The LINNEAN SOCIETY of London BOTANICAL Journal of the Linnean Society

Botanical Journal of the Linnean Society, 2013, **171**, 687–699. With 2 figures

DNA barcoding to analyse taxonomically complex groups in plants: the case of *Thymus* (Lamiaceae)

SILVIA FEDERICI[1], ANDREA GALIMBERTI[1], FABRIZIO BARTOLUCCI[2], ILARIA BRUNI[1], FABRIZIO DE MATTIA[1], PIERLUIGI CORTIS[3] and MASSIMO LABRA[1]*

[1]*ZooPlantLab, Dipartimento di Biotecnologie e Bioscienze, Università degli Studi di Milano Bicocca, Piazza della Scienza 2, 20126 Milano, Italy*
[2]*Centro Ricerche Floristiche dell'Appennino (Parco Nazionale del Gran Sasso e Monti della Laga – Scuola di Scienze Ambientali dell'Università di Camerino), San Colombo, Via Proviciale Km 4.2, 67021 Barisciano (AQ), Italy*
[3]*Dipartimento di Scienze della Vita e dell'Ambiente Macrosezione Botanica ed Orto Botanico, Università degli Studi di Cagliari, Viale S. Ignazio 13, 09123 Cagliari, Italy*

Received 30 December 2012; revised 10 January 2013; accepted for publication 18 January 2013

We evaluated the utility of the core barcode regions (*matK* and *rbcL*) and the plastid intergenic spacer *trnH-psbA* to distinguish between *Thymus* spp. This is a taxonomically complex group that has been investigated so far mainly using morphological approaches. Thirty-six samples representing nine different morphospecies were collected and used for molecular analysis. The three markers showed clear amplification and sequencing. However, the genetic variation and the resulting haplotype networks showed that only *Thymus capitatus* forms a well-defined 'barcoding gap' compared with the other taxa. The identification problems observed in the other *Thymus* spp. may be related to reduced gene flow among populations, resulting in high intraspecific and low interspecific genetic variation. This situation does not permit the definition of species-specific barcodes. A second hypothesis suggests that morphological traits used for the delimitation of *Thymus* spp. do not reflect real biological and molecular species boundaries. If this is the case, the taxonomy of *Thymus* should be revised through extensive sampling and analyses with different tools (i.e. molecular variability, morphology, geographical distribution, etc.) to define the natural units at the species level. *Botanical Journal of the Linnean Society*, 2013, **171**, 687–699.

ADDITIONAL KEYWORDS: *matK* – *rbcL* – species delimitation – *trnH-psbA*.

INTRODUCTION

The use of molecular identification techniques, such as DNA barcoding (Hebert *et al*., 2003), has recently taken on an important role in the definition of taxonomic status and evolutionary processes for almost all existing taxa (Newmaster, Ragupathy & Janovec, 2009). In particular, the promise of DNA barcoding is that it will provide a quick, simple and economic tool to identify and discover biological diversity (Casiraghi *et al*., 2010; De Mattia *et al*., 2012). In 2009, the Consortium for the Barcode of Life (CBOL) Plant Working Group (http://www.barcoding.si.edu/plant_working_group.html) proposed the combination of *matK* and *rbcL* as core barcode regions to universally identify plant species (CBOL Plant Working Group, 2009). These regions were chosen on the basis of the high level of recoverability of high-quality sequences combined with relatively high levels of species discrimination (Hollingsworth, Graham & Little, 2011). In spite of this first proposal, the application of DNA barcoding in plant taxonomy remains contentious, especially for taxa in which molecular variability does not match with morphological differences (e.g. Sass *et al*., 2007; Ford *et al*., 2009; Newmaster & Ragupathy, 2009; Seberg & Petersen,

*Corresponding author. E-mail: massimo.labra@unimib.it

2009; Starr, Naczi & Chouinard, 2009). This is evident in the case of taxonomically complex groups (TCGs) consisting of a genetic mixture of related individuals, often at more than one ploidy, in which biological diversity defies simple classification into discrete species (Ennos, French & Hollingsworth, 2005). Species in TCGs can be difficult to define and identify. This situation leads to practical difficulties in implementing conservation measures on endangered, rare or endemic taxa in TCGs. Secondly, a species-based approach based only on morphological analysis often fails to capture the whole diversity present in TCGs. Generally, the species belonging to TCGs are indeed hard to classify in stable and coherent taxa even for expert taxonomists (Ennos *et al.*, 2012).

In such a context, the taxonomy of TCGs presents two problems: (1) the identification of stable and discrete characters of each species; and (2) the choice of effective tools with which to analyse these characters and to distinguish each species (Rieseberg, Wood & Baack, 2006).

Thymus is one of the TCGs in the European flora, consisting of 215 accepted species, distributed across the Eurasian continent, northern Africa and southern Greenland, with high levels of polymorphism as a result of post-glacial colonization dynamics and hybridization events occurring even between taxa that are not closely related (Bartolucci, 2010). Other important factors influencing the evolution of the genus are polyploidy, disploidy/aneuploidy (Morales, 2002) and gynodioecy (the co-existence of hermaphrodite and 'female' plants with much-reduced anthers and little or no viable pollen) in natural populations (Darwin, 1877; Thompson, 2002).

Jalas (1971), based on morphological characters, divided *Thymus* into two subgenera [*Thymus*, *Coridothymus* (Rchb.f.) Borbás] and eight sections: *T.* section *Micantes* Velen., *T.* section *Mastichina* (Mill.) Benth., *T.* section *Piperella* Willk., *T.* section *Teucrioides* Jalas, *T.* section *Pseudothymbra* Benth., *T.* section *Thymus*, *Hypodromi* (A.Kern.) Halácsy and *T.* section *Serpyllum* (Mill.) Benth. In our work, nine species were analysed using the DNA core barcode markers (*rbcL* and *matK*) and the supplementary intergenic plastid region *trnH-psbA*. One species belongs to subgenus *Coridothymus* and the others belong to three sections of subgenus *Thymus*: *T.* section *Thymus* (one species), *T.* section *Hypodromi* (three species) and *T.* section *Serpyllum* (four species). For one species (*T. striatus* Vahl), we also sampled three different intraspecific taxa to investigate their putative molecular variability with DNA barcoding markers. To evaluate the usefulness of the DNA barcoding approach in TCGs, multiple accessions for each taxon were sampled to investigate: (1) the utility of the three candidate loci for DNA barcoding; (2) the presence of a 'barcoding gap' (Meyer & Paulay, 2005) between the ranges of genetic inter- and intraspecific distances; and (3) the congruence between traditional taxonomic assessments (based on morphological data) and DNA barcoding data.

MATERIAL AND METHODS

Sampling collection and taxonomic identification based on morphology

Thirty-six samples representing nine *Thymus* spp. were collected for morphological and molecular (DNA barcoding) analyses (Tables 1, 2). We studied individuals, mostly from Italy, belonging to one species of subgenus *Coridothymus* and three sections of subgenus *Thymus*. To maximize the chance of observing intraspecific geographical variation, conspecifics were sampled from distant sites. At least three individuals per species were sampled, with the exception of *T. paronychioides* Čelak, an endemic species from Sicily (Italy) with only two known populations (Bartolucci & Peruzzi, in press); this was represented by only two individuals in our dataset. In addition, three different intraspecific taxa of *T. striatus* were considered to evaluate the genetic variability at the intraspecific level.

Taxonomic identification of samples was conducted according to dichotomous keys in floras and taxonomic papers concerning the Mediterranean area (Jalas, 1972; Baden, 1991; Morales, 2010; Bartolucci & Peruzzi, in press). The most distinctive traits of each accession are described in Table 1. In addition, sampled individuals were compared with herbarium specimens kept in the Floristic Research Centre of the Apennines (APP) to confirm their identification. Each specimen was deposited at the same herbarium and a sample for each individual was stored for DNA extraction and analysis. These samples were vouchered as 'MIB:ZPL' following the protocols specified by the Registry of Biological Repositories (http://www.biorepositories.org) and the data standards for BARCODE Records in INSDC (http://barcoding.si.edu/PDF/DWG_data_standards-Final.pdf). A list of samples and voucher names is included in Table 2 and the distribution of sampling localities is depicted in Figure 1.

DNA barcoding analysis

DNA was isolated starting from 20 mg of plant material (young leaves) using the DNeasy Plant Mini kit (Qiagen, Milan, Italy). The concentration of extracted DNA for each sample was estimated both fluorometrically and by comparison of ethidium bromide-stained band intensities with a λ DNA standard. DNA barcoding analysis was performed with the *rbcL* and

Table 1. Principal diagnostic traits of the analysed *Thymus* taxa

Taxon	Supraspecific classification	Habitus	Stem indumentum (upper internode)	Leaf morphology	Leaf hairiness	Leaf lateral veins	Median cauline leaf (length × width) (mm)	Calyx length (mm)	Calyx morphology	Bract morphology	Inflorescence shape	Corolla colour
T. capitatus (L.) Hoffmanns. & Link	*T.* subgenus *Coridothymus* (Rchb.f.) Borbas	Dwarf shrub	Holotrichous	Linear to lanceolate, rarely with revolute margins	Subglabrous or puberulent	Weak	(4)6–10 × 1.5–2	3.5–5.5	Cylindrical, dorso-ventrally compressed with two lateral keels	Not similar to the leaves, ovate to lanceolate, greenish	Capitula	Pink or purple, rarely white
T. vulgaris L. ssp. *vulgaris*	*T.* subgenus *Thymus* section *Thymus* subsection *Thymus*	Plant erect with woody branches	Holotrichous	Linear to elliptical, with revolute margins	Tomentose	Weak	3.5–7(9) × 1–3	2.5–4.5	Campanulate	Similar to the leaves, greenish	Capitula	Pink or purple
T. striatus Vahl ssp. *acicularis* (Waldst. & Kit.) Ronninger	*T.* subgenus *Thymus* section *Hypodromi* (A. Kern.) Halácsy subsection *Subbracteati* (Koklov) Jalas	Plant with non-flowering long creeping branches, rarely with creeping branches ending in a terminal inflorescence	Holotrichous	Acicular to lanceolate, often falcate	Glabrous	Prominent, parallel	4–10 × 0.9–1.5	3.5–4.5	Cylindrical	Not similar to the leaves, trullate, purplish	Capitula, rarely elongated to interrupted	Pink or purple
T. striatus Vahl ssp. *striatus*	*T.* subgenus *Thymus* section *Hypodromi* (A. Kern.) Halácsy subsection *Subbracteati* (Koklov) Jalas	Plant with non-flowering long creeping branches, rarely with creeping branches ending in a terminal inflorescence	Holotrichous	Lanceolate to spathulate	Glabrous	Prominent, parallel	6.5–12(14) × 1.4–2.6	4–5.5	Cylindrical	Not similar to the leaves, trullate, greenish to purplish	Capitula, rarely elongated to interrupted	Pink or purple
T. striatus var. *ophioliticus* (Lacaita) Fiori	*T.* subgenus *Thymus* section *Hypodromi* (A. Kern.) Halácsy subsection *Subbracteati* (Koklov) Jalas	Plant with non-flowering long creeping branches, rarely with creeping branches ending in a terminal inflorescence	Holotrichous	Acicular to lanceolate, often falcate	Glabrous	Prominent, parallel	5.5–13(16) × 0.9–1.5	3.5–5	Cylindrical	Not similar to the leaves, trullate, purplish	Capitula, rarely elongated to interrupted	Pink or purple
T. spinulosus Ten.	*T.* subgenus *Thymus* section *Hypodromi* (A. Kern.) Halácsy subsection *Subbracteati* (Koklov) Jalas	Plant with non-flowering long creeping branches, rarely with creeping branches ending in a terminal inflorescence	Holotrichous	Lanceolate	Hirsute	Weak	6.5–15 × 1.1–2.2(2.7)	4–5(5.7)	Cylindrical	Not similar to the leaves, trullate, greenish	Elongate to interrupted, rarely capitula	White, rarely pink

Table 1. *Continued*

Taxon	Supraspecific classification	Habitus	Stem indumentum (upper internode)	Leaf morphology	Leaf hairiness	Leaf lateral veins	Median cauline leaf (length × width) (mm)	Calyx length (mm)	Calyx morphology	Bract morphology	Inflorescence shape	Corolla colour
T. paronychioides Čelak	*T.* subgenus *Thymus* section *Hypodromi* (A. Kern.) Halácsy subsection *Serpyllastrum* Villar	Plant with long, non-flowering, creeping branches	Holotrichous	Spathulate to lanceolate	Hirsute	Prominent, curved	(5)6–8.5(10.5) × (1.3)1.6–2.6(3.2)	4.5–6	Cylindrical	Not similar to the leaves, trullate, greenish	Capitula	Pink or purple
T. oenipontanus Heinr. Braun	*T.* subgenus *Thymus* section *Serpyllum* (Mill.) Benth. subsection *Isolepides* (Borbás) Halácsy	Plant with creeping branches ending in a terminal inflorescence	Holotrichous	Elliptic-lanceolate	Glabrous	Prominent, curved	5–17 × 1.4–4.5(7)	3–5	Cylindrical	Similar to the leaves, greenish	Elongate to interrupted, rarely capitula	Pink, rarely white
T. longicaulis C.Presl	*T.* subgenus *Thymus* section *Serpyllum* (Mill.) Benth. subsection *Pseudomarginati* (Heinr. Braun ex Borbás) Jalas	Plant with non-flowering, long, creeping branches	Amphitrichous	Linear-lanceolate to elliptical	Glabrous	Weak	(2)6–12(15) × (1.5)2–4.5(6)	2.5–3.5(4.5)	Campanulate	Similar to the leaves, greenish	Capitula	Pink or purple
T. praecox Opiz ssp. *polytrichus* (A.Kern. ex Borbàs) Jalas	*T.* subgenus *Thymus* section *Serpyllum* (Mill.) Benth. subsection *Pseudomarginati* (Heinr. Braun ex Borbás) Jalas	Plant with long, non-flowering, creeping branches	Holotrichous or amphitrichous	Obovate, spathulate to suborbicular	Glabrous to hirsute	Prominent, curved, anastomosate from a marginal vein	5–10(15) × 2–5(7)	3–5(5.6)	Cylindrical	Similar to the leaves, usually purplish	Capitula	Pink or purple
T. pulegioides L. var. *pulegioides*	*T.* subgenus *Thymus* section *Serpyllum* (Mill.) Benth. subsection *Alternantes* Klokov	Plant suberect to procumbent	Goniotrichous	Ovate to oblong-lanceolate	Glabrous	Weak to prominent, curved	4–15 × 1.5–10	3–4	Campanulate to cylindrical	Similar to the leaves, greenish	Elongate to interrupted, rarely capitula	Pink or purple, rarely white

Diagnostic traits for each taxon are provided according to Bartolucci (2010) and Bartolucci & Peruzzi (in press).

Table 2. Sampling table. List of the 36 samples analysed. For each sample, the morphospecies name, voucher number, geographical coordinates of the sampling locality and the haplotype for each barcode marker are provided

Taxon	Voucher sample	Latitude	Longitude	Haplotype		
				rbcL	*matK*	*trnH-psbA*
T. capitatus (L.) Hoffmanns. & Link	MIB:ZPL:04827	37°9′20.01″N	15°14′1.10″E	R5	M20	S33
	MIB:ZPL:04828	41°18′15.92″N	13°33′1.70″E	R5	M20	S33
	MIB:ZPL:04829	40°4′55.03″N	18°28′58.47″E	R5	M20	S34
T. vulgaris L. ssp. *vulgaris*	MIB:ZPL:04819	41°17′31.51″N	13°15′35.33″E	R1	M16	S23
	MIB:ZPL:04820	44°13′27.72″N	9°30′34.05″E	R1	M3	S24
	MIB:ZPL:04821	43°56′02.79″N	7°31′6.94″E	R1	M17	S25
T. striatus ssp. *acicularis* (Waldst. & Kit.) Ronninger	MIB:ZPL:04797	42°13′33.10″N	12°58′26.89″E	R1	M1	S1
	MIB:ZPL:04798	42°17′56.19″N	13°28′23.51″E	R1	M2	S2
	MIB:ZPL:04799	43°38′29.02″N	12°42′54.54″E	R1	M3	S3
T. striatus Vahl ssp. *striatus*	MIB:ZPL:04800	40°09′33.42″N	15°50′20.52″E	R1	M3	S4
	MIB:ZPL:04801	39°54′48.85″N	16°09′17.69″E	R2	M4	S5
T. striatus var. *ophioliticus* (Lacaita) Fiori	MIB:ZPL:04816	43°26′13.10″N	10°36′53.70″E	R1	M3	S20
	MIB:ZPL:04817	43°38′45.50″N	12°02′50.50″E	R4	M15	S21
	MIB:ZPL:04818	43°55′15.30″N	11°04′14.20″E	R3	M11	S22
T. spinulosus Ten.	MIB:ZPL:04814	41°42′10.56″N	15°39′46.61″E	R1	M11	S16
	MIB:ZPL:04815	39°51′26.46″N	16°06′53.39″E	R3	M12	S17
	MIB:ZPL:04843	37°51′31.15″N	14°0′51.07″E	R1	M13	S18
T. paronychioides Čelak	MIB:ZPL:04844	37°51′5.36″N	13°25′28.50″E	R1	M14	S19
	MIB:ZPL:04845	37°51′02.01″N	13°25′50.89″E	R1	M6	S7
T. oenipontanus Heinr. Braun	MIB:ZPL:04822	45°33′01.65″N	10°15′36.02″E	R1	M18	S26
	MIB:ZPL:04823	42°25′12.29″N	13°32′38.85″E	R1	M2	S27
	MIB:ZPL:04824	43°39′06.80″N	12°03′53.40″E	R1	M6	S28
	MIB:ZPL:04825	42°39′21.98″N	13°46′26.79″E	R1	M17	S29
T. longicaulis C.Presl	MIB:ZPL:04802	40°39′28.30″N	14°29′39.00″E	R1	M5	S6
	MIB:ZPL:04803	39°03′13.00″N	16°38′44.60″E	R1	M6	S7
	MIB:ZPL:04804	42°13′10.75″N	12°25′17.15″E	R1	M7	S8
	MIB:ZPL:04805	42°10′40.90″N	13°36′13.30″E	R1	M3	S9
	MIB:ZPL:04806	41°42′27.40″N	15°39′28.70″E	R1	M8	S10
T. praecox Opiz ssp. *polytrichus* (A.Kern. ex Borbàs) Jalas	MIB:ZPL:04846	42°27′2.86″N	13°33′20.59″E	R1	M17	S30
	MIB:ZPL:04848	46°29′43.30″N	11°48′42.16″E	R1	M5	S31
	MIB:ZPL:04826	46°26′3.97″N	11°44′12.40″E	R1	M19	S32
T. pulegioides L. var. *pulegioides*	MIB:ZPL:04807	42°36′46.89″N	0°32′9.16″E	R1	M3	S11
	MIB:ZPL:04808	43°29′2.46″N	10°26′55.88″E	R4	M15	S12
	MIB:ZPL:04809	45°47′30.33″N	14°13′31.57″E	R1	M5	S13
	MIB:ZPL:04810	44°16′04.53″N	9°49′01.90″E	R1	M9	S14
	MIB:ZPL:04811	46°35′50.20″N	6°14′43.52″E	R1	M10	S15

matK coding regions, and *trnH-psbA* intergenic spacer as an additional marker (Hollingsworth *et al*., 2011).

Polymerase chain reaction (PCR) for each marker was performed using puReTaq Ready-To-Go PCR beads (Amersham Biosciences, Milan, Italy) in a 25-μL reaction volume according to the manufacturer's instructions. PCR cycles consisted of an initial denaturation (7 min at 94 °C), 35 cycles of denaturation (45 s at 94 °C), annealing (30 s at 50 °C for *rbcL* and 53 °C for *matK* and *trnH-psbA*) and extension (1 min at 72 °C), and a final extension for 7 min at 72 °C. The genes *rbcL* and *matK* were amplified using primers rbcL1F–rbcL724R (Fay *et al*., 1998) and matK390F–matK1326R (Sun, McLewin & Fay, 2001; Cuènoud *et al*., 2002), respectively, and the non-coding region *trnH-psbA* primers psbA and trnH (Newmaster *et al*., 2009) were used. Amplicons were bidirectionally sequenced using an ABI 155 3730XL automated sequencer at Macrogen Inc., Seoul, South Korea. Manual editing of raw traces and subsequent alignments of forward and reverse sequences using the Bioedit sequence alignment editor (version 7.0.5; Hall, 1999) enabled us to assign edited sequences for

Figure 1. Collection sites of the nine *Thymus* morphospecies. Three different symbols were used to distinguish the three intraspecific taxa of *Thymus striatus*.

Table 3. Percentages of the average Kimura two-parameter (K2P) distance, range of K2P variation and standard error (SE) for each tested marker

Locus	Intraspecific comparison		Interspecific comparison		Overall mean % distance (SE%)
	Mean % variation (SE%)	Range (%)	Mean % variation (SE%)	Range (%)	
matK	0.46 (0.17)	0.00–0.80	0.71 (0.21)	0.30–1.70	0.60 (0.10)
rbcL	0.04 (0.03)	0.00–0.20	0.11 (0.06)	0.00–0.30	0.10 (0.10)
trnH-psbA	1.85 (0.52)	0.70–2.60	2.05 (0.48)	0.90–3.30	2.00 (0.40)

Overall mean distances, intraspecific and interspecific comparisons are considered.

most species. The 3′ and 5′ terminals were clipped to generate consensus sequences for each taxon. The identification of short inverted repeat regions in the *trnH-psbA* spacer was performed as reported by Whitlock, Hale & Groff (2010), using the EMBOSS Software package (Rice, Longden & Bleasby, 2000). The EINVERTED algorithm (Guindon & Gascuel, 2003) was used with default parameters to detect the occurrence of inversions in the *trnH-psbA* region.

All sequences were deposited in GenBank. The accession numbers of the sequences are listed in Supporting Information Appendix S1. To assess the ability of the three barcode regions selected to identify the analysed samples unequivocally, genetic distances among different samples were calculated using MEGA 5.0 (Tamura *et al*., 2011) to generate Kimura two-parameter (K2P) distance matrices. Intra- and interspecific genetic distances were evaluated for all taxa.

Based on sequence similarity, samples were grouped into haplotypes for each barcode marker. Haplotype network analysis was carried out using TCS v. 1.21 (Clement, Posada & Crandall, 2000) which implements the Templeton, Crandall & Sing (1992) statistical parsimony procedure. The maximum number of mutational steps that constitutes a parsimonious connection between two sequence types was calculated with 95% confidence. According to Hart & Sunday (2007), a discrimination approach based on the parsimony connection limit can be useful in a DNA barcoding context. This approach might be useful to assign unknown specimens to known well-sampled taxa (Hart & Sunday, 2007). With this approach, taxa are correctly identified on the basis of correspondence between the number of resulting subnetworks and the number of taxa. A network analysis using this operational species definition might be particularly useful for DNA barcoding studies in which morphological or ecological species markers are labile (Hart & Sunday, 2007). Furthermore, network analysis clarifies the relationships among the haplotypes and defines the correlation between morphological species and genetic diversity.

RESULTS

DNA BARCODING SUCCESS

DNA extraction was successful for all the 36 samples with high DNA quality and good yield (i.e. 30–40 ng μL^{-1}). On amplification, all tested loci, *rbcL*, *matK* and *trnH-psbA*, yielded a single band and exhibited 100% amplification success with standard primers. All the PCR products were easily sequenced and high-quality bidirectional sequences were obtained. After primer trimming and alignment, the same sequence length was observed in all the analysed samples for *rbcL* (599 bp), whereas differences were observed in *matK* and *trnH-psbA* alignments, for which the sequence lengths ranged from 789 to 795 bp and from 394 to 431 bp, respectively, mainly as a result of insertions/deletions (indels).

Sequence alignment was used to evaluate genetic differences among samples; the highest overall K2P distance was shown by *trnH-psbA* (2.00%), followed by *matK* (0.60%) and *rbcL* (0.10%). Analysis performed with EINVERTED (Guindon & Gascuel, 2003) did not find any short inverted repeat regions, and therefore excluded their putative influence on genetic distance values (data not shown).

To evaluate the 'DNA barcoding gap', inter- and intraspecific genetic divergences were computed for each tested marker. The results suggested that *trnH-psbA* exhibited the highest interspecific variation (3.30%), followed by *matK* and *rbcL* with 1.70% and 0.30%, respectively (Table 3). Consistent intraspecific genetic distances were observed for *trnH-psbA* and *matK* among the analysed species, with average K2P distances of 1.85% (range, 0.70–2.60%) and 0.46% (range, 0–0.80%), respectively. For *rbcL*, the intraspecific genetic variation ranged from 0% [*T. vulgaris* L., *T. praecox* Opiz ssp. *polytrichus* (A.Kern. ex Borbàs) Jalas, *T. paronychioides*, *T. oenipontanus* Heinr. Braun, *T. longicaulis* C.Presl and *T. capitatus* (L.) Hoffmanns. & Link] to 0.20% for *T. striatus* (Table 3).

Based on sequence diversity, five haplotypes were identified with *rbcL* and 20 with *matK*. The *rbcL* R1 haplotype was the most common and was shared

amongst all *Thymus* morphospecies, with the exception of *T. capitatus* (haplotype R5). R2 is the rarest, and R3 and R4 are shared by two morphospecies, respectively (Table 3, Fig. 2A). These haplotype distributions resulted in a poor structured species network in which only *T. capitatus* was distinguished from the other taxa (Fig. 2A). Among the *matK* haplotypes, seven were shared by more than one population (Table 3, Fig. 2B). Six haplotypes were detected in the eight *T. striatus* accessions without any genetic structuring among different intraspecific taxa. Thirteen haplotypes (M1, M4, M5, M7, M8, M9, M10, M12, M13, M14, M16, M18 and M19) were exclusive haplotypes (Table 3, Fig. 2B).

In the *matK* network, each haplotype is separated by low genetic distances, with the exception of M15, M20 and the clade including the haplotypes M16, M13, M5, M6 and M8. However, only M20 is private for a single morphospecies (*T. capitatus*).

In the case of *trnH-psbA*, 34 haplotypes were detected for 36 analysed samples (Table 3), suggesting that almost all samples analysed had a distinct *trnH-psbA* sequence, with the exclusion of haplotypes S7 and S33, shared by two samples belonging to the same species (S33) or by two species (S7). This variability did not allow us to define an exhaustive network to explain species relationships, as also revealed by a neighbor-joining tree based on K2P distances provided in Supporting Information Figure S1.

DISCUSSION

Our data show that the three tested DNA barcoding markers can be easily amplified and sequenced for all the analysed samples. However, their discriminating power seems to be inadequate to distinguish the morphospecies considered. Only *T. capitatus* (previously placed in subgenus *Coridothymus*) showed private haplotypes and a well-defined 'barcoding gap'. The separation between *T. capitatus* and the other taxa (subgenus *Thymus*) has also been emphasized by morphological data (i.e. calyx dorso-ventrally compressed with two lateral keels, ciliate), chemical profile (Figueiredo *et al*., 2008), isozymes (Ben El Hadj Ali, Guetat & Boussaid, 2012a) and random amplified polymorphic DNA (RAPD) profile (Ben El Hadj Ali, Guetat & Boussaid, 2012b). All this evidence led to a significant change in the taxonomic status of this species, which was therefore moved to the genus *Thymbra* L. Moreover, this consideration was also highlighted by recent phylogenetic analyses by Bräuchler, Meimberg & Heubl (2010) and Theodoridis *et al*. (2012) based on nuclear and plastid markers and considering different outgroups. Although the aims of our DNA barcoding approach did not include the clarification of phylogenetic issues, our molecular results further support the clear genetic distinction of *Thymbra capitata* (L.) Cav. (= *Thymus capitatus*) from the *Thymus* morphospecies in subgenus *Thymus* (Morales, 1986; Vila, 2002; Bartolucci, 2008; Govaerts *et al*., 2011).

The lower discrimination ability of DNA barcoding in the *Thymus* morphospecies (and intraspecific taxa) involved in this study agrees with the results obtained for different TCGs, such as *Euphrasia* L. (Ennos *et al*., 2005), *Crocus* L. (Seberg & Petersen, 2009), *Carex* L. (Starr *et al*., 2009) and *Dactylorhiza* Neck. ex Nevski (Ennos *et al*., 2005). In these studies, DNA barcoding could not resolve the relationships among closely related taxa resulting from recurrent ecotypic origins or arising through polyploidization or hybridization (Hollingsworth *et al*., 2011). In several cases, the core barcode markers and additional markers did not show any polymorphisms amongst most of the species of the TCGs, such as in the cases of some genera of Meliaceae (Muellner, Schaefer & Lahaye, 2011) and Lamiaceae (De Mattia *et al*., 2011). However, this is not true for *Thymus*, in which consistent intraspecific variability and a certain degree of haplotype diversity were detected using *matK* and *trnH-psbA*. The complex genetic structure observed in this genus could result from a high DNA mutation rate that characterizes *Thymus* populations, combined with interspecific hybridization and polyploidization events (Jalas & Kaleva, 1967; Morales, 1995; Stahl-Biskup & Sáez, 2002; Mahdavi & Karimzadeh, 2010). Another source of variability could also be caused by the conservation of ancestral polymorphisms or by recent speciation events among the morphospecies investigated here (Bräuchler *et al*., 2010).

Hollingsworth *et al*. (2011) have suggested that, for DNA barcoding to work successfully, it requires sufficient time since speciation for mutation to lead to a set of genetic characters grouping conspecific individuals together, separate from other species of the same genus. Although the *Thymus* samples in our study showed clear genetic mutations in the tested DNA barcode regions, sequence variation was not always shared among all the individuals of the same morphospecies. Thus, the genetic diversity showed by the three tested markers does not translate into a 'barcoding gap' situation or a species-specific barcode.

There are two possible explanations for these results. It could be that the analysed *Thymus* spp. are morphologically distinct, but genetic isolation between populations of each species has led to a reduced gene flow, so that mutations have not become fixed across the species. This situation can lead to the development of several population haplotypes not shared at the species level. This hypothesis is also

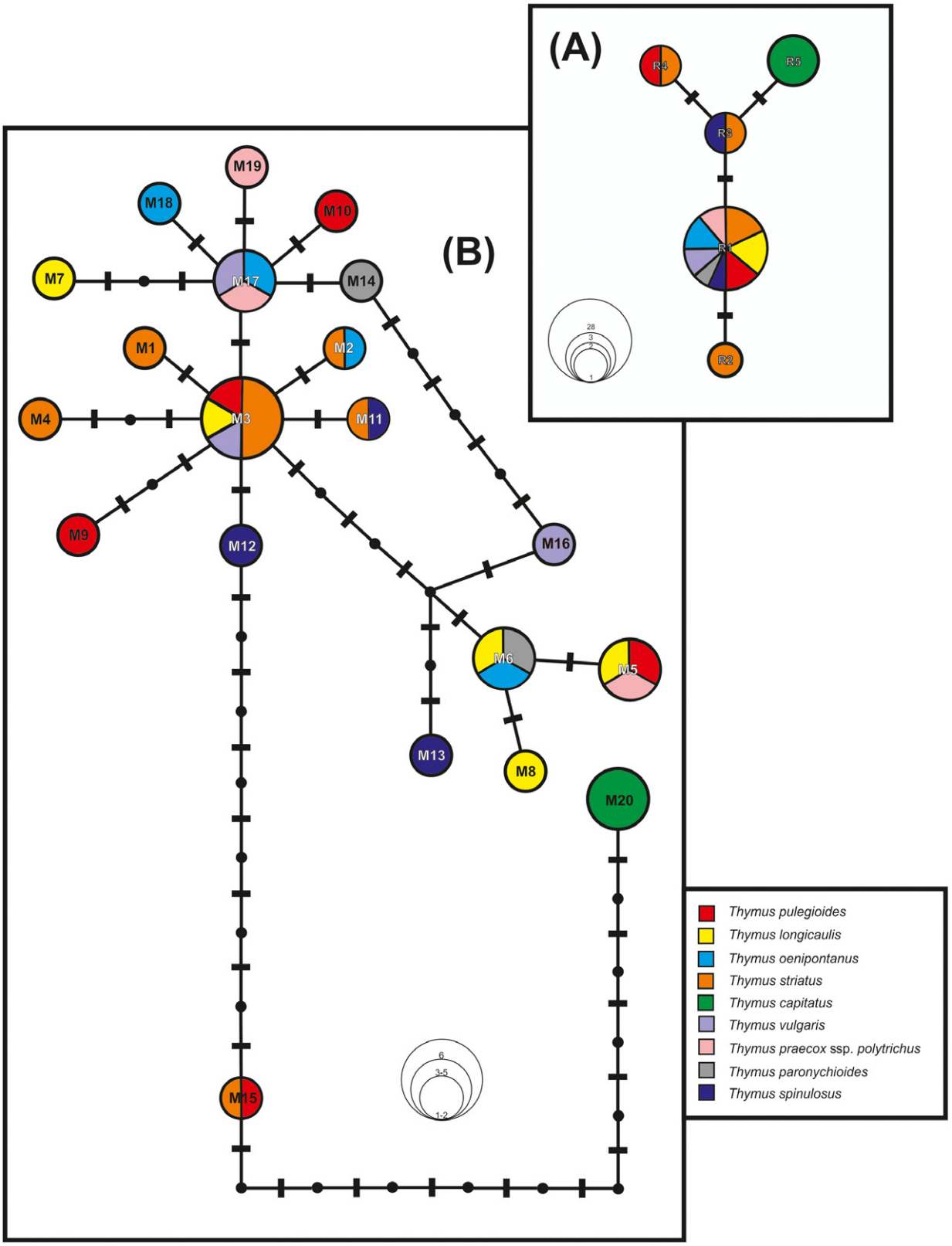

Figure 2. See caption on next page.

Figure 2. Haplotype network reconstruction based on *rbcL* (A) and *matK* (B) barcode markers. Each haplotype pie shows the percentage of sequences belonging to the different *Thymus* morphospecies considered (indicated with different colours). The size of the circles is proportional to the number of individuals sharing a given haplotype. Small black circles are unsampled intermediate haplotypes. Black bars represent single nucleotide substitutions.

supported by population-level molecular analyses of different species of this genus, such as *Thymus herba-barona* Loesel., which has strong among-population differentiation (Molins *et al*., 2011), and *T. quinque-costatus* Čelak. (Quan *et al*., 2009), *T. vulgaris* (Belhassen *et al*., 1993; Tarayre *et al*., 1997) and *T. loscosii* Willk. (López-Pujol *et al*., 2004), in which the genetic diversity is consistent at the population level. If this is the case, our data suggest that the DNA mutation rate in the barcode regions is not the only key element to be required to distinguish species using DNA barcoding. When high intraspecific genetic diversity is accompanied by low gene flow among populations, it makes the DNA barcoding approach ineffective.

Alternatively, it could be that the morphological variation used for the delimitation of *Thymus* morphospecies does not reflect real species boundaries. Such a situation has been discussed by Ennos *et al*. (2005), who suggested that TCGs are sometimes represented by artificial entities identified on the basis of a few, weak morphological traits. This hypothesis is indeed supported by a large number of critical revisions of *Thymus*, with controversial discussion about the values of morphological characters (Morales, 1997; Diklic & Vasic, 2000; Aytas, 2003, 2006; Gomes Pinto *et al*., 2006; Ložiene, 2006; Blanco Salas, Vazquez Pardo & Ruiz Tellez, 2007; Dentant, 2007; Riera, Guemes & Rossello, 2007; Bartolucci, 2010; Molins *et al*., 2011; Bartolucci & Peruzzi, in press).

This scenario explains the sharing of the same haplotypes among different morphospecies of *Thymus* and the failure of DNA barcoding as a tool for species identification. As suggested by Blaxter & Floyd (2003), DNA markers can be useful in systems in which species limits are either subtle or cryptic, but nonetheless clear-cut. This is particularly true for DNA barcoding (Hebert *et al*., 2003), which cannot distinguish any molecular group if clear taxonomic boundaries are not determined.

Based on these results, we conclude that, before the application of DNA barcoding to elucidate the taxonomy of TCGs, an objective and quantitative analysis of taxonomic characters is required to identify natural units at the species level (Rieseberg *et al*., 2006; Jacobs *et al*., 2011). In this context, population genetic markers [e.g. amplified fragment length polymorphisms (AFLPs) or microsatellites or single nucleotide polymorphisms (SNPs)] should be used to investigate the population/species structuring in *Thymus* to verify classification problems, as shown in *Solanum* section *Petota*, in which almost half of the morphologically recognized taxa were found to be not genetically supported (Jacobs *et al*., 2011; see also Ovchinnikova *et al*., 2011). We believe that this approach could be useful in several genera of Lamiaceae, in which relationships among close taxa are still unclear and boundaries among species are sometimes weak.

ACKNOWLEDGEMENTS

This work was supported by the Ministero dell'istruzione, dell'università e della Ricerca Italiano grant 20089BZYAH with the project entitled, 'Tassonomia integrata per lo studio della biodiversità vegetale: DNA barcoding e analisi morfologiche', and from RAS research grant cofinanced by PO Sardegna FSE 2007-2013 L.R.7/2007 with the project entitled, 'Tassonomia integrata (morfologia e DNA barcoding) per la salvaguardia della biodiversità vegetale della Sardegna'. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

Aytas T. 2003. Morphological and anatomical studies on some species of *Thymus* L. (Labiatae) distributed in Karadeniz region. *Ot Sistematik Botanik Degisi* **10:** 31–56.

Aytas T. 2006. Numerical taxonomic studies on some species of the genus *Thymus* L. (Labiatae) in Turkey. *Asian Journal of Plant Sciences* **5:** 782–788.

Baden C. 1991. *Thymus* L. In: Strid A, Tan K, eds. *Mountain flora of Greece*. Edinburgh: Edinburgh University Press, 139–165.

Bartolucci F. 2008. Novità nomenclaturali: Notula 1471. *Thymbra capitata* (L.) Cav. In: Nepi C, Peruzzi L, Scoppola A, eds. *Notulae alla checklist della flora vascolare italiana*. Firenze: Informatore Botanico Italiano, 40: 114.

Bartolucci F. 2010. Verso una revisione biosistematatica del genere *Thymus* L. in Italia: considerazioni nomenclaturali, sistematiche e criticità tassonomica. *Annali di Botanica* **Supplemento 2009:** 135–148.

Bartolucci F, Peruzzi L. In press. *Thymus paronychioides* Čelak. (Lamiaceae), a neglected species from Sicily belonging to *Thymus* section *Hypodromi*. *Folia Geobotanica*.

Belhassen E, Atlan A, Couvet D, Gouyon P-H, Quetier F. 1993. Mitochondrial genome of *Thymus vulgaris* L. (Labiatae) is highly polymorphic between and among natural populations. *Heredity* **71:** 462–472.

Ben El Hadj Ali I, Guetat A, Boussaid M. 2012a. Genetic diversity, population structure and relationships of *Thymus algeriensis* Boiss. et Reut. and *Thymus capitatus* Hoffm. et Link. revealed by isozyme markers. *Industrial Crops and Products* **36:** 149–163.

Ben El Hadj Ali I, Guetat A, Boussaid M. 2012b. Interspecific relationships among two Tunisian *Thymus* taxa: *Thymus capitatus* Hoffm. et Link. and *Thymus algeriensis* Boiss. et Reut. using molecular markers. *African Journal of Biotechnology* **11:** 8810–8819.

Blanco Salas J, Vazquez Pardo FM, Ruiz Tellez T. 2007. Revisión de los generos *Thymbra* L. y *Thymus* L. (Lamiaceae) en Extremadura (España). *Folia Botanica Extremadurensis* **1:** 27–53.

Blaxter M, Floyd R. 2003. Molecular taxonomics for biodiversity surveys: already a reality. *Trends in Ecology and Evolution* **18:** 268–269.

Bräuchler C, Meimberg H, Heubl G. 2010. Molecular phylogeny of Menthinae (Lamiaceae, Nepetoideae, Mentheae) – taxonomy, biogeography and conflicts. *Molecular Phylogenetics and Evolution* **55:** 501–523.

Casiraghi M, Labra M, Ferri E, Galimberti A, De Mattia F. 2010. DNA barcoding: a six-question tour to improve users' awareness about the method. *Briefings in Bioinformatics* **11:** 440–453.

CBOL Plant Working Group. 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 12794–12797.

Clement M, Posada D, Crandall KA. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9:** 1657–1660.

Cuènoud P, Savolainen V, Chatrou LW, Powell M, Grayer RJ, Chase MW. 2002. Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcL*, *atpB* and *matK* DNA sequences. *American Journal of Botany* **89:** 132–144.

Darwin CF. 1877. *The different forms of flowers on plants of the same species*. London: J. Murray.

De Mattia F, Bruni I, Galimberti A, Cattaneo F, Casiraghi M, Labra M. 2011. A comparative study of different DNA barcoding markers for the identification of some members of Lamiaceae. *Food Research International* **44:** 693–702.

De Mattia F, Gentili R, Bruni I, Galimberti A, Sgorbati S, Casiraghi M, Labra M. 2012. A multi-marker DNA barcoding approach to save time and resources in vegetation surveys. *Botanical Journal of the Linnean Society* **169:** 518–529.

Dentant C. 2007. Note sur la section *Serpyllum* (Miller) Bentham du genre *Thymus* L. dans le sud-est de la France. *Bulletin de la Société Linneenne de Provence* **58:** 63–77.

Diklic N, Vasic O. 2000. A taxonomic assessment of *Thymus dacicus*, a neglected member of the flora of Yugoslavia. *Botanika Chronika (Patras)* **13:** 179–186.

Ennos RA, French GC, Hollingsworth PM. 2005. Conserving taxonomic complexity. *Trends in Ecology and Evolution* **20:** 164–168.

Ennos RA, Whitlock R, Fay MF, Jones B, Neaves LE, Payne R, Taylor I, De Vere N, Hollingsworth PM. 2012. Process-based Species Action Plans: an approach to conserve contemporary evolutionary processes that sustain diversity in taxonomically complex groups. *Botanical Journal of the Linnean Society* **168:** 194–203.

Fay MF, Bayer C, Alverson WS, de Bruijn AY, Chase MW. 1998. Plastid *rbcL* sequence data indicate a close affinity between *Diegodendron* and *Bixa*. *Taxon* **47:** 43–50.

Figueiredo AC, Barroso JG, Pedro LG, Salgueiro L, Miguel MG, Faleiro ML. 2008. Portuguese *Thymbra* and *Thymus* species volatiles: chemical composition and biological activities. *Current Pharmaceutical Design* **14:** 3120–3140.

Ford CS, Ayres KL, Toomey N, Haider N, Van Alphen Stahl J, Kelly LJ, Wikström N, Hollingsworth PM, Duff RJ, Hoot SB, Cowan RS, Chase MW, Wilkinson MJ. 2009. Selection of candidate coding DNA barcoding regions for use on land plants. *Botanical Journal of the Linnean Society* **159:** 1–11.

Gomes Pinto C, Vázquez Pardo FM, Paiva-Ferreira R, Ramos S, Doncel E. 2006. Biosystematic study of the subsection *Thymastra* (Nyman ex Velen.) R. Morales of the genus *Thymus* L. (Lamiaceae). *Acta Botanica Gallica* **153:** 355–364.

Govaerts R, Paton A, Harvey Y, Navarro T, Garcia Peña MR. 2011. *World checklist of Lamiaceae*. Kew: Facilitated by the Royal Botanic Gardens. Published on the Internet. Available at: http://www.kew.org/wcsp/accessed [accessed 11 January 2012].

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

Hall TA. 1999. Bioedit: a userfriendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41:** 95–98.

Hart MW, Sunday J. 2007. Things fall apart: biological species form unconnected parsimony networks. *Biology Letters* **3:** 509–512.

Hebert PDN, Cywinska NA, Ball SL, deWaard JR. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences* **270:** 313–321.

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

Jacobs MMJ, Smulders MJM, van den Berg RG, Vosman B. 2011. What's in a name; genetic structure in *Solanum* section *Petota* studied using population-genetic tools. *BMC Evolutionary Biology* **11:** 42–62.

Jalas J. 1971. Notes on *Thymus* L. (Labiatae) in Europe. I. Supraspecific classification and nomenclature. *Botanical Journal of the Linnean Society* **64:** 199–235.

Jalas J. 1972. Thymus L. In: Tutin TG, Heywood VH, Burges NA, Moore DM, Valentine DH, Walters SM, Webb DA, eds. *Flora Europaea 3*. Cambridge: Cambridge University Press, 172–182.

Jalas J, Kaleva K. 1967. Chromosome studies in *Thymus*. *Annales Botanici Fennici* **4:** 74–80.

López-Pujol J, Bosch M, Simon J, Blanché C. 2004. Allozyme diversity in the tetraploid endemic *Thymus loscosii* (Lamiaceae). *Annals of Botany* **93:** 323–332.

Ložiene K. 2006. Instability of morphological features used for classification of *Thymus pulegioides* infraspecific taxa. *Acta Botanica Hungarica* **48:** 345–360.

Mahdavi S, Karimzadeh G. 2010. Karyological and nuclear DNA content variation in some Iranian endemic *Thymus* species (Lamiaceae). *Journal of Agricultural Science and Technology* **12:** 447–458.

Meyer CP, Paulay G. 2005. DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology* **3:** e422.

Molins A, Bacchetta G, Rosato M, Rosselló JA, Mayol M. 2011. Molecular phylogeography of *Thymus herba-barona* (Lamiaceae): insight into the evolutionary history of the flora of the western Mediterranean islands. *Taxon* **60:** 1295–1305.

Morales R. 1995. Híbridos de *Thymus* L. (Labiatae) en la Península Ibérica. *Anales del Jardín Botánico de Madrid* **53:** 199–211.

Morales R. 1997. Synopsis of the genus *Thymus* L. in the Mediterranean area. *Lagascalia* **19:** 249–262.

Morales R. 2002. The history, botany and taxonomy of the genus *Thymus*. In: Stahl-Biskup E, Sáez F, eds. *Thyme. The genus* Thymus. London: Taylor & Francis, 1–43.

Morales R. 2010. Thymus L. In: Castroviejo S, Morales R, Quintanar A, Cabezas F, Pujadas AJ, Cirujano S, eds. *Flora iberica. Plantas vasculares de la Península Ibérica e Islas Baleares, Vol. XII*. Madrid: Consejo Superior de Investigaciones Científicas, Real Jardín Botánico, 349–409.

Morales VR. 1986. Taxonomía de los generos *Thymus* (escluida la sección *Serpyllum*) y *Thymbra* en la peninsula iberica. *Ruizia* **3:** 1–324.

Muellner AN, Schaefer H, Lahaye R. 2011. Evaluation of candidate DNA barcoding loci for economically important timber species of the mahogany family (Meliaceae). *Molecular Ecology Resources* **11:** 450–460.

Newmaster SG, Ragupathy S. 2009. Testing plant barcoding in a sister species complex of pantropical *Acacia* (Mimosoideae, Fabaceae). *Molecular Ecology Resources* **9:** S172–S180.

Newmaster SG, Ragupathy S, Janovec J. 2009. A botanical renaissance: state-of-the-art DNA bar coding facilitates an automated identification technology system for plants. *International Journal of Computer Applications in Technology* **35:** 50–60.

Ovchinnikova A, Krylova E, Gavrilenko T, Smekalova T, Zhuk M, Knapp S, Spooner DM. 2011. Taxonomy of cultivated potatoes (*Solanum* section *Petota*: Solanaceae). *Botanical Journal of the Linnean Society* **165:** 107–155.

Quan J-P, Peng F, He SL, Zheng YH, Jiang YM, Li NW, Zhang MX, Xia B. 2009. Patch diversity and spatial structure of wild *Thymus quinquecostatus*. *Journal of Applied Ecology* **20:** 20–26.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends in Genetics* **16:** 276–277.

Riera J, Guemes J, Rossello JA. 2007. *Thymus richardii* (Lamiaceae) in the Iberian Peninsula. *Flora Montiberica* **37:** 77–80.

Rieseberg LH, Wood TE, Baack E. 2006. The nature of plant species. *Nature* **440:** 524–527.

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.

Seberg O, Petersen G. 2009. How many loci does it take to DNA barcode a crocus. *PLoS ONE* **4:** e4598.

Stahl-Biskup E, Sáez F. 2002. *Thyme*. London: Taylor and Francis.

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

Sun H, McLewin W, Fay MF. 2001. Molecular phylogeny of *Helleborus* (*Ranunculaceae*), with an emphasis on the East Asia–Mediterranean disjunction. *Taxon* **50:** 1001–1018.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28:** 2731–2739.

Tarayre M, Saumitou-Laprade P, Cuguen J, Couvet D, Thompson J. 1997. The spatial genetic structure of cytoplasmic (cpDNA) and nuclear (allozyme) markers within and among populations of the gynodioecious *Thymus vulgaris* (Labiatae) in southern France. *American Journal of Botany* **84:** 1675–1684.

Templeton AR, Crandall KA, Sing CF. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* **132:** 619–633.

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.

Thompson JD. 2002. Population structure and the spatial dynamics of genetic polymorphism in thyme. In: Stahl-Biskup E, Sáez F, eds. *Thyme. The genus* Thymus. London: Taylor & Francis, 44–74.

Vila R. 2002. Flavonoids and further polyphenols in the genus *Thymus*. In: Stahl-Biskup E, Sáez F, eds. *Thyme. The genus* Thymus. London: Taylor & Francis, 330.

Whitlock BA, Hale AM, Groff PA. 2010. Intraspecific inversions pose a challenge for the *trnH-psbA* plant DNA barcode. *PLoS ONE* **5:** e11533.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. Neighbor-joining (NJ) tree based on the *trnH-psbA* dataset generated using MEGA version 5 [Tamura *et al*., 2011 – options: tree inference method, neighbor-joining; phylogeny test and options, bootstrap (100 replicates); gaps/missing data, pairwise deletion; codon positions, 1st + 2nd + 3rd + non-coding; substitution model, Kimura two-parameter (K2P); substitutions to include, transitions + transversions; pattern among lineages, same (homogeneous); rates among sites, uniform rates]. Bootstrap values of < 75% are not shown.
Appendix S1. Voucher and GenBank accession numbers (*rbcL*, *matK*, *trnH-psbA*).