DNA barcoding of a large genus, *Aspalathus* L. (Fabaceae)

Dawn Edwards[1,2], Anne Horn[1,3], Diane Taylor[1], Vincent Savolainen[4,5] & Julie A. Hawkins[1]*

[1] *School of Biological Sciences, Plant Sciences Laboratories, University of Reading, Reading, Berkshire, RG6 6AS, U.K. *j.a.hawkins@reading.ac.uk (author for correspondence)*
[2] *Royal Horticultural Society Garden Wisley, Woking, Surrey, GU23 6QB, U.K.*
[3] *Conservation Ecology and Entomology, University of Stellenbosch, Private Bag X1, Matieland 7602, Republic of South Africa*
[4] *Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3DS, U.K.*
[5] *Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot, Berkshire, SL5 7PY, U.K.*

The DNA barcode potential of three regions (the nuclear ribosomal ITS and the plastid *psbA-trnH* and *trnT-trnL* intergenic spacers) was investigated for the plant genus *Aspalathus* L. (Fabaceae: Crotalarieae). *Aspalathus* is a large genus (278 species) that revealed low levels of DNA sequence variation in phylogenetic studies. In a 51-species dataset for the *psbA-trnH* and ITS regions, 45% and 16% of sequences respectively were identical to the sequence of at least one other species, with two species undiscriminated even when the two regions were combined. In contrast, *trnT-trnL* discriminated between all species in this dataset. In a larger ITS and *trnT-trnL* dataset, including a further 82 species, 7 species in five pairwise comparisons remained undiscriminated when the two regions were combined. Four of the five pairs of species not discriminated by sequence data were readily distinguished using a combination of qualitative and quantitative morphological data. The difficulty of barcoding in this group is increased by the presence of intraspecific variation in all three regions studied. In the case of *psbA-trnH*, three intraspecific samples had a sequence identical to at least one other species. Overall, *psbA-trnH*, currently a candidate for plant barcoding, was the least discriminatory region in our study.

KEYWORDS: *Aspalathus*, DNA barcode, ITS, *psbA-trnH*, species identification, *trnT-trnL*

INTRODUCTION

Since the first calls for the application of DNA sequences as a means of large-scale species identification (Tautz & al., 2002, 2003; Blaxter, 2003; Hebert & al., 2003a; Stoeckle, 2003), and the introduction of the term DNA "barcodes" to describe such sequences (Hebert & al., 2003a), case studies examining the feasibility of universal DNA barcoding in the animal kingdom have dominated the literature. Whilst many animal DNA barcoding studies suggest the mitochondrial cytochrome oxidase subunit I (*coxI*) alone is almost 100% accurate (e.g., Hebert & al., 2003a, 2004b; Barrett & Hebert, 2005; Hogg & Hebert, 2005), the plant barcoding community recognises a need for more than one barcoding region (Chase & al., 2005, 2007; Kress & al., 2005; Cowan & al., 2006; Newmaster & al., 2006; Kress & Erickson, 2007; Taberlet & al., 2007; Lahaye & al., 2008). The choice of barcoding regions has yet to be made, but the following regions have been suggested as candidates: the internal transcribed spacer region, ITS, *rbcL* and another plastid region such as *psbA-trnH*^{GUG} (Chase & al., 2005); ITS & *psbA-trnH* (Kress & al., 2005); *rbcL* (Newmaster & al., 2006); *rpoC1*, *rpoB* & *matK* or *rpoC1*, *matK* & *psbA-trnH* (Chase & al., 2007); *matK*, *atpF-atpH* & *psbA-trnH* or *matK* & *psbK-psbI* (proposed by Ki-Joong Kim and co-workers, University of Korea, Seoul; Pennisi, 2007); *psbA-trnH* & *rbcL* (Kress & Erickson, 2007); *psbA-trnH* (Shaw & al., 2007); *trnL*^{UAA} (Taberlet & al., 2007); *matK* & *psbA-trnH* (Lahaye & al., 2008).

In particular, the plastid intergenic spacer *psbA-trnH* has been highlighted as a prime candidate due to its high level of variation. In the first published flowering plant barcoding study, Kress & al. (2005) found *psbA-trnH* to have the highest mean percentage sequence divergences (1.24%) of the ten plastid regions examined, and based on this result, together with general concordance with their two additional criteria for barcode selection (sequence length and ease of amplification from universal primers), they suggested further testing should be undertaken for these regions. In subsequent publications (Chase & al., 2005, 2007; Kress & Erickson 2007; Shaw & al., 2007; Lahaye & al., 2008) *psbA-trnH* remains a strong candidate locus for plant barcoding. Current research is focussing on plastid regions for barcoding, although Kress & al. (2005) and Kress & Erickson (2007) recovered their highest mean percentage sequence divergence for ITS (2.81% and 5.7% respectively). Though the effectiveness of nuclear regions such as ITS has been questioned in the light of complex evolutionary patterns, some feel it should not be ruled

out (pers. comm., Second International Barcode of Life Conference, Taipei, Taiwan).

An opportunity for testing DNA barcoding in a large plant genus is provided by *Aspalathus* L. (Fabaceae: Crotalarieae). *Aspalathus* comprises 278 species, and is the largest plant genus endemic to South Africa, with 272 species distributed in the Cape Floristic Region, an area of just 90,000 km^2. A comprehensive key is available for *Aspalathus* (Dahlgren, 1988), and species delimitations are presented in an extremely thorough monograph (Dahlgren, 1960, 1961a, b, c, 1962, 1963a, b, c, d, 1965, 1966, 1967a, b, 1968a, b, c, d, 1969, 1971) and Flora treatment (Dahlgren, 1988), in which the genus was divided into 34 infrageneric groups. Nevertheless, the identification process is difficult and time consuming, and since the key is dependent on floral characters, the identification of sterile specimens is in most cases impossible. An electronic multi-access/interactive key could speed the identification process and permit identification from sterile material, even in the field using palm-top computers. Even electronic keys demand some user expertise, however, and this, together with the difficulties with identification through traditional morphological means and the sympatric distribution of a large number of species in a relatively small area, suggests that *Aspalathus* is a group that might also benefit from barcoding. Phylogenetic analysis of the regions tested here (ITS, *psbA-trnH* & *trnT-trnL*^{UAA}) recovered a poorly resolved phylogeny (Edwards, 2006). Reconstruction of a resolved phylogeny for *Aspalathus* may require the use of other regions which are not suitable for universal barcoding, such as the *CYCLOIDEA* nuclear region which has been used for phylogeny reconstruction in another closely related species rich group, the lupins (Hughes & Eastwood, 2006). The failure of the three regions investigated here to recover resolved phylogeny suggests that barcode identification of *Aspalathus* species using these three regions may prove challenging.

Two of the DNA regions used for the phylogeny reconstruction of *Aspalathus* (the nuclear ITS and plastid *psbA-trnH*) have been proposed as barcoding regions for plants (Chase & al., 2005; Kress & al., 2005; Shaw & al., 2007; Lahaye & al., 2008). The third region, the plastid *trnT-trnL*, was found to have a more than 95% probability of identifying the correct species of *Sinningia* s.l. (Gesneriaceae) in the barcoding study of Cowan & al., (2006) and also generally shows good concordance with the three criteria as proposed by Kress & al. (2005): *(1)* The region has been employed in a number of species-level studies and is often reported as amongst the most variable chloroplast regions (Böhle & al., 1994; Cronn & al., 2002; Shaw & al., 2005). *(2)* Although universal primers for the region were published alongside those for the *trnL-trnF* region (Taberlet & al., 1991), *trnT-trnL* has not been as widely employed in phylogenetic studies as the *trnL-trnF* region. Shaw & al. (2005) suggested this might be due to PCR amplification problems that had been encountered in a number of groups when using the universal primer A from the study of Taberlet & al. (1991), but reported that a new primer (Cronn & al., 2002) worked in all of the widely sampled taxa surveyed. *(3)* In a departure from one of the criteria of Kress & al. (2005), published *trnT-trnL* sequences range from ~400–1,500 bp (Shaw & al., 2005). Nevertheless, Kress & al. (2005) selected the *psbA-trnH* region as their most potential barcoding chloroplast region even though 41% of the *psbA-trnH* sequences in their study were outside their 300–800 bp criterion, with minimum and maximum outliers of 119 and 1,094 bp respectively. Thus, further investigation of the *trnT-trnL* region appears warranted.

The successful application of DNA barcoding for species identification depends on a threshold of sequence divergence above which species can be distinguished and below which intraspecific variation is encountered. In the plant kingdom, Cowan & al. (2006) reported that for 33 species, two to five accessions per species revealed no or very low levels of nucleic acid substitution intraspecific variation for *psbA-trnH* (0.0–0.8%). Kress & Erickson (2007) and Lahaye & al. (2008) also reported lower levels of genetic divergence within species than between species. Baldwin & al. (1995) showed low intraspecific variation in ITS, but see Feliner & al. (2004), Rauscher & al. (2004), Noyes (2006), and Koch & Matschinger (2007), who found remarkable amounts of intraspecific variation in ITS. Multiple sampling of *Aspalathus* species for the three targeted DNA regions will permit assessment of levels of intraspecific divergence for each region, and hence, their utility in barcoding.

MATERIALS AND METHODS

Sampling and molecular methods. — Total genomic DNA was isolated from field-collected leaf material preserved in silica-gel (Chase & Hills, 1991) for 141 specimens representing 133 *Aspalathus* taxa (and 30 of the 34 *Aspalathus* infrageneric groups, Dahlgren, 1988) using a small-scale hot CTAB DNA extraction protocol (Harris, 1995), a modification of the method of Doyle & Doyle (1987). The following amendments were made to the protocol of Harris (1995): the leaf material, sealed in an Eppendorf tube, was initially cooled in liquid nitrogen; a few grains of sand were added to the leaf material to aid grinding; and 0.2% of β-mercaptoethanol was added to the CTAB extraction buffer at the time of extraction. For troublesome extractions, 1% polyvinyl polypyrrolidine (PVPP) was added to the stock CTAB extraction buffer and samples were left overnight at –20°C at the precipitation stage.

PCR amplifications were performed in 50 µl reactions comprising: 1–5 µl of undiluted DNA; Reaction Buffer

×1 (Opti Buffer, 10X, Bioline); $MgCl_2$ 2 mM; dNTP mix 0.2 mM; 0.35 µM of each primer; 10 µg of bovine serum albumin (BSA, New England Biolabs); and 2 units of *Taq* DNA polymerase (Bioline, London, U.K.). For a few amplicons that were troublesome to optimise for the ITS region, 2.5 mM $MgCl_2$ ReddyMix™ PCR Master Mix (Abgene, Epsom, Surrey, U.K.) was used. The primers and PCR profiles used are listed in Table 1. PCRs were run on a GeneAmp® PCR system 2700 machine (Applied Biosystems). A Qiaquick® PCR Purification kit (Qiagen Ltd, Dorking, Surrey, U.K.) was used to clean the resultant PCR products, eluting them into 30 µl of EB buffer. Sequencing was carried out using ABI PRISM® BigDye™ Terminator 100 Reaction Ready kit (Applied Biosystems). The cycle-sequencing products were purified using PERFORMA® DTR V3 96-well short plates (Edge Biosystems, U.S.A.), and separated on an ABI PRISM® 3100 automated capillary sequencer (DNA sequencing service, School of Biological Sciences, University of Reading). Accession details for all taxa, indicating taxonomic position according to Dahlgren (1988), are presented in the Appendix in Taxon online issue.

Alignment and barcoding analyses. — Raw sequence data were assembled and edited using SeqMan™ II, one of the programmes of the Lasergene® software package (DNASTAR, Inc.), and verified by a Blastn search on GenBank. The edited sequences were aligned using the multiple alignment Clustal W algorithm (Thompson & al., 1994) as implemented in MegAlign™ (Lasergene®, DNASTAR, Inc.), with further adjustment by eye using MacClade v. 4.0 (Maddison & Maddison, 2001). All variable sites were rechecked to the original trace files. The alignments are available on TreeBase (Piel & al., 2002).

Study 1. – Alignments were generated for sequences of 51/278 (18%) *Aspalathus* species for each of the three DNA regions (ITS, *psbA-trnH* and *trnT-trnL*), to investigate the interspecific discriminatory power of the individual regions. Combined alignments were assembled to assess the discriminatory power of the three pairwise combinations and all three regions combined. A total of 22 of 34 infrageneric groups (Dahlgren, 1988) were represented, with more than one species sampled in eleven of the groups.

Study 2. – In a second examination of interspecific variation for ITS and *trnT-trnL*, individual and combined alignments were prepared for 133/278 (48%) *Aspalathus* species. In these two larger datasets, 30 of the infrageneric groups were represented, with more than one species sampled for 24 of these groups.

Study 3. – To provide a preliminary examination of levels of intraspecific DNA divergence in *Aspalathus*, a second taxon was sequenced for each of the three regions for three species and for a further five species for ITS and *trnT-trnL* only. Pairwise alignments were generated for each pair of conspecific sequences.

For studies 1 and 2, pairwise comparisons for each species pair in the MacClade multiple sequence alignments were made by calculating sequence divergence levels using the total character difference function in PAUP* v. 4.0b4a (Swofford, 2002). A method was devised to ensure that information from indels, as well as nucleotide substitutions was utilised. Dashes (–) in an alignment are treated as missing data in pairwise comparisons in PAUP*, and are not scored, so that species pairs differing only in the presence or absence of an indel are incorrectly identified as identical. To ensure indel information was utilised, each position in an indel was recoded to a number 1, whilst retaining terminal gaps as missing data. Unlike a dash, a 1 is treated as a fifth character state, and included in pairwise comparison calculations. For each region and combination of regions for studies 1 and 2 the number of species pairs which were not discriminated (i.e., were identical), was recorded. In study 3, pairwise comparisons

Table 1. PCR primers and profiles.

Region	Primer sequence (5′ to 3′)		Reference
ITS	17SE:	ACG AAT TCA TGG TCC GGT GAA GTG TTC G	Sun & al. (1994)
	26SE:	ATG AAT TCC CCG GTT CGC TCG CCG TTA C	
psbA-trnH	*psb* B:	GTT TAC TTT TGG GCA TGC TTC G	Hamilton (1999)
	psb F:	CGC AGT TCG TCT TGG ACC AG	
trnT-trnL	A:	CAT TAC AAA TGC GAT GCT CT	Taberlet & al. (1991)
	B:	TCT ACC GAT TTC GCC ATA TC	

Region	PCR profile					
	Initial denaturation temp./time	Denaturation temp./time	Annealing temp./time	Extension temp./time	Final extension temp./time	No. of cycles
ITS	97°C/1 min	97°C/1 min	52°C/1 min	72°C/2 min	72°C/7 min	30
psbA-trnH	96°C/5 min	96°C/45 s	53°C/1 min	72°C/30 s	72°C/7 min	35
trnT-trnL	94°C/5 min	94°C/1 min	56°C/1 min	72°C/1 min	72°C/7 min	25–30

were calculated for each conspecific taxon pair, and numbers of taxon pairs showing divergence were recorded. Where the additional sequences showed intraspecific variation, they were added to the appropriate single region alignments of studies 1 and 2 and interspecific divergence levels calculated. For each region, intraspecific variants that were identical to other species were recorded.

Morphological studies. — To test the species delimitations used by the author of the most complete treatment of the genus (Dahlgren, 1988), we examined herbarium specimens representing five species pairs that were not distinguished by *trnT-trnL* and ITS in concert (see below). Herbarium specimens examined included material determined by Dahlgren, as well as field collected material which we identified using the key presented in his treatment and checked for overall similarity against determined specimens. The published descriptions (Dahlgren, 1960, 1961a, b, c, 1962, 1963a, b, c, d, 1965, 1966, 1967a, b, 1968a, b, c, d, 1969, 1971, 1988) were examined to identify potentially diagnostic quantitative and qualitative characters. Qualitative characters were presence/absence characters which were not measurable; we were careful not to erroneously consider continuous variables as qualitative where botanical terminology obscures potentially overlapping quantitative variation (Stevens, 1991). Qualitative characters were scored from all specimens. Wherever quantitative data could be collected we made six measurements for each of the vegetative and calyx characters and three for the petal characters from each sheet. We considered quantitative characters as diagnostic if there were absolute gaps in the distribution of variation i.e., ranges were non-overlapping (Theile, 1993). Morphological characters which showed qualitative or diagnostic quantitative differences across all specimens for the species pairs examined were tabulated.

RESULTS

Interspecific variation. — Figure 1 shows the number of species pairs with identical sequences for each region and combination of regions for studies 1 and 2. In study 1, a total of 64 species pairs, representing 23 out of 51 species, were not discriminated in the *psbA-trnH* dataset, each of these species having a sequence identical to at least one other species. For ITS, the sequences of seven species pairs were identical representing eight species that could not be discriminated from at least one other species. Even when combined, *psbA-trnH* and ITS failed to separate one species pair. In contrast, the *trnT-trnL* region distinguished all species pairs by the presence of at least one base pair difference (including indels). However, in study 2, which employed a denser taxon sampling, both of the individual 133 species datasets for ITS and *trnT-trnL*

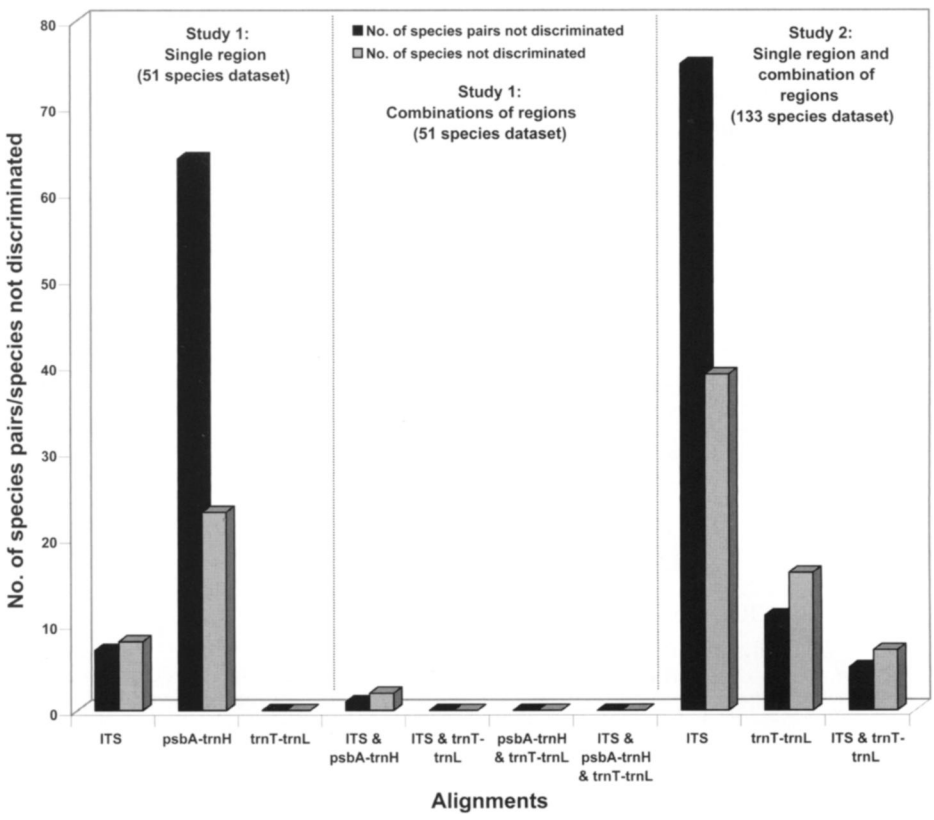

Fig. 1. Graph showing the number of species pairs with identical sequences for each region and combination of regions for studies 1 and 2.

and a combination of the two datasets included species pairs which could not be discriminated (ITS, 75 species pairs representing 39 species; *trnT-trnL*, 11 species pairs representing 16 species; ITS & *trnT-trnL*, 5 species pairs representing 7 species). Of the three regions studied, *trnT-trnL* was the most indel-rich region. In the 51 species dataset, indels were present for at least one species in 33% of base pair positions across the alignment, compared with 1% in ITS and 9% in *psbA-trnH*.

Intraspecific variation. — As is shown in Table 2, all three regions displayed intraspecific variation (ITS 2/8 species; *psbA-trnH* 3/3 species; and *trnT-trnL* 6/8 species). In the case of *psbA-trnH*, one of the intraspecific variants of each species sampled was identical to at least one other *Aspalathus* species, and for ITS one of the variants of one species showed identity to at least one other species. All of the intraspecific variants recovered for the *trnT-trnL* region were distinct from other species.

Morphological studies. — Qualitative and quantitative differences between species pairs are summarised in Table 3. Four of the five species pairs could be reliably distinguished.

DISCUSSION

Interspecific variation. — The strictest test of the utility of a DNA region for barcoding is whether discrimination is possible between all species pairs within the study group, in other words that there are no identical sequences. None of the three DNA regions tested (the nuclear ITS, and the plastid *psbA-trnH* and *trnT-trnL*) is individually able to discriminate all species pairs tested in *Aspalathus,* and sometimes may not even distinguish between species from different infrageneric groups sensu Dahlgren (1988): ITS 26/39, *psbA-trnH* 18/23, and *trnT-trnL* 5/16 undiscriminated species were identical to species across infrageneric boundaries. These results are in contrast to previously published studies examining the feasibility of DNA barcoding, which report 100% species discrimination from just one DNA region: *cox1* in animals (e.g., Hebert & al., 2003a, 2004b; Hogg & Hebert, 2004; Barrett & Hebert, 2005, although see Hebert & al., 2003b, Meyer & Paulay, 2005). The first studies to densely sample plant genera are those reported in Chase & al. (2005) and Cowan & al. (2006). Chase & al. (2005) report on surveys of 170 species of the

Table 2. Intraspecific results showing the numbers of species with intraspecific divergence and the numbers of species with at least one intraspecific variant identical to at least one other species.

3-region study	ITS	*psbA-trnH*	*trnT-trnL*
Number of species showing intraspecific divergence	1/3	3/3	2/3
Number of species with at least one intraspecific variant identical to at least one other species	0/3	3/3	0/3
2-region study			
Number of species showing intraspecific divergence	1/5	N/A	4/5
Number of species with at least one intraspecific variant identical to at least one other species	1/5	N/A	0/5

N/A indicates that *psbA-trnH* was not sampled for this study.

Table 3. Morphological qualitative and quantitative differences between species pairs.

Species 1	Species 2	Diagnostic qualitative characters	Diagnostic quantitative characters
A. barbata	*A. cordata*	Leaf margin ciliate/not	Keel length; ratio of leaf width to length
A. acifera	*A. corrudifolia*	Leaf tip pungent/not; calyx lobes pungent/not	Standard blade length; wing blade length; keel blade length
A. chenopoda	*A. incurva*	Leaf tip pungent/not; inflorescences terminal/on lateral short shoots; inflorescences unifloral or bifloral/3 or more flowers; bract tips pungent/not; calyx lobes pungent/not	Standard blade length; wing blade length; keel blade length
A. chenopoda	*A. spicata*	Leaf tip pungent/not; inflorescences terminal/on lateral short shoots; inflorescences unifloral or bifloral/3 or more flowers; bract tips pungent/not; calyx lobes pungent/not	None
A. spicata	*A. incurva*	None	None

250 species of *Moraea* (Iridaceae) and 88 taxa representing the ca. 85 species of *Protea* (Proteaceae) respectively. Cowan & al. (2006) used a complete species sampling for the 96 species of *Sinningia* s.l. (Gesneriaceae). These surveys showed more than 95% probability of identifying the correct species using *trnL-trnF*, *rbcL* & *rps16* for *Moraea*; *rps16* & *ncpGS* for *Protea*; and *trnS-trnG*, *trnT-trnL*, *rpl16*, *trnL-trnF*, *atpB-rbcL* & *ncpGS* for *Sinningia*). The lowest discrimination for *Protea* was using the *trnL-trnF* region (60%) and the *atpB-rbcL* (85%). Neither of these poorly performing regions has been proposed as candidate barcode regions. In contrast, in our study we find lower levels of discrimination (55% for *psbA-trnH* and 84% for ITS), although these, in particular *psbA-trnH*, have been advocated as barcode regions of choice (Chase & al., 2005, 2007; Kress & al., 2005; Kress & Erickson, 2007; Shaw & al., 2007; Lahaye & al., 2008).

The championing of Kress & al. (2005) of ITS as one of the regions suitable for barcoding in plants finds support here. The region generally satisfied all three of their criteria: ITS was the second most useful region under consideration, discriminating the second highest number of *Aspalathus* species; all taxa were successfully sequenced using universal primers (Sun & al., 1994), and further universal primers, nested within those used here, are also available for the region (White & al., 1990) and the length of the sequences, ranging from 698 to 829 bp, more or less conforms to the proposed length of 300–800 bp of Kress & al. (2005). ITS, nevertheless, is not a panacea for plant barcoding studies. Kress & Erickson (2007) showed that universal primers failed in gymnosperms, ferns and mosses. Although the most commonly used DNA region for species-level phylogeny reconstruction (Alvarez & Wendel, 2003), ITS has not always displayed sufficient variability to resolve species-level (Richardson & al., 2001; Edwards, 2006), or even higher-level relationships (Klak & al., 2004), and as been shown here, may not permit total species discrimination in groups with low DNA variation. There is also the undesirable potential of ITS to comprise paralogous sequences from processes such as gene duplication and incomplete concerted evolution (Alvarez & Wendel, 2003). In some cases, secondary peaks resulting from sequencing of paralogous copies can be so severe as to render some groups of plants unsequenceable without recourse to cloning. These are not universal problems—they were not apparent in our study, and there are other taxa where ITS identifications have not been problematical (e.g., Asteraceae tribes Lactuceae and Anthemideae; Gemeinholzer & al., 2006). In our case ITS presented no technical problems and was considerably more informative than *psbA-trnH*.

For the chloroplast *psbA-trnH* region, in concordance with the criteria of Kress & al. (2005), all *Aspalathus* taxa were successfully sequenced using universal primers (Hamilton, 1999). The range of sequence lengths was from 271–331 bp, with only eight sequences falling below the suggested minimum length of 300 bp of Kress & al. (2005). However, although the region has revealed sufficient variation to be successfully employed in species-level phylogenies (Sang & al., 1997; Chandler & al., 2001), and, with the inclusion of indel information, successfully discriminated all species pairs in the eight genera intrageneric examination of Kress & al. (2005), and showed the highest plastid region mean percent sequence divergence in the studies of Kress & Erickson (2007) and Lahaye & al. (2008), *psbA-trnH* showed the lowest discriminatory power of the three regions examined in this study, with 45% of species undiscriminated. This result in *Aspalathus* agrees with a three species survey of five chloroplast regions conducted during a phylogenetic study of *Aspalathus* (Edwards, 2006), which revealed *psbA-trnH* to have the lowest phylogenetic informativeness: *psbA-trnH* (using the primers from Hamilton, 1999), 9 potentially informative characters; *psbB-psbF* (using the primers from Hamilton, 1999), 11; *accD-psaI* (using the primers from Small & al., 1998), 17; *atpB-rbcL* (using the primers from Hoot & al., 1995), 18; and *trnT-trnL* (using the primers from Taberlet & al., 1991), 25, where potentially informative characters were calculated from numbers of nucleotide substitutions and indels following Shaw & al. (2005).

In contrast to *psbA-trnH*, *trnT-trnL* was revealed as a tractable candidate for the plastid region, displaying the highest discriminatory power of all three regions examined in *Aspalathus*. In further support of the utility of *trnT-trnL*, all *Aspalathus* taxa were sequenced using the universal primers from the study of Taberlet & al. (1991), and the lengths of the sequences ranged from 694 to 792 bp. The recent reports of high levels of variability in *trnT-trnL* (Small & al., 1998; Shaw & al., 2005), together with the development of a new primer (Cronn & al., 2002) to replace the A primer shown to be problematic by Taberlet & al. (1991), may lead to an increase in popularity for the application of *trnT-trnL* in species-level phylogenetic reconstruction. An increase of the generation of sequences for the region would further support a proposal for *trnT-trnL* to be considered as a potential DNA barcode in flowering plants.

Aspalathus, as suspected, due to the low levels of DNA sequence variation recovered for phylogeny reconstruction (Edwards, 2006), has proved a challenging barcoding test case. The molecular data's relatively low discriminatory power could be due to over-zealous taxonomic splitting. However, our morphological survey shows that four out of five species pairs with identical sequences can be readily distinguished using morphological characters. The formalisation of a traditional morphological species concept is sometimes attributed to Van Steenis (1957) who noted that "sets of characters (differences) occur in nature

discontinuously in combinations, facilitating the distinction of groups of specimens, species or other taxa characterised by these sets of characters". Operationally, Nixon & Wheeler (1990) appeal to a comparable procedure for species delimitation. Their phylogenetic species concept is also based on observations of pattern, makes no reference to ability to interbreed and does not require any analysis, phenetic or cladistic. The phylogenetic species sensu Nixon & Wheeler (1990) are the smallest aggregate of populations or lineages diagnosable by a unique combination of character states in comparable individuals. Thus for Nixon & Wheeler (1990), a single character difference is sufficient for species delimitation. Van Steenis (1957) argued that if the contrasted taxa only differ in one character then the separation will mostly prove to be artificial, and that good species are generally also characterised by additional ecological features. Four of the five species pairs we consider here are separate species sensu Van Steenis (1957).

For some authors, a phylogenetic species concept sensu Nixon & Wheeler (1990) results in the recognition of too many species, since it has been shown that the concept may lead to recognition of a far greater number of much less inclusive units. A review by Agapow & al. (2004) showed a 48% increase in species numbers when non-PSC species were revised using the PSC. In our study we found a trend in the opposite direction. Under the PSC fewer *Aspalathus* species were recognised than had been accepted by Dahlgren (1988). This suggests that Dahlgren was using a species concept which results in the recognition of notably large numbers of species. In this case, it might be argued that current taxonomy is oversplit. However, many in workers in South Africa have found Dahlgren's species to be robust, reliably identifiable, geographically coherent and valuable units for conservation planning and evolutionary analysis (e.g., Oliver & al., 1983; Edwards, 2006; Edwards & Hawkins, 2007; Forest & al., in prep.). There has been a debate about whether barcoding differences might be a useful in species delimitation (e.g., Hebert & al., 2004a; Hebert & Gregory, 2005; Meyer & Paulay, 2005; Will & al., 2005). Our morphological survey shows that barcodes have failed to discriminate species recognised in the most recent and thorough monograph, but which are not species sensu Nixon & Wheeler (1990). This is the case for one of our species pairs, *A. spicata* and *A. incurva*. These species have been described based on combinations of a number of character states with overlapping ranges (i.e., a non-explicit polythetic species concept). Further study using genetic markers, a more comprehensive morphometric dataset and consideration of distribution ranges are needed to provide insights into the status of these species.

The role of indels. — The contribution indels can make to barcoding has been highlighted here. The indel-rich *trnT-trnL* region proved to be the most useful region in the discrimination of *Aspalathus* species, in support of the report of Kress & al. (2005) that indels in the *psbA-trnH* region permitted discrimination of otherwise identical species pairs in *Solidago*, and their suggestion that indels will most likely improve resolving power in barcoding. However, if a criterion for the selection of DNA barcoding regions is that sequences are also to be used in phylogeny reconstruction (Seberg & al., 2003; Blaxter, 2004), then difficult to align indel-rich regions might be discounted. For the *Aspalathus* phylogeny reconstruction (Edwards, 2006), a number of poly A/T-rich indel regions were excluded from *trnT-trnL* phylogenetic analyses due to alignment difficulties. Nevertheless, even after the exclusion of these regions, *trnT-trnL* had a higher number of parsimony informative characters than ITS (98 parsimony informative characters for *trnT-trnL* and 91 for ITS in a 145 taxon dataset).

There is a large literature which describes the best use of indel regions in phylogeny reconstruction (Graham & al., 2000; Simmons & Ochoterena, 2000; Müller, 2006). This literature emphasises the number of presumed evolutionary events as informative when seeking to reconstruct phylogeny, and considers an entire gapped region in an alignment as a single binary character, not a collection of characters in which the number of differences depends on the total length of the indel, with the caveat that indels of different length in the same region of DNA necessarily represent more than one event. In contrast, in the case of barcoding, the absolute numbers of base pair differences are significant. In the case of barcoding, for example, confidence in an identification will be greater when a pair of species differ in length by two base pairs than when they differ by one, whether or not those indel differences are adjacent and the result of one or two events in evolutionary time. Species identification using a DNA barcode database should not be hampered by the use of an indel containing region. Indeed, the incorporation of indel rich regions could increase the discriminatory power of barcode regions (e.g., Kress & al., 2005; Cowan & al., 2006). However, problems arise when attempting to assess the potential utility of indel-rich barcoding regions.

The multiple sequence alignment comparison method used here recognises both nucleotide substitution and indel information in determining whether species pairs can be discriminated. Table 4 shows that using multiple sequence alignment might not accurately determine numbers of differences between species pairs when a sequence contains an indel that is ambiguously aligned. Although the multiple sequence alignment shown is reasonable when species 1 and 3 are considered as a single pair, an alternative pairwise sequence alignment is preferred which reduces the number of differences recorded for that species pair from two (a point mutation and a single base pair indel difference) to one (a single base pair indel difference). Potentially, all species pair differences above zero in multiple sequence alignments could be affected

by ambiguously aligned indels. For this reason we prefer to determine whether sequences differ or not, rather than to calculate the number of point mutation and indel differences by which species pairs are distinguished.

Whilst the strictest test was employed in this study (i.e., the determination of whether each region can discriminate all species pairs by at least one base pair difference, including indels), for the successful implementation of identification using DNA barcodes, it may be desirable to determine levels of divergence for selected regions in order to identify thresholds by which inter- and intraspecific variation can be determined (Hebert & al., 2004b). For accurate calculation of divergence levels between large numbers of species for indel containing regions, either the alignment must contain no ambiguously aligned indels, or pairwise comparisons must be made for all possible two-taxon alignments. In this study 18,831 two-taxon alignments and pairwise comparisons would have to be prepared for the single region datasets of studies 1 and 2 alone. This is clearly not a realistic proposition, and there is a need for an algorithm to calculate these comparisons in a more tractable way, so that the value of indel containing regions can be fully examined.

Intraspecific variation. — The intraspecific study conducted here has further highlighted the low utility of the *psbA-trnH* region as a DNA barcode for identification of *Aspalathus* species. All three of the pairwise intraspecific comparisons sampled for the region showed divergence. More problematically, in every case, one of the two intraspecific variants was identical to at least one other species. For example, the *Aspalathus costulata* pair displayed one base pair difference and the *A. crenata* pair four differences, with one of the variants in each case identical to another species, specifically a species from another infrageneric group. Similarly, *A. spinosa* subsp. *spinosa* showed five intraspecific differences in *psbA-trnH* between the two individuals sampled, but in this case one of the variants was not discriminated from eleven other species from eight different infrageneric groups.

The ITS region posed the same problems. Two intraspecific taxon pairs showed divergence, with one of the intraspecific variants of one of the species, *A. setacea*, identical to eight other species spread across three infrageneric groups. In contrast, although six of the eight species pairs sampled for *trnT-trnL* displayed intraspecific variation, all variants were distinct from other species.

These results can be compared to the reports in Cowan & al. (2006) of low intraspecific sequence variation. Cowan & al. (2006) do not report the distributions of the samples, although the phenomenon of intraspecific variation in space is well documented (e.g., Vargas, 2003; Lorenz-Lemke & al., 2005; Lihová & al., 2006 for the ITS region). Documented intraspecific variation in these and other studies, alongside our results, suggests a need for a more critical evaluation of the distribution of intraspecific samples. It is notable in that regard that all but one of the species pairs displaying intraspecific variation in our study are widely separated. Just for one species pair, *A. lebeckioides*, were the samples collected within a very short distance of each other. The other pairs were sampled from separate mountain ranges, were separated from each other by mountains, or in the case of the two samples of *A. linearis*, were collected from far ends of the Cedarberg mountain range.

Table 4. Simulated alignments showing how an ambiguously aligned indel, although optimally aligned in a multiple sequence alignment, may align differently in a pairwise alignment and give different numbers of differences between species. In the case of the multiple alignment there are two differences between species 1 and species 3, one is an indel mutation at position 9 (boxed) and the second is a point mutation at position 10 (in bold). However, when this species pair is aligned without reference to other species only one difference between species 1 and species 3 is recovered, this is an indel at position 10 (boxed).

Simulated multiple sequence alignment with indels treated as a 5th character state																								
Alignm. position no. →										1	1	1	1	1	1	1	1	1	1	2	2	2	**No. of differences between species 1 & 3**
Species ↓	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**0**	**1**	**2**	
Species 1	–	–	–	G	A	T	T	C	T	**A**	G	C	T	A	G	C	G	G	C	C	T	A	
Species 2	C	A	T	T	A	T	T	C	A	T	1	1	T	A	G	C	G	T	T	C	–	–	
Species 3	C	A	T	G	A	T	T	C	[1]	**T**	G	C	T	A	G	C	G	G	C	C	–	–	2
Species 4	C	A	T	T	A	T	T	C	A	T	1	1	T	A	G	C	G	T	T	C	–	–	
Species 5	A	A	T	G	A	C	T	C	A	T	G	C	T	A	G	C	G	G	C	C	–	–	
Simulated pairwise comparison between species 1 & 3 with indels treated as a 5th character state																								
Species 1	–	–	–	G	A	T	T	C	T	A	G	C	T	A	G	C	G	G	C	C	T	A	1
Species 3	C	A	T	G	A	T	T	C	T	[1]	G	C	T	A	G	C	G	G	C	C	–	–	

CONCLUSIONS

Our study suggests that at least three regions will be necessary for species discrimination in *Aspalathus*. A decrease has already been observed in the discriminatory power of ITS and *trnT-trnL* with increased sampling. Sampling was broad across the genus and the four unrepresented infrageneric groups are monotypic; thus, further sampling would mostly be gap-filling of species closely related to those already sampled, and failure to discriminate between species is predicted to become more problematic. For successful discrimination of all species within *Aspalathus* it is likely therefore that at least one more region with a greater level of variation than *psbA-trnH* will be required if ITS and *trnT-trnL* remain the most appropriate regions.

As suggested by other authors (Chase & al., 2005, 2007; Kress & Erickson 2007; Shaw & al., 2007; Lahaye & al., 2008), this study highlights that complete species identification for all plants from a small suite of just two or three universal barcoding regions may remain elusive. At least some plant genera will require additional barcode regions to those proposed as universal to achieve complete species identification, whether to address issues of complex evolution, technical issues regarding successful amplification, or as in the case of *Aspalathus*, low DNA sequence variation at species level, coupled with examples of high intraspecific variation, most notably in the region *psbA-trnH*, the most popular choice for inclusion as a universal plant barcoding region. The low level of barcoding success in *Aspalathus* may be uncommon, but further densely sampled studies are needed, including more case studies with 100% species sampling, to explore more fully the performance of the proposed universal regions and the extent to which groups would need specific case-by case investigation.

ACKNOWLEDGEMENTS

We are grateful to Nicola Bergh, Charlie Boucher, Leanne Dreyer, Félix Forest, Chris Hardy, Terry Hedderson, Amida Johns, Mark Johns, Peter Linder, Adrian Möhl, Philip Moline, A. Muthama Muasya, Tracey Nowell, Tony Rebelo, Mathia Schwegler, Terry Trinder-Smith, Timo van der Niet, Tony Verboom and Christopher Whitehouse for fieldwork assistance. We would also like to thank Terry Trinder-Smith for use of the Bolus Herbarium, Cape Town, South Africa, Chris Cupido at the Compton Herbarium, Cape Town, South Africa and Gwilym Lewis at the Royal Botanic Gardens, Kew, U.K. Financial support to Dawn Edwards was provided by the University of Reading and the British Council. Collecting in the Western Cape was carried out under the permit 22/1998 held by Julie A. Hawkins and 77/2001 held by Dawn Edwards.

LITERATURE CITED

Agapow, P.-M., Bininda-Emonds, O.R.P., Crandall, K.A., Gittleman, J.L., Mace, G.M., Marshall, J.C. & Purvis, A. 2004. The impact of species concept of biodiversity studies. *Quart. Rev. Biol.* 79: 161–179.

Alvarez, I. & Wendell, J.F. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molec. Phylog. Evol.* 29: 417–434.

Baldwin, B.G., Sanderson, M.J., Porter, J.M., Wojciechowski, M.F., Campbell, C.S. & Donoghue, M.J. 1995. The ITS region of nuclear ribosomal DNA—a valuable source of evidence on angiosperm phylogeny. *Ann. Missouri Bot. Gard.* 82: 247–277.

Barrett, R.D.H. & Hebert, P.D.N. 2005. Identifying spiders through DNA barcodes. *Canad. J. Zool.* 83: 481–491.

Blaxter, M. 2003. Molecular systematics—counting angels with DNA. *Nature* 421: 122–124.

Blaxter, M.L. 2004. The promise of a DNA taxonomy. *Philos. Trans., Ser. B.* 359: 669–679.

Böhle, U.-R., Hilger, H., Cerff, R. & Martin, W.F. 1994. Noncoding chloroplast DNA for plant molecular systematics at the infrageneric level. Pp. 391–403 in: Schierwater, B., Streit, B., Wagner, G.P. & DeSalle, R. (eds.), *Molecular Ecology and Evolution: Approaches and Applications*. Birkhäuser, Basel.

Chandler, G.T., Bayer, R.J. & Crisp, M.D. 2001. A molecular phylogeny of the endemic Australian genus *Gastrolobium* (Fabaceae: Mirbelieae) and allied genera using chloroplast and nuclear markers. *Amer. J. Bot.* 88: 1675–1687.

Chase, M.W., Cowan, R.S., Hollingsworth, P.M., Van den Berg, C., Madriñán, S., Petersen, G., Seberg, O., Jørgensen, T., Cameron, K.M., Carine, M., Pedersen, G., Hedderson, T.A.J., Conrad, F., Salazar, G.A., Richardson, J.E., Hollingsworth, M.L., Barraclough, T.G., Kelly, L. & Wilkinson, M. 2007. A proposal for a standardised protocol to barcode all land plants. *Taxon* 56: 295–299.

Chase, M.W. & Hills, H.H. 1991. Silica gel—an ideal material for field preservation of leaf samples for DNA studies. *Taxon* 40: 215–220.

Chase, M.W., Salamin, N., Wilkinson, M., Dunwell, J.M., Kesanakurthi, R.P., Haidar, N. & Savolainen, V. 2005. Land plants and DNA barcodes: short-term and long-term goals. *Philos. Trans., Ser. B.* 360: 1889–1895.

Cowan, R.S., Chase, M.W., Kress, W.J. & Savolainen V. 2006. 300,000 species to identify: problems, progress, and prospects in DNA barcoding of land plants. *Taxon* 55: 611–616.

Cronn, R.C., Small, R.L., Haselkorn, T. & Wendel, J.F. 2002. Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. *Amer. J. Bot.* 89: 707–725.

Dahlgren, R. 1960. Revision of the genus *Aspalathus*, I. The species with flat leaflets. *Opera Bot.* 4: 1–393.

Dahlgren, R. 1961a. Additions to a revision of the *Aspalathus* species with flat leaflets. *Bot. Not.* 114: 313–321.

Dahlgren, R. 1961b. Revision of the genus *Aspalathus*, II. The species with ericoid and pinoid leaflets, 1. The *Aspalathus nigra* group. *Opera Bot.* 6: 1–69.

Dahlgren, R. 1961c. Revision of the genus *Aspalathus*, II. The species with ericoid and pinoid leaflets, 2. The *Aspalathus triquetra* group. *Opera Bot.* 6: 70–120.

Dahlgren, R. 1962. Some new species of *Aspalathus* from the Cape Province. *Bot. Not.* 115: 465–474.

Dahlgren, R. 1963a. The genus *Borbonia* L. incorporated in *Aspalathus* L. *Bot. Not.* 116: 185–192.

Dahlgren, R. 1963b. Revision of the genus *Aspalathus*, II. The species with ericoid and pinoid leaflets, 3. The *Aspalathus ciliaris* group and some related groups. *Opera Bot.* 8: 1–183.

Dahlgren, R. 1963c. Studies on *Aspalathus* and some related genera in South Africa., *Opera Bot.* 9: 1–301.

Dahlgren, R. 1963d. Studies on *Aspalathus*. Phytogeographical aspects. *Bot. Not.* 116: 505–534.

Dahlgren, R. 1965. Revision of the genus *Aspalathus*, II. The species with ericoid and pinoid leaflets, 4. The *Aspalathus ericifolia, parviflora, calcarata, desertorum, micrantha, pinea, rostrata, filicaulis, laricifolia*, and *longifolia* groups. *Opera Bot.* 10: 1–231.

Dahlgren, R. 1966. Revision of the genus *Aspalathus*, II. The species with ericoid and pinoid leaflets, 5. The *Aspalathus carnosa, aciphylla, pachyloba, arida, pinguis, spinosa*, and *sanguinea* groups and some other groups. *Opera Bot.* 11: 1–266.

Dahlgren, R. 1967a. Chromosome numbers in some South African genera of the tribe Genisteae s. lat. (Leguminosae). *Bot. Not.* 120: 149–160.

Dahlgren, R. 1967b. Some new and rediscovered species of *Aspalathus* (Leguminosae). *Bot. Not.* 120: 26–40.

Dahlgren, R. 1968a. Distribution and substrate in the South African genus *Aspalathus* L. (Leguminosae). *Bot. Not.* 121: 505–534.

Dahlgren, R. 1968b. Revision of the genus *Aspalathus*, II. The species with ericoid and pinoid leaflets, 6. The *Aspalathus frankenioides, nivea, juniperina, rubens*, and *divaricata* groups and some other groups. *Opera Bot.* 21: 1–309.

Dahlgren, R. 1968c. Revision of the genus *Aspalathus*, II. The species with ericoid and pinoid leaflets, 7. Subgenus *Nortieria*. With remarks on Rooibos tea cultivation. *Bot. Not.* 121: 165–208.

Dahlgren, R. 1968d. Revision of the genus *Aspalathus*, III. The species with flat and simple leaves. *Opera Bot.* 22: 1–125.

Dahlgren, R. 1969. Comprehensive key to the species of *Aspalathus* (Leguminosae). *Bot. Not.* 122: 512–548.

Dahlgren, R. 1971. Chromosome numbers in the South African genus *Aspalathus* L. (Fabaceae). *Bot. Not.* 124: 383–398.

Dahlgren, R. 1988. *Aspalathus*. Pp. 1–430 in: Leistner, O.A. (ed.), *Flora of Southern Africa*. National Botanical Institute, Pretoria.

Doyle, J.J. & Doyle, J.L. 1987. A rapid DNA isolation procedure for small amounts of fresh leaf tissue. *Phytochem. Bull.* 19: 11–15.

Edwards, D. 2006. Systematic studies of the South African endemic *Aspalathus* L. (Fabaceae) and the genistoid s.l. clade. PhD Thesis, University of Reading, Reading, U.K.

Edwards, D. & Hawkins, J.A. 2007. Are Cape floral clades the same age? Contemporaneous origins of two lineages in the genistoids s.l. (Fabaceae). *Molec. Phylog. Evol.* 45: 952–970.

Feliner, G.N., Larena, B.G. & Aguilar, J.F. 2004. Fine-scale geographical structure, intra-individual polymorphism and recombination in nuclear ribosomal Internal Transcribed Spacers in *Armeria* (Plumbaginaceae). *Ann. Bot.* 93: 189–200.

Gemeinholzer, B., Oberprieler, C. & Bachmann, K. 2006. Using GenBank data for plant identification: possibilities and limitations using the ITS 1 of Asteraceae species belonging to the tribes Lactuceae and Anthemideae. *Taxon* 55: 173–187.

Graham S.W., Reeves, P.A., Burns, A.C.E. & Olmstead, R.G. 2000. Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. *Int. J. Pl. Sci.* 161: S83–S96.

Hamilton, M.B. 1999. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molec. Ecol.* 8: 521–523.

Harris, S.A. 1995. Systematics and randomly amplified polymorphic DNA in the genus *Leucaena* (Leguminosae, Mimosoideae). *Pl. Syst. Evol.* 197: 195–208.

Hebert, P.D.N., Cywinska, A., Ball, S.L. & DeWaard, J.R. 2003a. Biological identifications through DNA barcodes. *Philos. Trans., Ser. B.* 270: 313–321.

Hebert, P.D.N. & Gregory, T.R. 2005. The promise of DNA barcoding for taxonomy. *Syst. Biol.* 54: 852–859.

Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H. & Hallwachs, W. 2004a. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc. Natl. Acad. Sci. U.S.A.* 101: 14812–14817.

Hebert, P.D.N., Ratnasingham, S. & DeWaard, J.R. 2003b. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Philos. Trans., Ser. B.* 270: S96–S99.

Hebert, P.D.N., Stoeckle, M.Y., Zemlak, T.S. & Francis, C.M. 2004b. Identification of birds through DNA barcodes. *PLoS Biol.* 2(10): e312 doi:10.1371/journal.pbio.0020312.

Hogg, I.D. & Hebert, P.D.N. 2004. Biological identification of springtails (Hexapoda: Collembola) from the Canadian Arctic, using mitochondrial DNA barcodes. *Canad. J. Zool.* 82: 749–754.

Hoot, S.B., Culham, A. & Crane, P.R. 1995. The utility of *atpB* gene sequences in resolving phylogenetic relationships: comparisons with *rbcL* and 18S ribosomal DNA sequences in the Lardizabalaceae. *Ann. Missouri Bot. Gard.* 82: 194–207.

Hughes, C. & Eastwood, R. 2006. Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. *Proc. Natl. Acad. Sci. U.S.A.* 103: 10334–10339.

Klak, C., Reeves, G. & Hedderson,T. 2004. Unmatched tempo of evolution in Southern African semi-desert ice plants. *Nature* 427: 63–65.

Koch, M.A. & Matschinger, M. 2007. Evolution and genetic differentiation among relatives of *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 104: 6272–6277.

Kress W.J. & Erickson D.L. 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(6): e508. doi:10.1371/journal.pone.0000508.

Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A. & Janzen, D.H. 2005. Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. U.S.A.* 102: 8369–8374.

Lahaye, R., Van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., Maurin, O., Duthoit, S., Barraclough,

T.G. & Savolainen, V. 2008. DNA barcoding the floras of biodiversity hotspots. *Proc. Natl. Acad. Sci. U.S.A.* 2008: 105: 2923–2928.

Lihova, J., Marhold, K., Kudoh, H. & Koch, M.A. 2006. Worldwide phylogeny and biogeography of *Cardamine flexuosa* (Brassicaceae) and its relatives. *Amer. J. Bot.* 93: 1206–1221.

Lorenz-Lemke, A.P., Muschner, V.C., Bonatto, S.L., Cervi, A.C., Salzano, F.M. & Freitas, L.B. 2005. Phylogeographic inferences concerning evolution of Brazilian *Passiflora actinia* and *P. elegans* (Passifloraceae) based on ITS (nrDNA) variation. *Ann. Bot.* 95: 799–809.

Maddison, D.R. & Maddison, W.P. 2001. MacClade. Sinauer Associates, Inc. Sunderland, Massachusetts.

Meyer, C.P. & Paulay, G. 2005. DNA barcoding; error rates based on comprehensive sampling. *PLoS Biol.* 3(12): e422 doi:10.1371/journal.pbio.0030422.

Müller, K. 2006. Incorporating information from length-mutational events into phylogenetic analysis. *Molec. Phylog. Evol.* 38: 667–676.

Newmaster, S.G., Fazekas, A.J. & Ragupathy, S. 2006. DNA barcoding in land plants: evaluation of *rbcL* in a multigene tiered approach. *Canad. J. Bot.* 84: 335–341.

Nixon, K.C. & Wheeler, Q.D. 1990. An amplification of the phylogenetic species concept. *Cladistics* 6: 211–223.

Noyes, R.D. 2006. Intraspecific nuclear ribosomal DNA divergence and reticulation in sexual diploid *Erigeron strigosus* (Asteraceae). *Amer. J. Bot.* 93: 470–479.

Oliver, E.G.H., Linder, H.P. & Rourke, J.P. 1983. Geographical distribution of present-day Cape taxa and their phytogeographical significance. *Bothalia* 14: 427–440.

Pennisi, E. 2007. Taxonomy. Wanted: a barcode for plants. *Science* 318: 190–191.

Piel, W.H., Donoghue, M.J. & Sanderson, M.J. 2002. TreeBASE: a database of phylogenetic information. Pp. 41–47 in: Shimura, J., Wilson, K.L & Gordon, D. (eds.), *To the Interoperable "Catalog of Life" With Partners – Species 2000 Asia Oceanea.* Research Report from the National Institute for Environmental Studies 171, National Institute for Environmental Studies, Tsukuba, Japan.

Rauscher, J.T., Doyle, J.J. & Brown, A.H.D. 2004. Multiple origins and nrDNA Internal Transcribed Spacer homeologue evolution in the *Glycine tomentella* (Leguminosae) allopolyploid complex. *Genetics* 166: 987–998.

Richardson, J.E., Pennington, R.T., Pennington, T.D. & Hollingsworth, P.M. 2001. Rapid diversification of a species-rich genus of neotropical rain forest trees. *Science* 293: 2242–2245.

Sang, T., Crawford, D.J. & Stuessy, T.F. 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *Amer. J. Bot.* 84: 1120–1136.

Seberg, O., Humphries, C.J., Knapp, S., Stevenson, D.W., Petersen, G., Scharff, N. & Andersen, N.M. 2003. Shortcuts in systematics? A commentary on DNA-based taxonomy. *Trends Ecol. Evol.* 18: 63–65.

Shaw, J., Lickey, E.B., Beck, J.T., Farmer, S.B., Liu, W.S., Miller, J., Siripun, K.C., Winder, C.T., Schilling, E.E. & Small, R.L. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Amer. J. Bot.* 92: 142–166.

Shaw, J., Lickey, E.B., Schilling, E.E. & Small, R.L. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise & the hare III. *Amer. J. Bot.* 94: 275–288.

Simmons, M.P. & Ochoterena, H. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.* 49: 369–381.

Small, R.L., Ryburn, J.A., Cronn, R.C., Seelanan, T. & Wendel, J.F. 1998. The tortoise and the hare: choosing between noncoding plastome and nuclear ADH sequences for phylogeny reconstruction in a recently diverged plant group. *Amer. J. Bot.* 85: 1301–1315.

Stevens, P.F. 1991. Character states, morphological variation, and phylogenetic analysis: a review. *Syst. Bot.* 16: 553–583.

Stoeckle, M. 2003. Taxonomy, DNA, and the bar code of life. *BioScience* 53: 796–797.

Sun, Y., Skinner, D.Z., Liang, G.H. & Hulbert, S.H. 1994. Phylogenetic analysis of *Sorghum* and related taxa using internal transcribed spacers of nuclear ribosomal DNA. *Theor. Appl. Genet.* 89: 26–32.

Swofford, D.L. 2002. *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods).* Sinauer Associates, Inc, Sunderland, Massachusetts.

Taberlet, P., Gielly, L., Pautou, G. & Bouvet, J. 1991. Universal primers for amplification of 3 noncoding regions of chloroplast DNA. *Pl. Molec. Biol.* 17: 1105–1109.

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., Vermat, T., Corthier, G., Brochmann, C. & Willerslev. E. 2007. Power and limitations of the chloroplast *trnL* (UAA) introns for plant DNA barcoding. *Nucleic Acids Res.* 35: e14 doi:10.1093/nar/gkl938.

Tautz, D., Arctander, P., Minelli, A., Thomas, R.H. & Vogler, A.P. 2002. DNA points the way ahead of taxonomy. – In assessing new approaches, it's time for DNA's unique contribution to take a central role. *Nature* 418: 479–479.

Tautz, D., Arctander, P., Minelli, A., Thomas, R.H. & Vogler, A.P. 2003. A plea for DNA taxonomy. *Trends Ecol. Evol.* 18: 70–74.

Theile, K. 1993. The holy grail of the perfect character: the cladistic treatment of morphometric data. *Cladistics* 9: 275–304.

Thompson, J.D., Higgins, D.G. & Gibson, T.J. 1994. ClustalW—improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673–4680.

Van Steenis, C.G.G.J. 1957. Specific and infraspecific delimitation. Pp. 167–234 in: Van Steenis, C.G.G.J. (ed.), *Flora Malesiana*, vol. I (5), *Spermatophyta.* P. Noordhoff, Leiden.

Vargas, P. 2003. Molecular evidence for multiple diversification patterns of alpine plants in Mediterranean Europe. *Taxon* 52: 463–476.

White, T.J., Bruns, T., Lee, S. & Taylor, J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pp. 315–322 in: Innis, M.A., Gelfand, D.H., Sninsky, J.J. & White, T.J. (eds.), *PCR Protocols: A Guide to Methods and Application.* Academic Press, San Diego.

Will, K.W., Mishler, B.D. & Wheeler, Q.D. 2005. The perils of DNA barcoding and the need for integrative taxonomy. *Syst. Biol.* 54: 844–851.