MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2014) **14**, 272–285 doi: 10.1111/1755-0998.12175

Evaluation of 11 single-locus and seven multilocus DNA barcodes in *Lamium* L. (Lamiaceae)

K. KRAWCZYK, M. SZCZECIŃSKA and J. SAWICKI
Department of Botany and Nature Protection, University of Warmia and Mazury in Olsztyn, Plac Lodzki 1, 10-727, Olsztyn, Poland

Abstract

The aim of this work was to evaluate the suitability of selected DNA regions in the barcoding of plants, based on the species belonging to the genus *Lamium* (Lamiaceae). For this purpose, nine chloroplast barcodes, that is, *accD*, *matK*, *rbcL*, *rpoA*, *rpoB*, *rpoC1*, *rpoC2*, *trnH-psbA*, *trnL-trnF*, as well as ITS nuclear region, and intron of mitochondrial *nad5* gene were tested. Among the single-locus barcodes, most effective in the identification of *Lamium* species was the *trnH-psbA* spacer and *matK* gene. The high level of variability and resolving power was also observed in the case of *rpoA* and *rpoC2* genes. Despite the high interspecies variability of ITS region, it turned out to be inapplicable in *Lamium* identification. An important disadvantage of ITS as a barcode is a limitation of its use in polyploid plants, samples contaminated with fungal material or samples with partially degraded DNA. We have also evaluated five-two-locus and two-three-locus barcode regions created from a combination of most effective single loci. The best-performing barcode combinations were *matK* + *trnH-psbA* and *matK* + *rpoA*. Both of them had equally high discriminative power to identify *Lamium* species.

Keywords: DNA barcoding, *Lamium*, molecular diagnostics, *rpo* genes, species identification

Received 31 January 2013; revision received 6 September 2013; accepted 13 September 2013

Introduction

The idea of DNA barcoding, as a method introducing the use of short DNA sequences for phylogenetic purposes, was first presented on the example of *Tetrahymena* (protozoa) by Nanney (1982). However, the real development of the research on DNA barcoding leading to the emergence of a new trend in this branch of science happened only 20 years later, mainly due to publications of Hebert *et al.* (2003a,b). This method has proven to be extremely useful in the identification of species that cannot be determined based on the morphological characteristics (e.g. juveniles), in forensic analyses where experimental material is usually scarce, as well as in the studies on biodiversity (Hebert *et al.* 2003a; Kress *et al.* 2005). The potential use of this technique was also pointed out as a taxonomic tool, supporting traditional research approach that is based on the analysis of morphological features (Hebert *et al.* 2003a).

DNA barcoding has rapidly developed in the identification of species of animals and red algae, where a comparative analysis of mitochondrial *cox*1 (cytochrome oxidase subunit 1) gene variability is successfully applied (Hebert *et al.* 2003b; Robba *et al.* 2006). *Cox*1 gene was also attempted to be used as a barcode for plants, but studies have shown that most species of higher plants showed insufficient variability of this gene in order to use it as a species-specific sequence (Newmaster *et al.* 2008). Discussion on the selection of the best-performing DNA region or a combination of them that could serve as an effective barcode for plants is still ongoing. A fragment of the genome is still searched for in plants, which would most accurately reflect interspecies variability, and at the same time would show a lesser extent of individual's variability within a species (Fazekas *et al.* 2009; Liu *et al.* 2010; Zhang *et al.* 2012). What is more an ideal barcode should be relatively short (ca. 700 bp or less), so that it could be easily amplified and sequenced. A small size of the amplicon is also of great importance in the case of museum and herbarium materials, which potentially contain degraded DNA. Barcoding loci should also be characterized by the presence of conserved flanking regions that allow for the design of universal primers (Kress *et al.* 2005). Most of the candidate DNA regions for plant barcodes are fragments of the chloroplast genome (cpDNA). This genome shows a higher rate of evolution compared with mitochondrial

Correspondence: Katarzyna Krawczyk, Fax: +48 089 5233546; E-mail: katarzyna.wasowicz@uwm.edu.pl

DNA (mtDNA) (Chase & Fay 2009), as well as stable structure and lack of recombinations (Kress *et al.* 2005); hence, the search for barcode loci within the cpDNA seems substantiated.

The purpose of this work is to evaluate the suitability of selected DNA regions in the barcoding of the genus *Lamium* L. (Lamiaceae). Depending on the taxonomical recognition, a different number of species belonging to this genus is reported; nevertheless, it can be assumed that currently it includes at least 25 species of herbaceous plants (Govaerts *et al.* 2010). Until now, studies on barcoding considered only several species of this genus (Ferri *et al.* 2009; Han *et al.* 2012; Theodoridis *et al.* 2012). *Lamium* is an interesting object of phylogenetic studies and taxonomic discussion, which among other things concerns infrageneric taxa. The current division of this genus into subgenera and sections, determined on the basis of morphological features (Mennema 1989), has been undermined by phylogenetic studies (Bendiksby *et al.* 2011); however, a new division has not yet been proposed.

Lamium comprises three following tetraploid species, which are considered hybrids: *L. confertum* Fr., *L. incisum* Willd. and *L.* × *holsaticum* Prahl. *L. amplexicaule* L. and *L. purpureum* L. are regarded as parental species of *L. confertum* (Jörgensen 1927; Jones & Jones 1965; Bendiksby *et al.* 2011), and *L. incisum* is probably derived from *L. bifidum* Cirillo and *L. purpureum* (Bernström 1955; Taylor 1991; Dvořáková 2000; Bendiksby *et al.* 2011). *L.* × *holsaticum* is commonly believed to be a hybrid between *L. album* L. and *L. maculatum* (L.) L. (Mennema 1989; Govaerts *et al.* 2010). However, molecular data gave no evidence that could support a hybrid origin of *L.* × *holsaticum*. Only its high genetic similarity to *L. maculatum* has been shown (Bendiksby *et al.* 2011). The hybrid species, especially *L. confertum* and *L. incisum*, are difficult to identify on the basis of morphological features. Therefore, definition of genetic markers allowing their identification is of particular importance.

From many regions of cpDNA that have been proposed in the previous studies as barcodes for plants, we analysed nine loci in this work. Seven coding regions were tested including *matK* and *rbcL*. The usefulness of these loci in barcoding is confirmed by the results of numerous studies (Newmaster *et al.* 2006; Chase *et al.* 2007; Lahaye *et al.* 2008; Ford *et al.* 2009; Hollingsworth *et al.* 2009; Newmaster & Ragupathy 2009; Starr *et al.* 2009; Theodoridis *et al.* 2012). The two-locus combination of *matK* and *rbcL* is recommended as the core plant barcode (CBOL Plant Working Group *et al.* 2009; Zhang *et al.* 2012). Among coding regions, we tested also *rpoA*, *rpoB*, *rpoC1* and *rpoC2* genes. The *rpo* genes that encode subunits of plastid-encoded plastid RNA polymerase (PEP) (Little & Hallick 1988; Serino & Maliga 1998) are relatively fast evolving sequences (Gabor *et al.* 1990). For this reason, they have been applied as markers in phylogenetic studies (Petersen & Seberg 1997; Logacheva *et al.* 2007). Until now, only *rpoB* and *rpoC1* of the *rpo* genes have been tested in terms of utility in barcoding studies (Chase *et al.* 2007; Newmaster *et al.* 2008; CBOL Plant Working Group *et al.* 2009; Hollingsworth *et al.* 2009; Liu *et al.* 2010). In the literature, there are scarce data available on the usefulness of *rpoA* and *rpoC2* in plants' barcoding; nevertheless, in recent years, there has been a growing interest in their use (Logacheva *et al.* 2007; Ford *et al.* 2009; Kuang *et al.* 2011). Until now, the *rpoA* gene has been used primarily in the identification of bacterial species (Naser *et al.* 2005, 2007; De Vuyst & Vancanneyt 2007; Park *et al.* 2010). Another cpDNA locus we tested was *accD* gene, which similarly to *rpo* genes is evolving at a fast rate (Yamane *et al.* 2003). We have also included two intergenic regions in our study: *trnH-psbA* and *trnL-trnF*. The usefulness of both spacers for barcoding is the subject of a broad debate (Kress *et al.* 2005; Fazekas *et al.* 2008; Newmaster *et al.* 2008; Hollingsworth *et al.* 2009; Liu *et al.* 2010; Theodoridis *et al.* 2012; Zhang *et al.* 2012). Due to good universality and high discrimination power of *trnH-psbA*, CBOL encourages continued assessment of these loci as supplement to *rbcL* + *matK* (CBOL Plant Working Group *et al.* 2009).

Current work also includes the analysis of a single nuclear and a mitochondrial locus. The tested nuclear sequence is the internal transcribed spacer (ITS), which is considered an effective barcode locus (Kress *et al.* 2005; Chase *et al.* 2007; Zhang *et al.* 2012). In the mitochondrial genome, intron of *nad5* gene was analysed as it is believed to be one of the most variable regions in mtDNA of plants (Wahrmund *et al.* 2009).

By testing the selected loci as potential barcode sequences, we wanted to address the following issues: (i) which loci or their combination would enable the most effective identification of species belonging to the genus *Lamium*, (ii) whether any of the tested loci allow the identification of hybrid species and (iii) whether the application of selected barcodes enables identification of subspecies.

Material and methods

Plant material

In this work, the applicability of selected loci as DNA barcodes was analysed in the comparison of 19 species belonging to *Lamium*. Within certain species, the specimens studied represented different subspecies and varieties, so the total number of taxa amounted to 29 (Table 1). Nomenclature was adopted from Govaerts *et al.* (2010). The exceptions were as follows:

Table 1 The taxa used in the study with the number of analysed specimens

Taxon name	No. of specimens
Lamium album L.	4
Lamium album subsp. *barbatum* (Siebold & Zucc.) Mennema	5
Lamium amplexicaule L.	4
Lamium bifidum Cirillo	5
Lamium bifidum subsp. *balcanicum* Velen.	1
Lamium confertum Fr.	6
Lamium coutinhoi J.G.García	2
Lamium flexuosum Ten.	4
Lamium galactophyllum Boiss. & Reut.	1
Lamium galeobdolon subsp. *argentatum* (Smejkal) J.Duvign.	3
Lamium galeobdolon subsp. *flavidum* (F.Herm.) Á.Löve & D.Löve	4
Lamium galeobdolon subsp. *galeobdolon* (L.) Crantz	3
Lamium galeobdolon subsp. *montanum* (Pers.) Hayek	4
Lamium garganicum subsp. *corsicum* (Gren. & Godr.) Mennema	2
Lamium garganicum subsp. *garganicum* L.	1
Lamium garganicum subsp. *laevigatum* Ces., Pass. & Gibelli	3
Lamium garganicum subsp. *striatum* (Sm.) Hayek	2
Lamium garganicum var. *armenum* (Boiss.) Mennema	1
Lamium gevorense (Gómez Hern.) Gómez Hern. & A.Pujadas	2
Lamium × *holsaticum* Prahl	2
Lamium incisum Willd.	5
Lamium lycium Boiss.	2
Lamium macrodon Boiss. & A.Huet	2
Lamium maculatum (L.) L.	6
Lamium moschatum Mill.	2
Lamium moschatum var. *rhodium* (Gand.) R.R.Mill.	1
Lamium orvala L.	5
Lamium purpureum L.	4
Lamium tomentosum Willd.	3

L. garganicum subsp. *laevigatum* Ces. Pass. & Gibelli, *L. lycium* Boiss., *L. incisum* Willd. and *L. moschatum* var. *rhodium* (Gand.) R.R.Mill., which were not considered by these authors as distinct taxonomic units. The names of these taxa were used as described in the herbarium sheets.

Each barcode has been tested on a pool of 89 individuals. Most of the samples were herbarium specimens, and some were collected by the authors during field research (Appendix S1, Supporting information). The analysis of *matK*, *rbcL*, *trnH-psbA* and *trnL-trnF* comprised also sequences from GenBank. Thanks to this, the collection of species tested was increased by three species in the case of *trnL-trnF* (*L. eriocephalum*, *L. multifidum and L. orientale*) and one species in *matK* locus (*L. orientale*).

Experimental procedures

DNA extraction from leaves was performed with the Genomic Mini AX Plant SPIN kit (A&A Biotechnology) following the manufacturer's protocols. Amplification and sequencing of analysed regions were conducted with the same primers that are given with their references in the Appendix S2 (Supporting information). PCR conditions can be found in Appendix S3 (Supporting information). Purified PCR products were sequenced in both directions using ABI BigDye 3.1 Terminator Cycle Kit (Applied Biosystems®, Foster City, CA, USA) and then visualized using an ABI Prism 3130 Automated DNA Sequencer (Applied Biosystems).

Data analysis

Electropherograms were edited and assembled using Sequencher 4.1.4 (Gene Codes Corporation, Ann Arbor, MI, USA). All sequence data were submitted to GenBank under the following numbers: JX073929-JX074044, KC350510-KC350937, KF055025-KF055208, KF188536-KF188570. The assembled sequences were aligned and manually adjusted with BioEdit 7 (Hall 1999). Regions of ambiguous alignment and incomplete data (*i.e.* at the beginning and end of sequences) were excluded from the analyses. We evaluated 11 single-locus DNA regions, five selected two-locus combinations: *matK* + *trnH-psbA*, *rpoA* + *matK*, *rpoA* + *rpoC2*, *rpoA* + *trnH-psbA*, *rpoC2* + *trnH-psbA* and two-three-locus combinations: *rpoA* + *rpoC2* + *matK* and *rpoA* + *trnH-psbA* + *matK*. Multi locus barcodes were created from a combination of best-performing barcodes in *Lamium* single loci. The effectiveness of each of the analysed barcodes was evaluated with the following methods.

Genetic-distance-based methods. The distribution of intra- and interspecific variability for the studied loci was examined in accordance with the CBOL guidelines pertaining to barcode loci assessment methods. MEGA5 software (K. Tamura, D. Peterson & N. Peterson, submitted) was used to calculate genetic distance over sequence pairs between and within species, based on the number of base differences between sequences (Nei & Kumar 2000). The same calculations were performed using Kimura 2-parameter model (K2P) of nucleotide substitution (Kimura 1980). The variation rate among sites was modelled with a gamma distribution (shape parameter = 0.25).

Based on the obtained results, mean values for each of the tested loci were calculated. For each of the loci, we calculated the number of pairs of species for which the estimated genetic distance was 0.00, that is, they could not be distinguished from each other on the basis of the sequence tested. In addition, the percentage of the species with the occurrence of the so-called barcoding gap was calculated. The barcoding gap is a measure of the effective barcode locus, indicating its high diversity between species with a low intraspecies variability. The barcoding gap is present when the minimum K2P-distance involving a species is larger than its maximum intraspecific distance (Kress & Erickson 2007; Hollingsworth *et al.* 2009; Liu *et al.* 2010).

The usefulness of the tested loci in barcoding was also verified by a method that is based on a direct comparison of DNA sequences. The analysis was carried out with the use of The SpeciesIdentifier 1.7.7 program from TAXON-DNA software package (Meier *et al.* 2006). This tool annotated the analysed DNA fragments as 'best match' and 'best close match', indicating whether the tested sequence allows for the correct identification of the species. The program compared each successive sequence with all the other sequences present in our data set and matched them with the most similar ones. If the indicated pair came from one species, then in the 'best match' analysis, the result was considered positive. If the sequences did not belong to the same species, the result was referred to as 'incorrect'. Several equally good best matches from different species were considered 'ambiguous'. The 'best close match' method was more rigorous, as the tool applied additional intraspecies variability criterion. Pairwise distances were calculated and the threshold was set to 95% (Meier *et al.* 2006). All the results above the threshold were classified as 'no match'. In the case of pairwise distances below the threshold, the sequences were analysed as in the 'best match' method. 'Best match' and 'best close match' analyses were carried out with the use of K2P corrected distances and a minimum sequence overlap of 250 bp.

A tree-based method. A tree-based method was used to assess whether sequences in our data sets form species-specific clusters. Neighbour-joining (NJ) trees were constructed for each individual locus and their combinations. The evolutionary distances were computed in MEGA5 using the number of differences method (Nei & Kumar 2000). Relative support for the branches of the NJ tree was assessed via 1000 bootstrap replicates. We considered a species to be successfully identified if all the specimens representing the species formed a monotypic clade in NJ tree with a bootstrap value above 60%.

Results

Sequence characteristics

The number of sequences obtained is given in Table 2. Success rate of PCR amplification and product sequencing ranged from 64.05% (ITS) to 100% (*trnL-trnF*) (Appendix S4, Supporting information). Obtaining high-quality sequences was difficult for ITS region. Of 89

Table 2 Characteristics of the DNA barcodes evaluated in this study

Barcode	*N* individuals/ *N* species/*N* taxa	Species represented by multiple individuals	Length of sequence [bp]	Alignment length [bp]
accD	88/19/28	18	324	324
ITS	57/12/20	12	662–683	704
matK	88/18/25	17	786–792	792
nad5	72/19/28	18	868	868
rbcL	58/12/16	11	505	505
rpoA	86/19/28	18	713–728	728
rpoB	87/19/28	18	1269	1269
rpoC1	81/19/28	18	701	701
rpoC2	86/19/28	18	676	676
trnH-psbA	105/19/28	19	256–308	375
trnL-trnF	90/22/31	22	314–328	340
matK + trnH-psbA	68/17/24	17	1032–1094	1166
matK + rpoA	68/17/24	17	1499–1520	1520
rpoA + rpoC2	84/19/28	18	1389–1404	1404
rpoA + trnH-psbA	83/19/28	18	984–1036	1103
rpoC2 + trnH-psbA	84/19/28	18	932–984	1051
matK + rpoA + rpoC2K	68/17/24	17	2175–2190	2196
matK + rpoA + trnH-psbA	68/17/24	17	1770–1813	1893

samples, only 57 were positively amplified and qualified for further analysis. One of the reasons for the worse quality of ITS amplicons was a contamination of herbarium materials with the DNA of fungal origin. The universality of the primer was a disadvantageous feature in this case, enabling possible simultaneous amplification of plant and fungal DNA. As a result, overlapping sequencing results of these organisms were unreadable and impossible to separate. Similar issue was encountered with some of the polyploidal species. Sequences of amplicons from different haplotypes of hybrid species were impossible to read. Poor quality of the genetic material was another limitation, in particular for the oldest herbarium specimens where nuclear DNA degradation made successful amplification of ITS region highly problematic. Low success of PCR and sequencing (71.76%) was also obtained for the gene *rbcL* (Appendix S4, Supporting information). In many cases, PCR resulted not in one but two amplification products of different lengths, making it impossible to correctly interpret the results of sequencing. Due to the indicated problems, the ITS and *rbcL* sequences were not acquired for a high number of species. Therefore, we resigned from the analysis of ITS and *rbcL* in combination with other barcodes.

Nine of the sequences we tested had an optimal length for a barcode sequence (Table 2). Only *nad5* (868 bp) and *rpoB* (1269 bp) amplicons exceeded the preferred size. The shortest amplicon, and also the most variable in length, was *trnH-psbA* locus, which ranged from 256 to 308 bp. The variability of *accD*, *nad5*, *rbcL*, *rpoB*, *rpoC1* and *rpoC2* regions was caused solely by the presence of nucleotide substitutions. Within the sequences of remaining barcodes, indels were also reported.

Intra- and interspecific barcode variation

Mean intra- and interspecific genetic distances of evaluated DNA regions are shown in Fig. 1. Among single-locus barcodes, the highest interspecific genetic distance was found in ITS region (No.diff. = 22.21; K2P = 0.042), while the lowest in *nad5* (No.diff. = 0.97; K2P = 0.001). The greatest variation between species in the coding

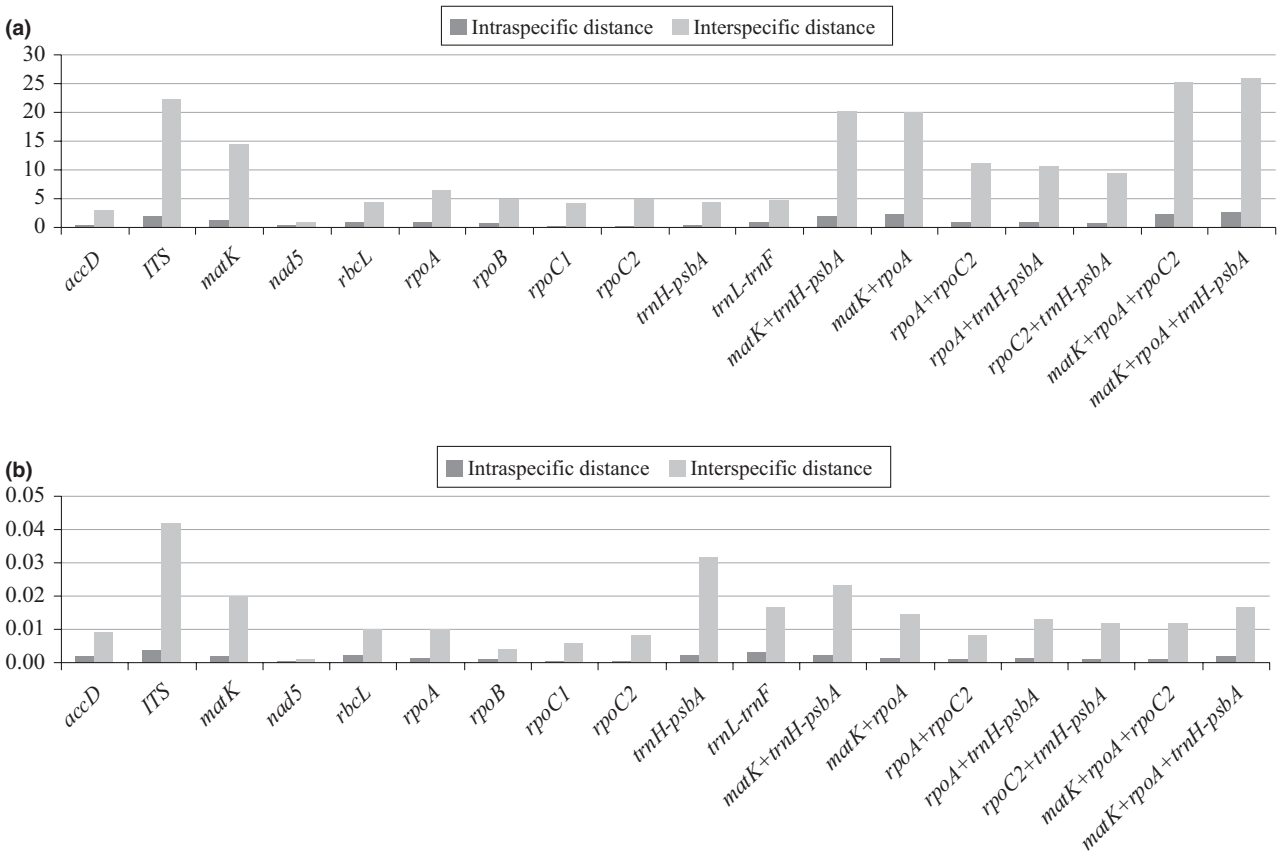

Fig. 1 Mean intra- and interspecific genetic distances of evaluated DNA regions based on (a) the number of base differences between sequences and (b) Kimura 2-parameter model of nucleotide substitution.

regions was found for the *matK* gene (No.diff. = 14.49; K2P = 0.020), followed by the *rpoA* gene (No.diff. = 6.52; K2P = 0.010). Relatively high interspecific genetic distance was found in *trnH-psbA* only using Kimura's model (K2P = 0.032), while the number of base differences between sequences was quite low (No.diff. = 4.62). The most variable two-locus barcode between species was *matK* + *trnH-psbA* (No.diff. = 20.20; K2P = 0.023). High values of the parameters were also reached by the combination of *matK* + *rpoA* (No.diff. = 20.12; K2P = 0.014). Of the two tested three-locus barcodes, slightly higher intra- and interspecific variability was demonstrated for *matK* + *rpoA* + *trnH-psbA* (Fig. 1).

The number of pairs of species, for which the estimated genetic distance was 0.00, is shown in Appendix S5 (Supporting information). The lowest result was obtained for ITS and *rbcL*. For these loci, there was not a single pair of species for which barcode sequences would be identical. Lack of differences between the sequences of the *rpoB* gene occurred only once, that is, for the pair of species *L. confertum* and *L. purpureum*. The same result was obtained for the *matK* gene and for multilocus barcodes containing *matK*. No differences in the sequence of this gene were found only in *L. confertum* and *L. coutinhoi*.

Barcoding gap

The tested barcode sequences varied considerably by the number of species with a barcoding gap. On the graphs depicting the minimum interspecific K2P distances vs. maximum intraspecific K2P distances, the points corresponding to these species can be seen above the 1:1 line (Fig. 2). The percentage of species with barcoding gap is lowest for *nad5* locus (5.56%) while it is highest in the case of ITS region (75.00%) (Fig. 3). The relatively high values of this parameter were observed for *rpoB* (55.56%), as well as for combined barcodes: *matK* + *rpoA* (52.94%), *matK* + *rpoA* + *trnH-psbA* (52.94)%) and *matK* + *rpoA* + *rpoC2* (50.00%). For other barcodes, the number of species characterized by the presence of barcoding gap did not exceed 50% (Fig. 3).

A tree-based method

An evaluation of barcode sequences, based on the phylogenetic tree, turned out to be more restrictive than the method based on the presence of barcoding gap. The results of testing the reliability of clades suggest lower ability to identify the species using the tested barcodes (Fig. 3). This relationship did not apply to genes *rpoA*, *rpoC2*, a combination of *matK* + *trnH-psbA* and *rpoA* + *trnH-psbA*, for which the tree method identified the same number of species as the barcoding gap technique (22.22%, 16.67%, 41.18% and 38.89%, respectively). Most species were identified using the analysis of the ITS tree (58.33%). The relatively high success of species identification (41.18%) was also demonstrated in the case of *matK* gene and the following multilocus barcodes: *matK* + *trnH-psbA*, *matK* + *rpoA*, *matK* + *rpoA* + *trnH-psbA*. Due to the large number of tested barcodes, we decided to include in the paper the trees obtained only for *matK* + *rpoA* (Fig. 4a) and for *matK* + *trnH-psbA* (Fig. 4b). Other phylogenetic trees were added as Supplementary Material (Appendix S6, Supporting information).

The analysis of NJ trees allowed us to conclude that the investigated barcodes permitted the molecular identification of *Lamium* subspecies. The results demonstrate that it was successful only in a few cases. The easiest to identify was *L. album* subsp. *barbatum*. This subspecies can be identified with the following analyses: *accD*, *ITS*, *matK*, *rbcL*, *rpoB*, *matK* + *rpoA*, *matK* + *trnH-psbA*, *rpoA* + *trnH-psbA* and with the analysis of examined three-locus barcodes. *L. garganicum* var. *armenum* was also a distinctive taxon. Its identification is possible with an analysis of all examined chloroplast barcodes, excluding *rpoC1*, *rpoC2* and *trnH-psbA*. The distinction of *L. garganicum* subsp. *corsicum* from the other representatives of this species was only feasible with the use of *rpoA* applied alone or in combination with other investigated loci. The identification of nine other investigated subspecies was not possible.

'Best match' and 'best close match' analysis

An analysis performed with SpeciesIdentifier, as well as two tests discussed above, showed high efficiency of the ITS region (94.73%) in species identification (Table 3). Among the single-locus barcodes, high discriminatory power had also *matK* gene (73.87–75.00%) and the *trnH-psbA* region (61.54%). The best-performing two-locus barcodes were *matK* + *trnH-psbA* (77.95%) and *matK* + *rpoA* (70.59%). Combining *matK* + *rpoA* with *rpoC2* or *trnH-psbA* as the third loci increased its resolving power only by 1.47% and 2.93%, respectively. Including *rpoA* gene to a combination of *matK* + *trnH-psbA* did not increase the effectiveness of the barcode and even decreased it by 4.43%.

Discussion

Single-locus barcodes

Our results indicate a very high efficiency of the ITS region in the identification of species of the genus *Lamium*. Thus, they correspond well with the results of studies carried out on other groups of plants that also

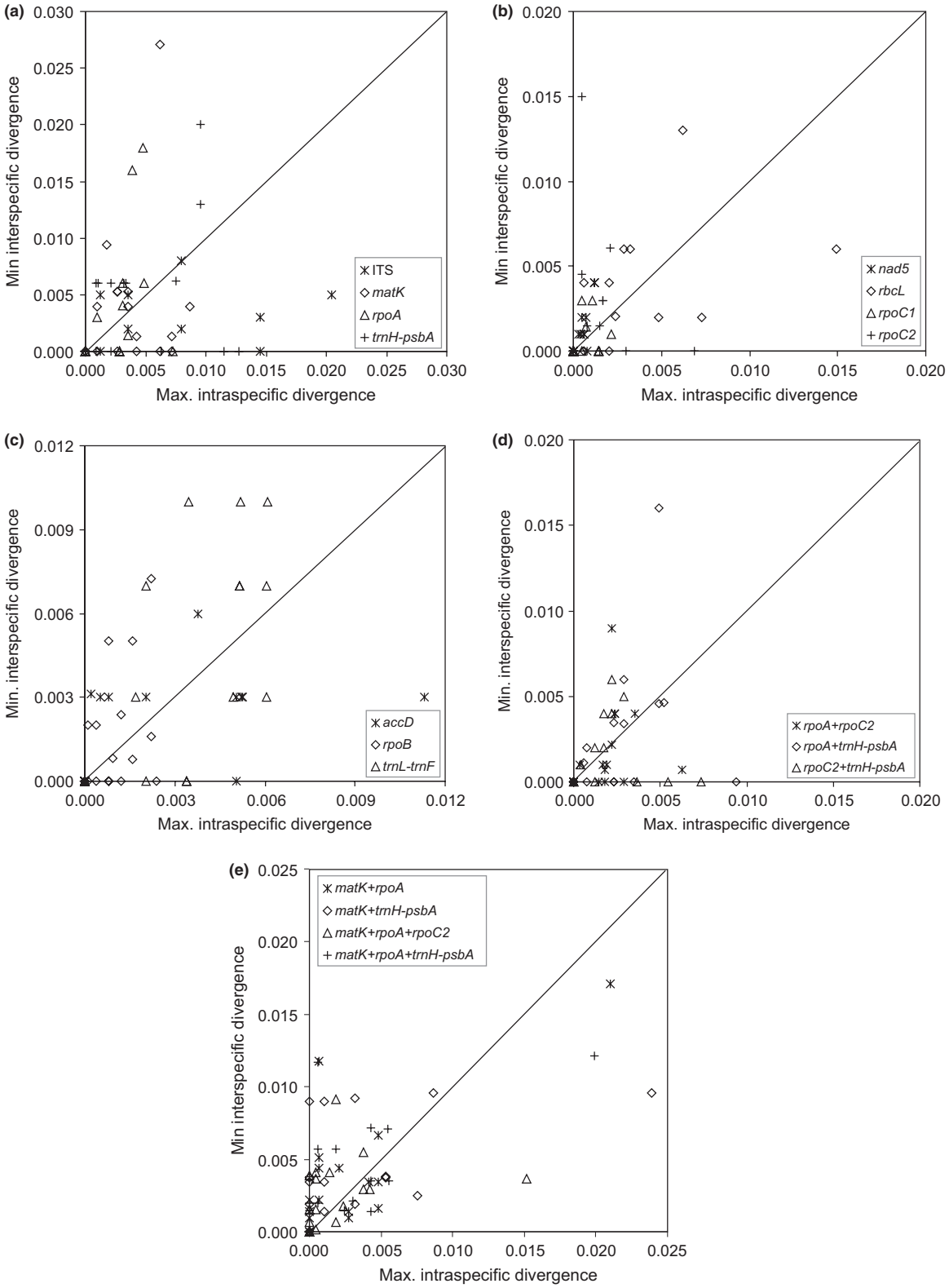

Fig. 2 Minimum interspecific K2P distances vs. maximum intraspecific K2P distances for the individual (a, b, c) and the combined (d, e) barcode regions. Each data point represents a species for which two or more individuals were sampled. Species that fall above the 1:1 line exhibit a barcoding gap. Due to the range of distances between analysed regions, the graphs presented are drawn to individual scales, according to the loci being compared.

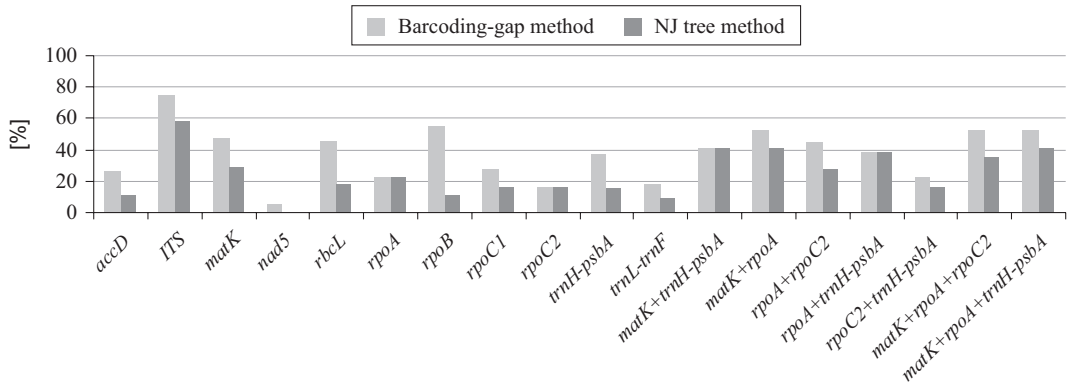

Fig. 3 Identification success rates obtained using barcoding gap and NJ tree methods for the DNA barcodes evaluated in this study.

showed high suitability of ITS region in barcoding (Kress *et al.* 2005; Sass *et al.* 2007; Shi *et al.* 2011; Xiang *et al.* 2011; Zhang *et al.* 2012). However, when comparing the results of our experiments on ITS region with other loci, it should be noted that the set of specimens covered by ITS analysis was significantly smaller. It did not include species of hybrid origin, for which we could not acquire a good-quality sequence. ITS sequence analysis was also impossible for specimens contaminated with the fungal DNA and samples with degraded DNA. Similar difficulties in the analysis of ITS sequences have also been reported by other authors (Alvarez & Wendel 2003; Kress *et al.* 2005; Hollingsworth *et al.* 2011). This low efficiency of amplification and sequencing of the ITS is very disadvantageous from the standpoint of barcoding experiments. One of the key features of this method is its simplicity and ease of use, and therefore, it is crucial for the barcode sequences to be highly applicable. When this criterion is taken into account, in our opinion, it disqualifies ITS region as a useful barcode for *Lamium*.

The problem of low efficiency of amplification and sequencing was also visible in the analysis of the gene *rbcL*. This problem significantly reduces the usefulness of this region in the studies of *Lamium*, despite its high discriminatory power. When comparing the efficacy of the gene *rbcL* with the results obtained in our study for other loci, one needs to bear in mind the difference in the number of species analysed, similarly as for the ITS region. The efficiency of *rbcL* as a barcode in *Lamium* (53.5%) is greater than the one observed in a cosmopolitan and diverse genus *Lysimachia* (38.1%) (Zhang *et al.* 2012) and even in the family Lamiaceae (48.8%) (Theodoridis *et al.* 2012).

Among the analysed loci, *matK* gene and the *trnH-psbA* spacer turned out to be most useful for DNA barcoding. Both of these barcodes demonstrated a high interspecies variability and at the same time a low variation within species. Both *matK* and *trnH-psbA* presented relatively high efficiency in the identification of the species tested, regardless of the method used. The results of 'best match' analysis carried out for the region *trnH-psbA* in *Lamium* are very similar to the results reported for other plant genera. The percentage of correctly identified samples, obtained in this study, was 61.5%, while in *Lysimachia,* it amounted to 56.7% (Zhang *et al.* 2012). In *Compsoneura* genus, it amounted to 65.8% (Newmaster *et al.* 2008). The studies encompassing a larger taxonomic group (family Lamiaceae) have proven even greater discriminative power of *trnH-psbA* barcode, reaching up to 77.6% (Theodoridis *et al.* 2012). In the case of the *matK*, the success of gene identification calculated for *Lamium* (75.0%) was greater than for the aforementioned taxonomic groups. In the genus *Lysimachia,* it amounted to 60.8% (Zhang *et al.* 2012); in the genus *Compsoneura,* it was 48.6% (Newmaster *et al.* 2008); and in the family Lamiaceae, it was 65.9% (Theodoridis *et al.* 2012). The presence of homopolymers that interfere with obtaining high-quality sequences is considered one of the main obstacles in the successful analysis of *trnH-psbA* region variability (Shaw *et al.* 2005; Chase *et al.* 2007; Fazekas *et al.* 2008; CBOL Plant Working Group *et al.* 2009). In *Lamium,* only one thymidine homopolymer was detected, ranging in length from 7 to 12 base pairs and one chain of adenosine repeats, which was six base pairs long. While the short poly-A homopolymer did not interfere with the sequencing reaction, the thymidine repeats decreased quality of reads causing a problem in determining the precise number of repeated bases.

Among the *rpo* genes, greater variability and greater discriminatory power was exerted so far, by rarely investigated genes *rpoA* and *rpoC2*. Both of these genes were of desired length for barcode sequences, and their alignment was not problematic. An unfavourable circumstance, reducing the usefulness of *rpoA* in DNA barcoding, is the phenomenon of its loss from the chloroplast genome of bryophyta (Sugiura *et al.* 2003), and

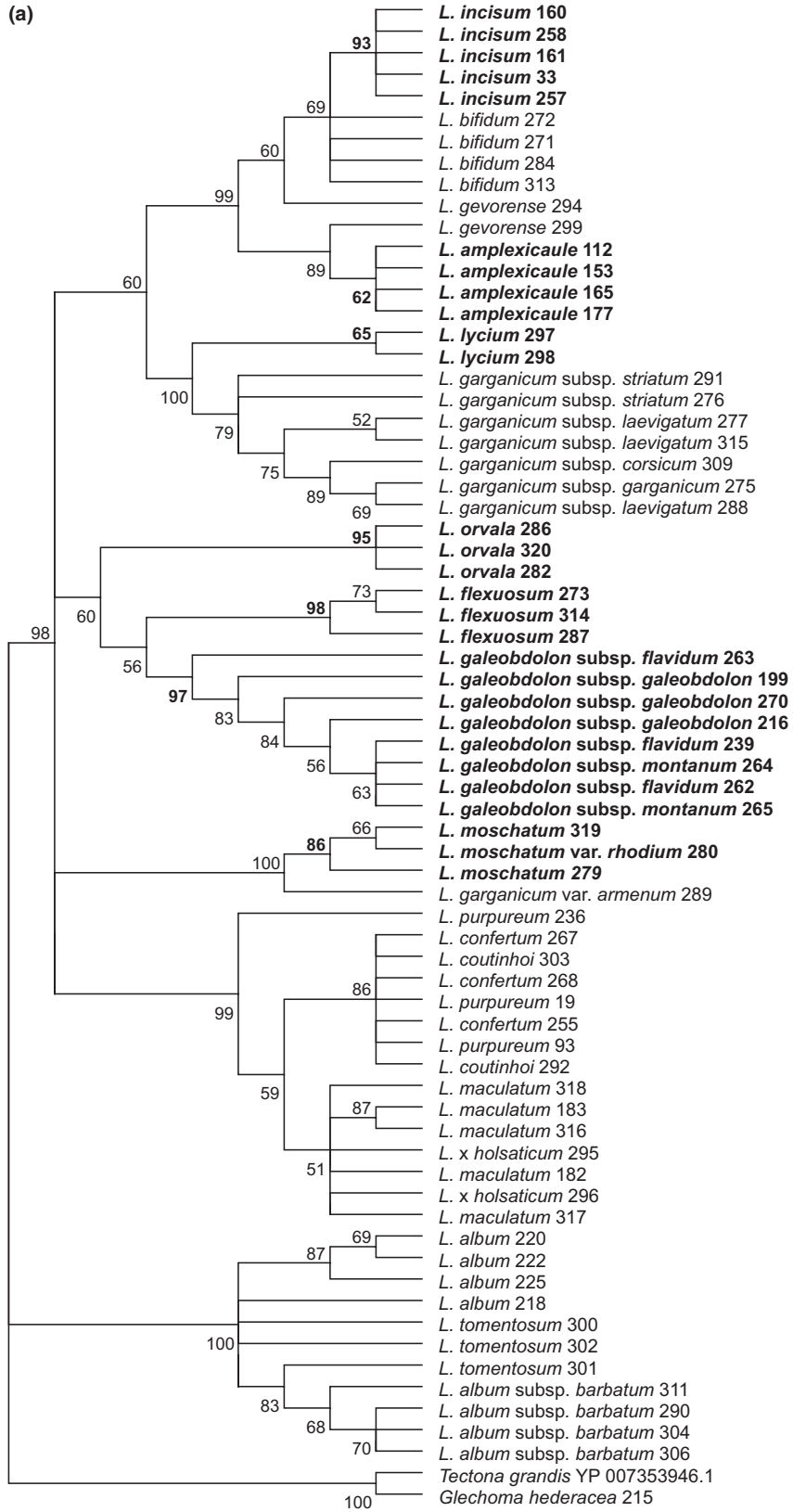

Fig. 4 Neighbour-joining 50% majority-rule consensus tree for *Lamium* generated using (a) the *matK* + *rpoA* combination and (b) the *matK* + *trnH-psbA* combination. Monophyletic species with bootstrap values above 60% are given in bold.

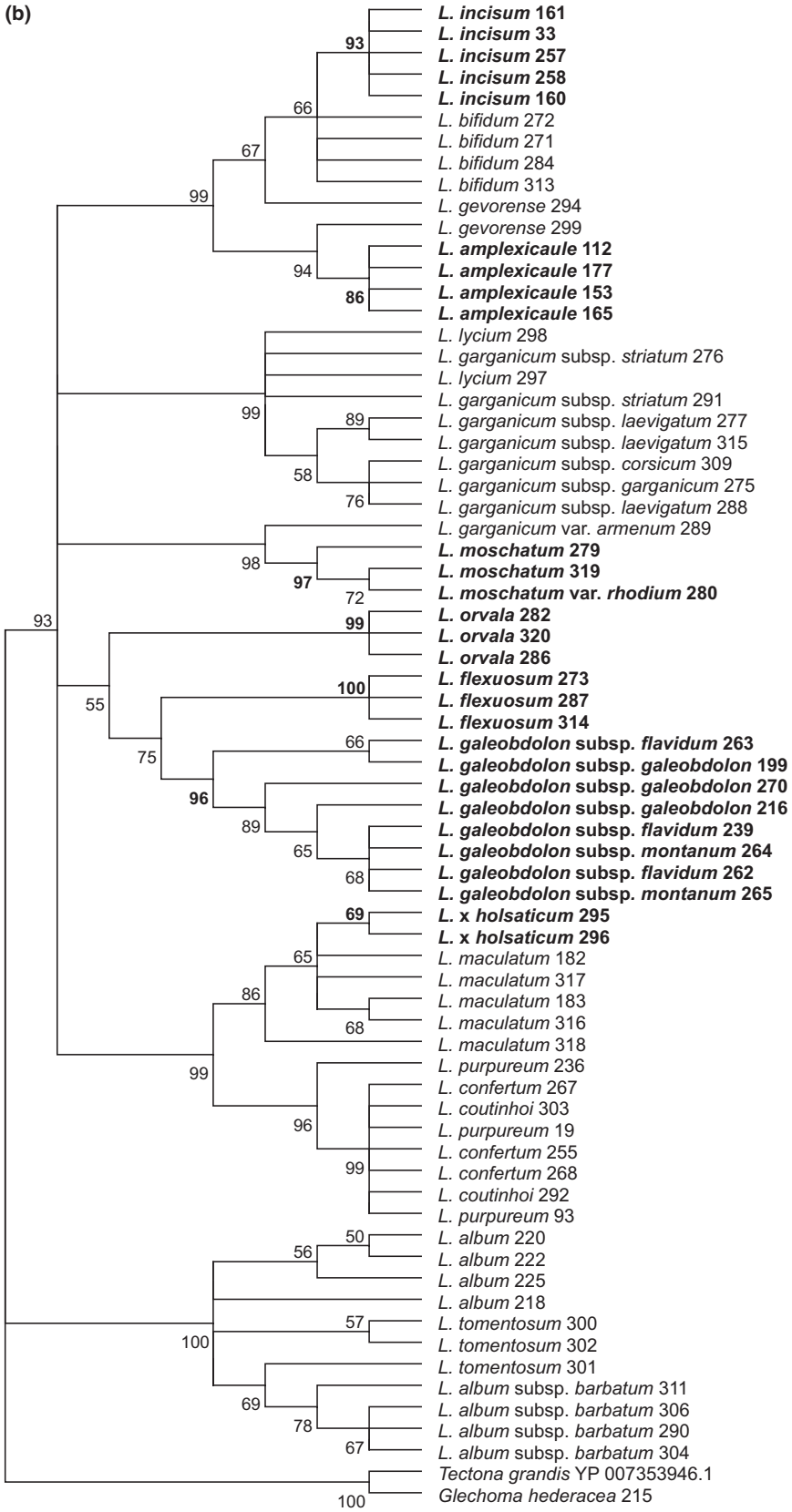

Fig. 4 (Continued)

Table 3 Identification success of analysed barcodes using SpeciesIdentifier 1.7.7 program under 'best match' and 'best close match' methods (Meier *et al.* 2006)

	Best match [%]				Best close match [%]				
Barcode	No. sequences*	Correct	Ambiguous	Incorrect	Correct	Ambiguous	Incorrect	No match [%]	Threshold
accD	88 (87)	13.63	77.28	9.09	13.63	77.28	9.09	0.00	0.93
ITS	57 (57)	94.73	0.00	5.27	94.73	0.00	5.27	0.00	2.47
matK	88 (87)	75.00	19.32	5.68	73.87	18.19	4.54	3.40	0.38
nad5	72 (71)	2.78	93.05	4.17	2.78	91.67	4.17	1.38	0.23
rbcL	58 (56)	53.45	36.21	10.34	53.45	36.21	10.34	0.00	1.39
rpoA	86 (85)	38.38	53.49	8.13	38.38	52.33	8.13	1.16	1.11
rpoB	87 (85)	24.14	68.97	6.89	24.14	67.82	6.90	1.14	0.47
rpoC1	81 (80)	32.10	67.90	0.00	32.10	67.90	0.00	0.00	0.28
rpoC2	86 (85)	37.21	60.47	2.32	37.21	60.47	2.32	0.00	0.44
trnH-psbA	105 (104)	61.54	31.73	6.73	61.54	30.77	5.77	1.92	1.36
trnL-trnF	90 (87)	15.56	75.55	8.89	15.56	75.55	8.89	0.00	0.94
matK + trnH-psbA	68 (68)	77.95	16.17	5.88	77.95	16.17	5.88	0.00	1.76
matK + rpoA	68 (68)	70.59	19.11	10.30	70.59	19.11	10.30	0.00	1.81
rpoA + rpoC2	84 (83)	52.39	40.47	7.14	52.39	40.47	7.14	0.00	0.43
rpoA + trnH-psbA	83 (82)	66.67	21.42	11.91	66.67	20.23	10.71	2.39	0.78
rpoC2 + trnH-psbA	84 (83)	59.52	28.58	11.90	59.52	27.38	10.71	2.39	0.52
matK + rpoA + rpoC2	68 (68)	72.06	19.11	8.83	72.06	19.11	8.83	0.00	1.43
matK + rpoA + trnH-psbA	68 (68)	73.52	14.71	11.77	73.52	14.71	11.77	0.00	1.62

*Numbers in parentheses indicate the number of sequences belonging to species that are represented by >1 individuals.

some representatives of higher plants such as *Pelargonium hortorum* (Palmer *et al.* 1990) and *Epifagus virginiana* (Wolfe *et al.* 1992). Nevertheless, the efficiency level of *rpoA* and *rpoC2* genes in the barcoding of *Lamium* species suggests that these regions should gain greater attention and should be studied in more detail in the future projects.

The *rpoB* and *rpoC1* fragments described here, when compared with *rpoA* and *rpoC2* regions, were less variable and less effective in the molecular identification of the *Lamium* genus species. However, the effectiveness of these DNA regions as barcodes was not particularly low, compared with the results obtained for different taxonomic groups. For instance, *rpoB* gene fragment was tested in a group of 260 species belonging to 30 genera of the family Zingiberaceae and was able to identify more than 45% of the samples (Shi *et al.* 2011), while the identification of closely related *Lamium* species was successful in 24.14%. The results obtained for *rpoC1* region are comparable. In the study of a wide range of species, including both vascular and lower plants, identification success was 43% (Hollingsworth *et al.* 2009), while the result for *Lamium* was 32.10%.

Intron of the *nad5* gene did not meet the criteria imposed on the barcode sequences. Within the studied genus, a low level of interspecies variability was found in this region and thus, it had a low discriminatory power to identify species. All the tests we have used clearly showed that the analysis of the mitochondrial sequence was not an effective way to identify the species we studied.

Multilocus barcodes

It is commonly accepted that an approach based on multilocus chloroplast barcodes is currently the most effective strategy for identification of plant species (Chase *et al.* 2005; Kress *et al.* 2005; Newmaster *et al.* 2008; CBOL Plant Working Group *et al.* 2009). Similar conclusions can be drawn from our results. None of the above-described single loci can be regarded as reliable barcode sequences. However, their usefulness can be increased by combining the most efficient barcode sequences into two-locus barcodes.

Among the five-two-locus barcodes tested, most effective were *matK + trnH-psbA* and *matK + rpoA*. It can be assumed that both have equally high discriminative power to identify the species of *Lamium*. Their resolving power, which for *Lamium* was over 70%, was comparable with the efficiency of *matK + trnH-psbA* (59.61–73.07%) tested within the genus *Holcoglossum* (Xiang *et al.* 2011) and exceeded the resolving power of *matK + rbcL* barcode (59.8–60.8%) in the *Lysimachia* genus (Zhang *et al.* 2012).

The application of *matK* + *rpoA* or *matK* + *trnH-psbA* barcode allowed us to distinguish the hybrid species *L. incisum* from its parent species. In the case of *L. incisum* and *L. bifidum* (which is the so-called organellar parent), the presence of three mutations in *matK* and three mutations in *trnH-psbA* permitted the identification of the species. The fixation of these mutations in *L. incisum* is a result of the evolution of this taxon that arose by hybridization, yet today, it functions as a separate species with the capacity for sexual reproduction (Taylor 1991).

The use of *matK* + *trnH-psbA* also enabled identification of *L.* × *holsaticum*. This species differed from *L. maculatum* in two mutations in *matK* and in one mutation in *trnH-psbA*. In relation to *L. album*, the numbers of preserved mutations were 21 and 10, respectively. The occurrence of these mutations may be interpreted similarly as in *L. incisum*. Although it is reported that in *L.* × *holsaticum*, the pollen is, in part, incorrectly developed (Mennema 1989), there are no data on the sterility of this species.

Both *matK* + *rpoA* and *matK* + *trnH-psbA* indicated a very strong similarity of *L. tomentosum* and *L. album*, making it difficult to distinguish between these taxa. This situation occurred not only in this analysis, but also in that of other chloroplasts and nuclear loci (Bendiksby *et al.* 2011; Krawczyk & Sawicki 2013). Therefore, this rather points to the incomplete lineage sorting of the species discussed and not to too low effectiveness of the barcodes tested.

Worthy of special attention in this study is also clear distinction of *L. garganicum* var. *armenum* from other representatives of *L. garganicum*. This result also should not to be construed as weakness of the barcodes used, because it was previously confirmed by other studies on *Lamium* (Krawczyk & Sawicki 2013).

The division of two investigated representatives of *L. gevorense* between two clades is an interesting result (Fig. 4). The specimen number 294 grouped together with *L. bifidum*, and *L. incisum* is an isoneotype originating from the centre of *L. gevorense* geographic range. The second investigated species (299) was collected in Turkey that is situated outside the range of *L. gevorense* (Govaerts *et al.* 2010). It is supposed that its genetic similarity to *L. amplexicaule* is a result of cryptic speciation.

Of three-locus barcodes discussed in this paper, *matK* + *rpoA* + *trnH-psbA* was able to identify *Lamium* species most effectively. Interestingly, the combination of the most effective two-locus barcode loci in three-locus barcode did not result in a significant improvement of the results obtained. Therefore, the use of three-locus barcode seems cost-ineffective. This study is first to propose *matK* + *rpoA* as well as *matK* + *rpoA* + *trnH-psbA* as barcode loci. Both of them require further studies to determine their effectiveness in other systematic groups of plants.

Acknowledgements

We are grateful to Herbarium Curators of the following institutions for making their collections available: Aarhus University (AAU); National Botanic Garden of Belgium (BR); Botanical Garden in Cordoba (COA); Herbarium des Steiermark. Landesmuseum Joanneum (GJO); Finish Museum of Natural History, Botanical Museum (H); Institut Botanic De Barcelona Herbari BC (HS); Herbarium Instituti Biologiae Academiae Scientarum RSS Latviensis (LATV); Nationaal Herbarium Nederland (L); Uniwersity of Lodz, Department of Geobotany (LOD); University of Warmia and Mazury in Olsztyn, Department of Botany and Nature Protection (OLS); Department of Biology, University of Turku (TUR); Nicolaus Copernicus University (TRN). Financial support for this study came from the National Science Center (Grant 2011/01/N/NZ8/02208) and the European Union through the European Social Fund.

References

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

Bendiksby M, Brysting AK, Thorbek L, Gussarova G, Ryding O (2011) Molecular phylogeny and taxonomy of the genus *Lamium* L. (Lamiaceae): disentangling origins of presumed allotetraploids. *Taxon*, **60**, 986–1000.

Bernström P (1955) Cytogenetic studies on relationships between annual species of *Lamium*. *Hereditas*, **41**, 1–122.

CBOL Plant Working Group, Hollingsworth PM, Forrest LL *et al.* (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences, USA*, **106**, 12794–12797.

Chase MW, Fay MF (2009) Barcoding of plants and fungi. *Science*, **325**, 682–683.

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 standardized protocol to barcode all land plants. *Taxon*, **56**, 295–299.

De Vuyst L, Vancanneyt M (2007) Biodiversity and identification of sourdough lactic acid bacteria. *Food Microbiology*, **24**, 120–127.

Dvořáková M (2000) *Lamium* L.–hluchavka. In: *Květena České Republiky* (ed Slavík B), Volume 6, pp. 596–603, Academia: Praha, Czech Republic.

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

Fazekas AJ, Kesanakurti PR, Burgess KS *et al.* (2009) Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Molecular Ecology Resources*, **9**(Suppl. 1), 130–139.

Ferri G, Alù M, Corradini B, Beduschi G (2009) Forensic botany: species identification of botanical trace evidence using a multigene barcoding approach. *International Journal of Legal Medicine*, **123**, 395–401.

Ford CS, Ayers 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.

Gabor LI, Meinke A, Döry I, Kössel H (1990) Nucleotide sequence of the maize chloroplast *rpoB*/CI/C2 operon: comparison between the derived protein primary structures from various organisms with respect to functional domains. *Molecular and General Genetics*, **221**, 379–394.

Govaerts R, Paton A, Harvey Y, Navarro T (2010) World checklist of Lamiaceae and Verbenaceae. The Board of Trustees of the Royal Botanic Gardens, Kew, Richmond [online]. Available from http://www.kew.org/wcsp/lamiaceae [accessed 15 November 2012].

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

Han JP, Shi LC, Chen XC, Lin YL (2012) Comparison of four DNA barcodes in identifying certain medicinal plants of Lamiaceae. *Journal of Systematics and Evolution*, **50**, 227–234.

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

Hebert PDN, Ratnasingham S, de Waard JR (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **270**, S96–S99.

Hollingsworth M, Clark AA, Forrest LL *et al.* (2009) Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Molecular Ecology Resources*, **9**, 439–457.

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

Jones SB Jr, Jones CA (1965) Status of *Lamium hybridum* Vill. (Labiatae). *American Midland Naturalist*, **74**, 503–505.

Jörgensen CA (1927) Cytological and experimental studies on the genus *Lamium*. *Hereditas*, **9**, 126–136.

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111–120.

Krawczyk K, Sawicki J (2013) The uneven rate of the molecular evolution of gene sequences of DNA-dependent RNA polymerase I of the genus *Lamium* L. *International Journal of Molecular Sciences*, **14**, 11376–11391.

Kress JW, 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*, **2**, e508.

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

Kuang DY, Wu H, Wang YL, Gao LM, Zhang SZ, Lu L (2011) Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implication for DNA barcoding and population genetics. *Genome*, **54**, 663–673.

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, USA*, **105**, 2923–2928.

Little MC, Hallick RB (1988) Chloroplast *rpoA*, *rpoB*, and *rpoC* genes specify at least three components of a chloroplast DNA-dependent RNA polymerase active in tRNA and mRNA transcription. *The Journal Biological Chemistry*, **263**, 14302–14307.

Liu Y, Yan HF, Cao T, Ge XJ (2010) Evaluation of 10 plant barcodes in Bryophyta (Mosses). *Journal of Systematics and Evolution*, **48**, 36–46.

Logacheva MD, Penin AA, Samigullin TH *et al.* (2007) Phylogeny of flowering plants by the chloroplast genome sequences: in search of a 'lucky gene'. *Biochemistry*, **72**, 1324–1330.

Meier R, Kwong S, Vaidya GN, Peter KL (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology*, **55**, 715–728.

Mennema J (1989). *A taxonomic revision of Lamium*. PhD thesis, Leiden University Press.

Nanney DL (1982) Genes and phenes in Tetrahymena. *BioScience*, **32**, 783–788.

Naser S, Thompson FL, Dawyndt P *et al.* (2005) Application of multilocus sequence analysis (MLSA) for rapid identification of *Enterococcus* species based on *rpoA* and *pheS* genes. *Microbiology*, **151**, 2141–2150.

Naser S, Dawyndt P, Hoste B *et al.* (2007) Identification of lactobacilli by *pheS* and *rpoA* gene sequence analyses. *International Journal of Systematic and Evolutionary Microbiology*, **57**, 2777–2789.

Nei M, Kumar S (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.

Newmaster SG, Ragupathy S (2009) Testing plant barcoding in a sister species complex of pantropical *Acacia* (Mimosoideae, Fabaceae). *Molecular Ecology Resources*, **9**(Suppl. 1), 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.

Palmer JD, Calie PJ, de Pamphilis PC *et al.* (1990) An evolutionary genetic approach to understanding plastid gene function: lessons from photosynthetic and nonphotosynthetic plants. In: *Current Research in Photosynthesis*. (ed Baltscheffsky M), Vol. III, pp. 475–482. Kluwer Academic Publishers, the Netherlands.

Park HK, Yoon JW, Shin JW, Kim JY, Kim W (2010) *rpoA* is a useful gene for identification and classification of *Streptococcus pneumoniae* from the closely related viridans group streptococci. *FEMS Microbiology Letters*, **305**, 58–64.

Petersen G, Seberg O (1997) Phylogenetic analysis of the Triticeae (Poaceae) based on *rpoA* sequence data. *Molecular Phylogenetics and Evolution*, **7**, 217–230.

Robba L, Russell SL, Barker GL, Brodie J (2006) Assessing the use of the mitochondrial cox1 marker for use in DNA barcoding of red algae (Rhodophyta). *American Journal of Botany*, **93**, 1101–1108.

Sass C, Little DP, Stevenson DW, Specht CD (2007) DNA Barcoding in the Cycadales: testing the potential of proposed barcoding markers for species identification of cycads. *PLoS ONE*, **2**, e1154.

Serino G, Maliga P (1998) RNA polymerase subunits encoded by the plastid *rpo* genes are not shared with the nucleus-encoded plastid enzyme. *Plant Physiology*, **117**, 1165–1170.

Shaw J, Lickey EB, Beck JT *et al.* (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany*, **92**, 142–166.

Shi LC, Zhang J, Han JP *et al.* (2011) Testing the potential of proposed DNA barcodes for species identification of Zingiberaceae. *Journal of Systematics and Evolution*, **49**, 261–266.

Starr JR, Naczi RF, Chouinard BN (2009) Plant DNA barcodes and species resolution in sedges (*Carex*, Cyperaceae). *Molecular Ecology Resources*, **9**(Suppl. 1), 151–163.

Sugiura C, Kobayashi Y, Aoki S, Sugita C, Sugita M (2003) Complete chloroplast DNA sequence of the moss *Physcomitrella patens*: evidence for the loss and relocation of *rpoA* from the chloroplast to the nucleus. *Nucleic Acids Research*, **31**, 5324–5331.

Taylor R (1991) The origin of *Lamium hybridum*, a case study in the search for the parents of hybrid species. *Northwest Science*, **65**, 116–124.

Theodoridis S, Stefanaki A, Tezcan M, Aki C, Kokkini S, Vlachonasios KE (2012) DNA barcoding in native plants of the Labiatae (Lamiaceae) family from Chios Island (Greece) and the adjacent Çeşme-Karaburun Peninsula (Turkey). *Molecular Ecology Resources*, **12**, 620–633.

Wahrmund U, Rei T, Müller KF, Groth-Malonek M, Knoop V (2009) Fifty mosses on five trees: comparing phylogenetic information in three types of non-coding mitochondrial DNA and two chloroplast loci. *Plant Systematics and Evolution*, **282**, 241–255.

Wolfe KH, Morden CW, Palmer JD (1992) Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proceedings of the National Academy of Sciences, USA*, **89**, 10648–10652.

Xiang XG, Hu H, Wang W, Jin XH (2011) DNA barcoding of the recently evolved genus *Holcoglossum* (Orchidaceae: Aeridinae): a test of DNA barcode candidates. *Molecular Ecology Resources*, **11**, 1012–1021.

Yamane K, Yasui Y, Ohnishi O (2003) Intraspecific cpDNA variations of diploid and tetraploid perennial buckwheat, *Fagopyrum cymosum* (Polygonaceae). *American Journal of Botany*, **90**, 339–346.

Zhang CY, Wang FY, Yan HF, Hao G, Hu CM, Ge XJ (2012) Testing DNA barcoding in closely related groups of *Lysimachia* L. (Myrsinaceae). *Molecular Ecology Resources*, **12**, 98–108.

K.K. - plant material collection, laboratory analyses, data analysis, text editing and M.S., J.S. - research supervisors.

Data Accessibility

Sample information: Supporting Information.
Sequence Alignments: doi:10.5061/dryad.v7q42.

Supporting Information

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

Appendix S1 Plant material, collection details and GenBank accession numbers of material used for evaluation of 11 single-locus and seven multilocus DNA barcodes in *Lamium* L.

Appendix S2 Primers used for the amplification and sequencing of 11 analysed regions with their references

Appendix S3 PCR conditions used for evaluation of 11 single-locus and seven multilocus DNA barcodes in *Lamium* L.

Appendix S4 Success rate of PCR and sequencing of analysed single-loci barcodes.

Appendix S5 Number of pairs of species for which genetic distance was 0.00.

Appendix S6 Neighbour-joining 50% majority-rule consensus tree for *Lamium* generated using 11 single-locus and five multilocus DNA barcodes in *Lamium* L. Monophyletic species with bootstrap values above 60% are given in bold.