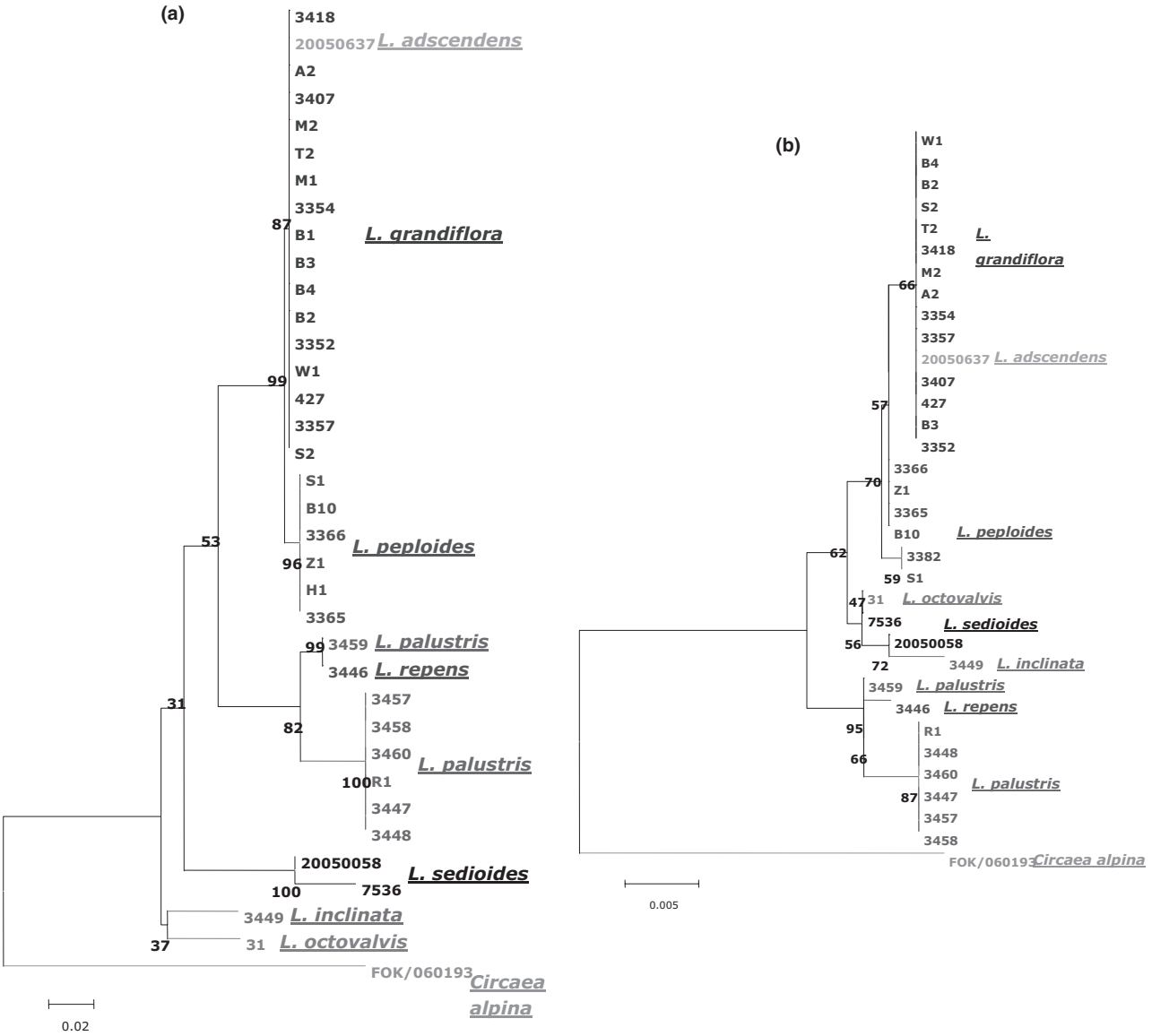

Fig. 2 Neighbour-joining trees (unrooted) of *Ludwigia* species based on ClustalW alignment of plastid sequences *trnH-psbA* (a) and *rbcL* (b) using MEGA version4.

in *Cabomba* spp. We were able to sequence 66% and 70% of amplified samples for *trnH-psbA* and *rbcL*, respectively (Table 1). All intraspecific distances were zero, and minimum interspecific genetic distances for each species were above zero (Fig. S2a,b, Supporting information). Thus, both loci enabled distinction of invasive *C. caroliniana* from the other species (see Fig. 4a,b). The *rbcL* sequence again was less informative than *trnH-psbA*: it showed only two variable sites that only just sufficed to distinguish the three *Cabomba* species tested. Contrary to the other aquatic plant groups tested, there were no indels in *trnH-psbA* in the *Cabomba* species examined and its length (246 bp) was shorter than in the other species tested.

Discussion

We investigated three potential barcoding loci, *trnH-psbA*, *rbcL* and *matK*, for their ability to distinguish invasive exotic from native and/or exotic (imported) non-invasive aquatic species. Although our focus was on the Netherlands, the species groups examined, *Myriophyllum* spp., *Ludwigia* spp., genera from the Hydrocharitaceae, and *Cabomba* spp., have shown invasive behaviour as a result of aquarium and pond trade in several parts of the world. From the three loci, *trnH-psbA* proved to be the most effective in distinguishing the species.

The Consortium for the Barcode of Life Plant Working Group recommended a compromise two-locus

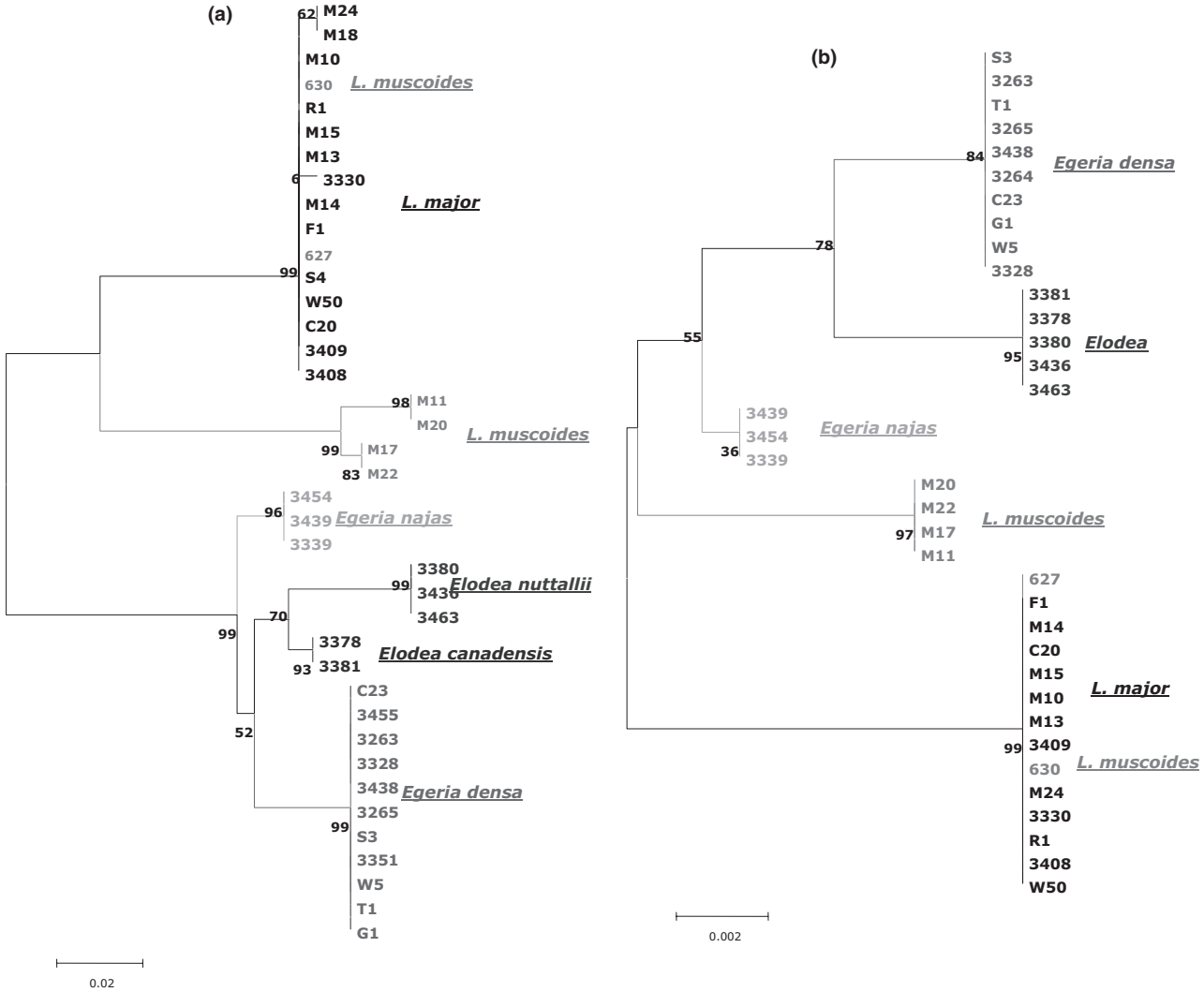

Fig. 3 Neighbour-joining trees (unrooted) of members of the Hydrocharitaceae family based on ClustalW alignment of plastid sequences *trnH-psbA* (a) and *rbcL* (b) using MEGA version4.

standard barcode (*rbcL* + *matK*) for initiating the barcoding process of plant species (CBOL Plant Working Group, Hollingsworth *et al.* 2009), as no single sequence could be identified that was universally effective. In the present study, amplification and sequencing success with *matK* was low. As far as we could assess, *matK* was able to distinguish invasive plant species from non-invasive related species, but in comparison with the other two loci, the efficiency was low with regard to the number of variable sites in relation to the large size of the sequence (800 bp), which contributed to the large sequencing problems. In an earlier study, we were also not able to obtain good *matK* sequences from another invasive aquatic species, *Hydrocotyle ranunculoides*, and its congeners (Van de Wiel *et al.* 2009). There are other mixed reports in the literature about PCR success and sequencing using *matK*, depending upon the use of particular primers and PCR conditions (cf. Roy *et al.* 2010). Although *matK* has been shown to provide a high level of species recovery in several plant DNA barcoding studies on various floristic or biodiversity hotspots (Chase *et al.* 2007; Lahaye *et al.* 2008; CBOL Plant Working group, Hollingsworth *et al.* 2009), this locus was not found useful in several other studies (Kress & Erickson 2007; Chen *et al.* 2010). Moreover, in some complex groups, even the combination of *matK* with *rbcL* was not sufficient to distinguish all species, for example, in the genus *Berberis* (Roy *et al.* 2010). In a recent review of the most optimal barcode for plants, Hollingsworth *et al.* (2011) indicated that none of the barcodes proposed is perfect in every respect and that *matK* still needed optimization of primer combinations, probably to be adapted to specific taxonomic groups.

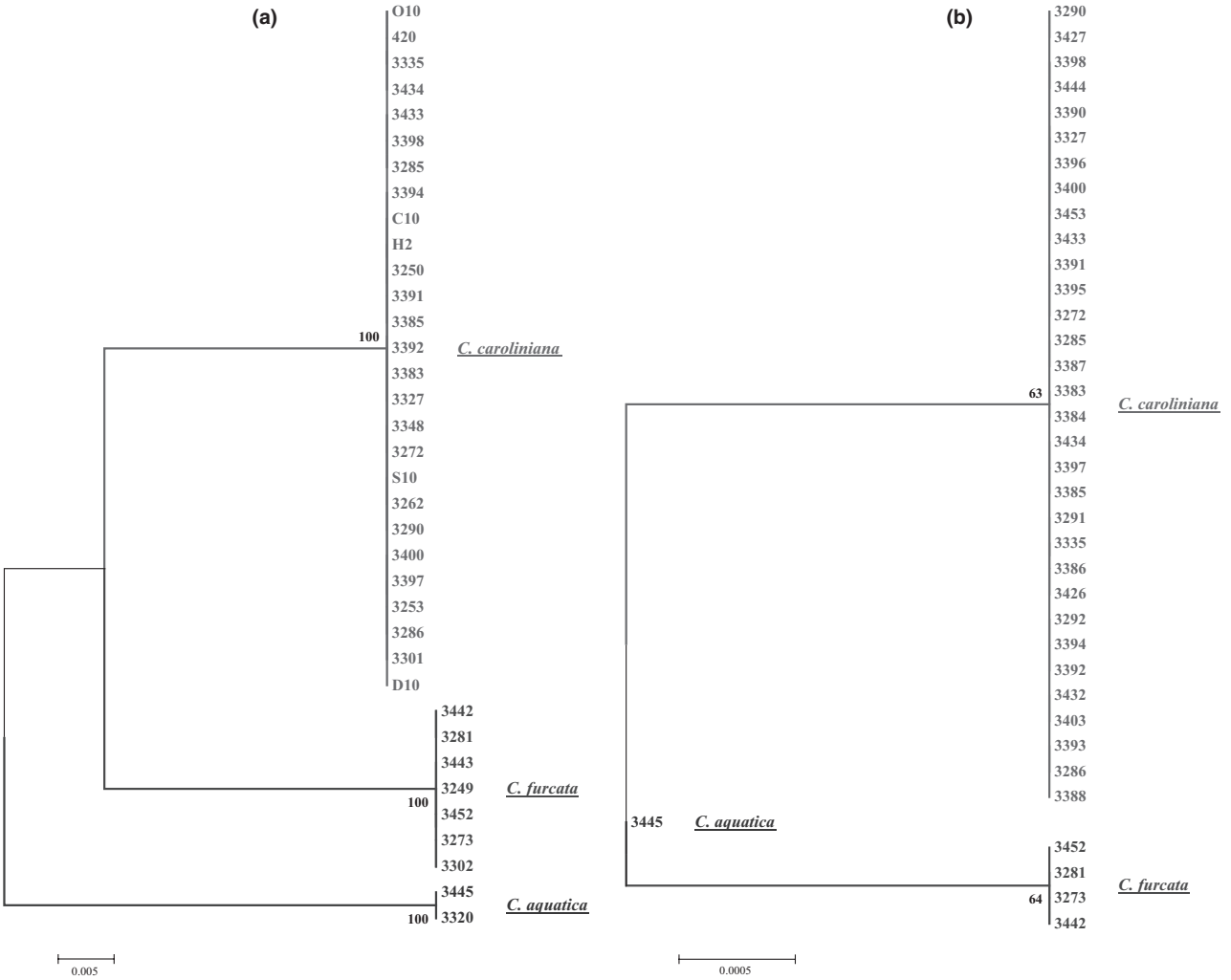

Fig. 4 Neighbour-joining trees (unrooted) of *Cabomba* species based on ClustalW alignment of plastid sequences *trnH-psbA* (a) and *rbcL* (b) using MEGA version4.

The noncoding *trnH-psbA* spacer shows dramatically higher sequence variability than the *rbcL* (and *matK*) coding regions due to both single-base mutations (SNPs) and insertions and deletions (indels) (Kress & Erickson 2007), and *trnH-psbA* was also mentioned as the most popular and obvious locus to be used in addition to the *rbcL* + *matK* core in the recent review by Hollingsworth *et al.* (2011). In our study, *trnH-psbA* was clearly more informative than *rbcL* in all species. Although, for the most part, amplification and sequencing success was slightly better with *rbcL* than with *trnH-psbA*, distinguishing species was significantly better with *trnH-psbA*: it could basically identify all species sequenced, and it did that in a more robust manner, that is, it showed a larger number of variable sites than *rbcL* (*rbcL* failed with eight of the 26 species sequenced in total).

The sequencing of plastid loci proved helpful in identifying specimens that gave problems on morphological grounds, particularly in the genus *Myriophyllum*, where nonflowering plants of closely related species are hard to tell apart. On top of that, in trade, regularly names of uncertain taxonomic status were used for the plants. We were able to show that samples carrying labels, such as '*propinquum*' or '*scabratum*' (synonym of '*pinnatum*'), actually belonged to other species, such as *M. heterophyllum* or *M. simulans*. However, most of the samples labelled '*scabratum*' turned out to belong to a cluster separate from all other species of which names were known on a morphological basis (for the time being labelled as 'sp. 1' in this study). On the same morphological grounds, '*scabratum*' or '*pinnatum*' (a species from North America) could be ruled out as correct species names for these samples. They most likely belong to a species of SE Asian origin, of which we do not know the correct name yet. This is subject of an ongoing study. For *Myriophyllum* in the USA, Thum *et al.* (2012) recently also uncovered

considerable mislabelling of species in trade using ITS sequences. In addition, as with our 'sp. 1', Thum *et al.* (2012) also came across samples from species as yet only known from trade for which identity and origins were unclear; they were similar to *M. aquaticum* and corresponded to M. sp. 'red 1' and M. sp. 'red 2' identified by Moody & Les (2010). In the Hydrocharitaceae group, we could also identify a few morphologically unidentified samples with the aid of the plastid sequences; in this case, it concerned samples belonging to either *Lagarosiphon major* or *L. muscoides* (see Table S1, Supporting information). The value of DNA barcoding for assessing plant identity in the horticultural trade was also shown for other plant groups, such as a fern of the genus *Cheilanthes*, in which case *rbcL*, plus the less often used loci, *atpA* and *trnG-R*, were used (Pryer *et al.* 2010).

In the Hydrocharitaceae, a problem remained with species identification in the genus *Lagarosiphon*: in most cases, *L. major* could be separated from *L. muscoides* by both *rbcL* and *trnH-psbA*, but two samples that were assigned the species name *L. muscoides* on morphological traits double-checked with existing species keys showed sequences identical to all the samples identified as *L. major*. Hybridization could be an explanation for this, but there was not any morphological clue for hybrid origins of the two samples. Further study using nuclear markers could shed light on this. Problems with hybridizing species complexes were also apparent in the *Ludwigia* genus (*L. repens* and *L. palustris*), although too few samples were assessed in these cases to reach firm conclusions. Problems from hybridization were also included in the challenges to DNA barcoding in plants listed by Hollingsworth *et al.* (2011). Another problem arose with the identity of *Myriophyllum robustum*. On the basis of only *trnH-psbA* from the barcoding loci, we could distinguish four trade samples as different from the *M. aquaticum* cluster (see Fig. 1); on morphological grounds, these four samples could be identified as *M. robustum* (see http://www.q-bank.eu/Plants/, Van Valkenburg 2011). However, the only *M. robustum* sample that we obtained from its area of origin, H20 from New Zealand, clustered elsewhere, with *M. verticillatum* in *rbcL*, and in *trnH-psbA*, with an *M. spicatum* sample showing a sequence aberrant from the other *M. spicatum* samples for this locus. In the interpretation of Orchard (1985) by Moody & Les (2010), *M. aquaticum* and *M. robustum* are combined in one 'alliance' separate from other alliances containing *M. verticillatum* or *M. spicatum*, which would be in line with our finding on the four trade samples. In contrast, Moody & Les (2010) showed a closer relationship of *M. robustum* with *M. verticillatum* and *M. spicatum* than with *M. aquaticum*, which would be more in line with our findings on the H20 sample. With this, one should first bear in mind that both the limited number of loci tested here and the limited number of variable sites in them preclude any precise phylogenetic conclusions as all species in this cluster are clearly closely related. With regard to species identification, it is difficult to envisage hybridization as explanation for our observations, as *M. robustum* is native to Australia and New Zealand, whereas *M. verticillatum* and *M. spicatum* are native to Eurasia and *M. aquaticum* to South America. However, as worldwide invasives, *M. spicatum* has been reported for Australia and *M. aquaticum* for New Zealand, respectively.

In conclusion, the *trnH-psbA* spacer proved the most effective barcoding locus for the aquatic angiosperm species groups tested in our study, as was also the case in our pilot study on the aquatic genus *Hydrocotyle*, with its invasive species *H. ranunculoides* (Van de Wiel *et al.* 2009). With *trnH-psbA*, the invasive aquatic species tested could be distinguished from related species, except for *Hydrilla verticillata*, which failed to produce enough sequences of good quality. The combination of *matK* and *rbcL* suggested by the CBOL plant workgroup (Hollingsworth *et al.* 2009) was not working well in our study. The *matK* locus was difficult to amplify and sequence and therefore resulted in a poor return on investments. The *rbcL* locus was the easiest to sequence and align, but showed too little variation to enable identifying all species tested. These results taken together are in line with recent reports on barcoding efforts in plants, *for example,* Costion *et al.* (2011). The *trnH-psbA* locus did show a well-known disadvantage, namely that significant length variations due to insertions, deletions, and simple sequence repeats may hamper sequencing and alignments. Only in *Cabomba* spp., there were no large indels but it still had simple sequence repeats that sometimes gave difficulties with sequencing; the other groups did have large indels among and within species. Nevertheless, within all groups studied, *trnH-psbA* sequences could be aligned, albeit not always unequivocally along the whole length, which meant that occasionally a variable site in the form of an SNP was lost from the analysis due to its occurrence inside an indel. However, this generally did not hamper species identification as can be seen in the trees of Figs 1–4. The only exception was *Ludwigia adscendens*, where the only variable site distinguishing it from *L. grandiflora* was in an indel and therefore was not included in the calculation of genetic distance. Hence, for species identifications within groups, *trnH-psbA* sequence alignment did not pose a serious problem. For wider applications, one could think of alignment-free methods as proposed by, for example, Sims *et al.* (2009) and Little (2011). Barcoding identifications could be helpful in enforcing a ban on the import of invasive plants, a code of conduct for which is already in place in the Netherlands, particularly for cases where

morphology is inadequate to assess species identity. This will become even more so once DNA barcoding would be turned into machinery routinely operable by nonspecialists in botany and molecular genetics (Chase *et al.* 2005).

Acknowledgements

The research was funded by the Netherlands' Ministry of Economic Affairs, Agriculture and Innovation, FES-programme 'Versterking Infrastructuur Plantgezondheid'. Data analysis was partly supported by the same Ministry's project KB-14-003-016. We thank J. Bruinsma, J. Coetzee, L. Triest, D.G. Martin, P. van der Wiel, A. Watterlot, I. Stiers, F. Slik, S. Lee, A. Hussner, D. Hofstra, S. Schooler, P. Clifton, T. Dugdale and D. Officer for providing plant samples.

References

Baars JR, Coetzee JA, Martin G, Hill MP, Caffrey JM (2010) Natural enemies from South Africa for biological control of *Lagarosiphon major* (Ridl.) Moss ex Wager (Hydrocharitaceae) in Europe. *Hydrobiologia*, **656**, 149–158.

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 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 S, Yao H, Han J *et al.* (2010) Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS One*, **5**, e8613.

Costion C, Ford A, Cross H, Crayn D, Harrington M, Lowe A (2011) Plant DNA barcodes can accurately estimate species richness in poorly known floras. *PLoS One*, **6**, e26841.

Dandelot S, Robles C, Pech N, Cazaubon A, Verlaque R (2008) Allelopathic potential of two invasive alien *Ludwigia* spp. *Aquatic Botany*, **88**, 311–316.

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus*, **12**, 13–15.

Hebert PDN, Cywinska A, Ball SL, de Waard 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.

Hollingsworth PM, Graham SW, Little DP (2011) Choosing and using a plant DNA barcode. *PLoS One*, **6**, e19254.

Kress WJ, Erickson DL (2007) A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *PLoS One*, **2007**, 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.

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.

Little DP (2011) DNA barcode sequence identification incorporating taxonomic hierarchy and within taxon variability. *PLoS One*, **6**, e20552.

Mack RN, Simberloff D, Lonsdale WM, Evans H, Clout M, Bazzaz FA (2000) Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications*, **10**, 689–710.

Mackey AP, Swarbrick JT (1997) The biology of Australian weeds 32. *Cabomba caroliniana* Gray. *Plant Protection Quarterly*, **12**, 154–165.

McNeely JA, Mooney HA, Neville LE, Schei P, Waage JK (2001) *A Global Strategy on Invasive Alien Species, X + 50 pp*. IUCN, Gland, Switzerland, and Cambridge, UK.

Meier R, Zhang GY, Ali F (2008) POINTS OF VIEW – The use of mean instead of smallest interspecific distances exaggerates the size of the "barcoding gap" and leads to misidentification. *Systematic Biology*, **57**, 809–813.

Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology*, **3**, 2229–2238.

Moody ML, Les DH (2010) Systematics of the aquatic Angiosperm genus *Myriophyllum* (Haloragaceae). *Systematic Botany*, **35**, 121–139.

Mooney HA, Hobbs RJ (2000) *Invasive Species in a Changing World*. Island Press, Washington, DC, 457 pp.

Orchard AE (1985) Myriophyll (Haloragaceae) in Australasia. II. The Australian species. *Brunonia*, **8**, 173–291.

Pimentel D, Zuniga R, Morrison D (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics*, **52**, 273–288.

Pryer KM, Schuettpelz E, Huiet L *et al.* (2010) DNA barcoding exposes a case of mistaken identity in the fern horticultural trade. *Molecular Ecology Resources*, **10**, 979–985.

Roy S, Tyagi A, Shukla V *et al.* (2010) Universal plant DNA barcode loci may not work in complex groups: a case study with Indian *Berberis* species. *PLoS One*, **5**, e13674.

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

Sims GE, Jun SR, Wua GA, Kim SH (2009) Alignment-free genome comparison with feature frequency profiles (FFP) and optimal resolutions. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 2677–2682.

Siti-Munirah MY, Chew MY (2010) Cabombaceae, a new family record for Peninsular Malaysia. *Malayan Nature Journal*, **62**, 241–248.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596–1599.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL-W - Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–4680.

Thum RA, Mercer AT, Wcisel DJ (2012) Loopholes in the regulation of invasive species: genetic identifications identify mislabeling of prohibited aquarium plants. *Biological Invasions*, **14**, 929–937.

Van de Wiel CCM, Van der Schoot J, Van Valkenburg JLCH, Duistermaat H, Smulders MJM (2009) DNA barcoding discriminates the noxious invasive plant species, floating pennywort (*Hydrocotyle ranunculoides* L.f.), from non-invasive relatives. *Molecular Ecology Resources*, **9**, 1086–1091.

Van Valkenburg JLCH (2011) *Invasieve Waterplanten in Nederland*. Veldgids Nieuwe Voedsel en Waren Autoriteit, Wageningen, 48 pp. (in Dutch).

R.G. conducted this work as part of her PhD thesis under supervision of S.H.M., H.D., J.L.C.H.vV., C.C.M.vdW. and M.J.M.S. conceived and designed the study. H.D. and J.L.C.H.vV. were responsible for sample collection and identification. G.E. and L.P.K. were involved in generating sequences. C.C.M.vdW. and M.J.M.S. supervised the work.

Data Accessibility

DNA sequences: GenBank accessions JX100462–JX100805. Alignments of the sequences are available as online Supporting Information (Appendix S1).

Supporting Information

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

Figure S1. Google map of the worldwide collection of samples.

Figure S2. Minimum interspecific distance with the most similar species against maximum intraspecific distance according to Kimura 2 parameter model for 21 species from four groups of aquatic Angiospermae.

Table S1 Overview of plant accessions, their species assignments, origins, vouchers and Genbank sequence accessions

Table S2 PCR primers used for amplification of plastid DNA sequences

Table S3 PCR amplification protocols used for the *matK* locus

Table S4 Sequences of M13 primers used for amplification and sequencing of the *matK* locus

Table S5 Percentages of matK locus amplification success in four groups of aquatic Angiospermae. N.A: no amplification

Appendix S1. Sequence alignments

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