MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2015) **15**, 42–56 doi: 10.1111/1755-0998.12292

Molecular taxonomic identification in the absence of a 'barcoding gap': a test with the endemic flora of the Canarian oceanic hotspot

RUTH JAÉN-MOLINA,*,† ÁGUEDO MARRERO-RODRÍGUEZ,* J. ALFREDO REYES-BETANCORT,‡ ARNOLDO SANTOS-GUERRA,‡ JOSÉ NARANJO-SUÁREZ* and JULI CAUJAPÉ-CASTELLS*

**Jardín Botánico Canario "Viera y Clavijo"-Unidad Asociada CSIC, Cabildo de Gran Canaria, Apartado de correos 14 de Tafira Alta, 35017, Las Palmas de Gran Canaria, Spain, †Fundación Canaria Amurga Maspalomas, Avda. Tirajana 39, II 6, 35100 San Bartolomé de Tirajana, Spain, ‡Unidad de Botánica Aplicada, Jardín de Aclimatación de La Orotava (ICIA), c/Retama 2, 38400 Puerto de la Cruz, Tenerife, Spain*

Abstract

We use a comprehensive subset of Canarian angiosperms corresponding to 23 families, 35 genera and 60 Canarian endemic taxa to test whether this flora is suitable to taxonomic identification with the two proposed plant DNA barcode sequences and whether these sequences may reveal the existence of cryptic species overlooked by morphology. The rate of discrimination success between the insular congeneric samples using the *rbcL+matK* combination and a 'character-based' approach (where we use only the combination of nucleotide positions in an alignment that allows unambiguous species identification) is higher (82.29%) than that obtained with the 'distance-based' approach (80.20%) used by the CBOL Plant Working Group in 2009 and also when compared with tests conducted in other floras. This suggests that the molecular identification of the Canarian endemic flora can be achieved as successfully as in other floras where the incidence of radiation is not as relevant. The facts that (i) a distance-based criterion was unable to discriminate between congeneric and conspecific comparisons and (ii) only the character-based discrimination criterion resolved cases that the distance-based criterion did not, further support the use of a character discrimination approach for a more efficient DNA barcoding of floras from oceanic islands like the Canaries. Thus, a barcoding gap seems not to be necessary for the correct molecular characterization of the Canarian flora. DNA barcodes also suggest the possible existence of cryptic taxa to be further investigated by morphology and that the current taxonomic status of some of the taxa analysed may need revision.

Keywords: angiosperms, barcoding gap, Canarian endemics, diagnostic characters, DNA barcoding, hotspot, oceanic islands

Received 11 October 2013; revision received 22 May 2014; accepted 4 June 2014

Introduction

To provide an universal framework for the routine use of DNA sequence data in species identification is important from a wide range of biological perspectives, encompassing (i) the rapid assignation of specimens to their correct species when morphological traits are not sufficient (Kress *et al.* 2005; Savolainen *et al.* 2005); (ii) the identification of possible cryptic species overlooked by morphology (Hebert *et al.* 2003, 2004; Ragupathy *et al.* 2009; Gao *et al.* 2011); (iii) the assistance to classical taxonomy in the elaboration of censuses of plant biodiversity (Hajibabaei *et al.* 2007; Lahaye *et al.* 2008; de Vere *et al.* 2012) or (iv) the application of these data to conservation and management strategies (Hollingsworth 2008; Kress *et al.* 2009; Bruni *et al.* 2012). These and other distinctive areas of interest have triggered 'DNA barcoding' for species-level identification (i.e. the use of short-standardized DNA sequences to tag species; Barret & Hebert 2005).

Unlike animals, where the application of cytochrome-oxidase-I (COI) has been highly successful in a wide range of taxa (Smith *et al.* 2008; Hebert *et al.* 2010), the quest for efficient 'DNA barcodes' in plants is still under way (Fazekas *et al.* 2009, 2010; Chen *et al.* 2010). Following the assessment of 12 potential plant DNA barcoding regions (Kress *et al.* 2005; Chase *et al.* 2007; Kress & Erickson 2007; Fazekas *et al.* 2008; Newmaster *et al.* 2008; Ford *et al.* 2009), some members of the Plant Working Group within the International Consortium for

Correspondence: Ruth Jaén-Molina, Fax: +34-928-219-581; E-mail: ruthjaen@gmail.com

the Barcode of Life (PWG from now on) recommended the cpDNA two-locus combination *rbcL-matK* as the universal DNA barcode for land plants (CBOL Plant Working Group 2009). The potential applicability of plant DNA barcodes has been evaluated using these and other regions [such as *trnH-psbA* or internal transcribed spacer region *(ITS)*] and in different taxonomic and floristic contexts (Chase *et al.* 2007; Kress & Erickson 2007; Lahaye *et al.* 2008; Hollingsworth *et al.* 2009; Burgess *et al.* 2011; Li *et al.* 2011; de Mattia *et al.* 2011). Also, although the existence of a barcoding gap (i.e. a clear separation between the average intraspecific distances relative to the interspecific distances) is usual in animals and has resulted in high species discrimination success rates, it is not frequent in plants. Certainly, some authors have used a barcoding gap to distinguish plant groups above the species level (de Vere *et al.* 2012; Zhang *et al.* 2013) or have detected the gap in congeneric species (e.g. within the lichen family *Parmeliae*, Leavitt *et al.* 2013). However, there are also numerous studies in plants where the detection of a barcoding gap did not result in species discrimination (e.g. Pettengill & Neel 2010; Jiang *et al.* 2011; Yang *et al.* 2012; Ashfaq *et al.* 2013). Notably, Fazekas *et al.* (2009) found out that the degree of overlap between intra- and interspecific distances is generally much greater in plants than in animals, and they suggested that the higher the overlap, the more reduced the ability to discriminate species is.

In plants, the discriminatory power of the *rbcL+matK* sequence combination is of ca. 72% (CBOL Plant Working Group 2009; Hollingsworth *et al.* 2011) and thus still very far from the usually higher (though variable) rate of over 90% success of COI in animals, fungi and red algae (Hebert *et al.* 2010; Ran *et al.* 2010; Buchheim *et al.* 2011; Assunção *et al.* 2012). Consequently, many of the recent efforts led by PWG members are driven toward complementing this cpDNA 'core barcode' with other sequence regions (e.g. Hollingsworth *et al.* 2011; Li *et al.* 2011).

Among other reasons, elusiveness in the discrimination of ca. 30% of land plant species by cpDNA barcodes can be explained by highly speciose genera, by closely related species belonging to recently evolved lineages (radiation) or by processes such as hybridization and/or introgression (Hansen *et al.* 2003; Hoban *et al.* 2009). Radiation (i.e. the rapid evolution of phenotypic diversity within a lineage, with or without the stimulus of natural selection) is of concern for molecular identification because often conspicuous morphological differences are not paralleled by comparable levels of molecular polymorphism. Likewise, hybridization (i.e. the process that begets offspring by cross-fertilization between individuals belonging to different taxa or populations) plays an important role in plant and animal evolution (Harrison 1990; Arnold 2006). However, the species of hybrid origin may be elusive for plant barcoding because cpDNA sequences would not detect differences between the ancestral mother species and the derived ones.

Oceanic island floras host a much higher incidence of these two processes that may hamper the applicability of DNA barcodes (Whittaker & Fernández-Palacios 2007; Losos & Ricklefs 2009; Caujapé-Castells 2011) and may also give rise to 'false negatives' when little or no sequence variation in the barcoding fragments is found between different biospecies (Wiemers & Fiedler 2007). Thus, oceanic insular floras are important and challenging test grounds for the applicability of molecular barcodes in land plants.

The Canary Islands are a Spanish oceanic archipelago formed by seven major islands situated ca. 100 km off NW Africa that feature the greatest number of endemic plant species per unit area compared with any continental or insular area of its size in the northern quarter of Africa and Europe (Martín-Esquivel *et al.* 2005). The flora of the Canary Islands is highly diverse, harbouring around 1457 native taxa, 23 endemic genera (encompassing 46 endemic Canarian species; Moreno Saiz 2011), and between 592 (Fernández Palacios & Martín Esquivel 2001) and 680 (Reyes-Betancort *et al.* 2008) terrestrial endemic plants. As in many oceanic archipelagos worldwide, a considerable proportion of the Canarian endemic flora consists of species belonging to morphologically radiating lineages (Bramwell 1975) where molecular polymorphism is moderate or low, at best, for example: *Argyranthemum* (Francisco-Ortega *et al.* 1996), *Aeonium* (Mort *et al.* 2002), *Convolvulus* (Carine *et al.* 2004), *Sideritis* (Barber *et al.* 2007), *Sonchus* (Kim *et al.* 2007) or *Echium* (García-Maroto *et al.* 2009). Examples of interspecific hybridization also have been reported for a wide representation of the Canarian endemic flora, for example, in *Micromeria* (Pérez de Paz 1978), *Argyranthemum* (Brochmann 1984), *Aeonium* (Bañares 1990), *Sideritis* (Marrero 1992), *Sonchus* (Kim *et al.* 1996), *Rubus* (Alice *et al.* 2001), *Pericallis* (Van Hengstum *et al.* 2012) or the *Bencomia* and the *Gonospermum* alliances (Francisco-Ortega *et al.* 2001).

In this investigation, we evaluate the 'core' markers for plant identification (*matK* and *rbcL*, CBOL Plant Working Group 2009) in a sample of Canarian endemic angiosperms to answer three crucial questions concerning the applicability of these two cpDNA regions in a barcoding context: (i) does a DNA barcoding gap exist in the Canarian endemic species considered in this study? (ii) how efficiently does a character-based discrimination criterion retrieve the identities of different taxa with respect to the distance-based criterion used by the CBOL Plant Working Group (2009)? and (iii) may the plant barcode regions reveal the existence of cryptic species overlooked by morphology?

This investigation represents the first comprehensive test of the two plant barcode regions encompassing a significant number of endemic angiosperms of the Canarian Flora and can thus be a relevant contribution to the literature on DNA barcoding of insular plants.

Material and methods

Taxon sampling

We analysed 75 Macaronesian plant taxa containing 60 Canarian endemics (45 species, 12 subspecies and three varieties) that belong to 35 genera (eight of them endemic to Macaronesia). This sample represents 23 families that encompass 314 taxa (264 species and 50 subspecies), of which 179 are endemic of the Canary Islands (see Tables S1 and S2 for details, Supporting information). Remarkably, this sample contains representatives the taxonomically richest families in the Canarian flora (especially Asteraceae, Brassicaceae, Crassulaceae, Fabaceae, Lamiaceae and Poaceae), all the ecosystems in the archipelago, radiating and nonradiating genera, and endangered and nonendangered species (see Tables S1 and S3 for details, Supporting information).

Overall, we sampled 140 populations (118 Macaronesian [102 Canarian, 11 Madeiran and five Azorean] and 22 non-Macaronesian; Table S2, Supporting information), collecting between two and five individuals in each of them (herbarium voucher numbers shown in Table S1, Supporting information). We refrained from sampling larger numbers of individuals because the general lack of resolution in most available phylogenies of Canarian endemics based on *rbcL* and *matK* (Doebley *et al.* 1990; Soltis *et al.* 1993; Kim *et al.* 1999) led us to expect low intraspecific polymorphism levels for these two cpDNA sequences. The main goal of our sampling was to have a subset of the Canarian endemic angiosperms sufficiently representative to test the performance of the barcode sequences in the identification of the rich plant diversity of this oceanic archipelago.

Although the DNA barcode sequences are tools for identification and not taxonomy, we were especially interested in assessing whether they provide additional information to assist morphology in the evaluation of taxonomically conflictive cases (see Table 1 for definitions). Thus, the 75 Macaronesian endemic angiosperms analysed include 35 taxa unambiguously identified by classical taxonomy (nonconflictive cases, see Table 1 for definitions) and 40 species, conspecific populations and/or subspecies that are difficult to identify based only on available morphological characters (conflictive cases, see Table 1 for definitions). We also considered four Macaronesian and 22 non-Macaronesian populations of taxa unambiguously identified by classical taxonomy (sequences available in GenBank, see Table S2, Supporting information) with the aim of assessing the robustness of eventual discrimination.

Table 1 Definitions of basic concepts used throughout the paper

Concept	Definition
Taxonomically nonconflictive cases	Species unambiguously identified by taxonomy, based on morphological characters
Taxonomically conflictive cases	Species with conspecific populations difficult to identify based on available morphological characters
Potentially cryptic taxa	Taxonomically nonconflictive conspecific populations which are morphologically indistinguishable, but where we detected a substantial number of diagnostic characters with the barcode sequences
p-distances	Proportion of nucleotide differences detected between two aligned sequences, excluding indels
Diagnostic character	Each of the positions in an alignment (including indels) that allows the differentiation between sequences belonging to individuals of either of two congeneric species within a given sample
Polymorphic character	Each of the variable positions in an alignment that is shared by two or more taxa, and therefore cannot be used for species identification
Barcoding gap	Inter-specific genetic variation exceeding intra-specific variation to such an extent that a clear gap exists which enables the assignment of unidentified individuals to their species with a negligible error rate (Hebert *et al.* 2003; Meyer & Paulay 2005)

DNA isolation, amplification and sequencing

DNA extractions were performed from either silica gel dried or fresh material using the CTAB protocol (Doyle & Doyle 1987; Palmer *et al.* 1988). The concentrations of the extracted DNA were measured with the spectrophotometer *ND-1000* (NanoDrop), and aliquots were deposited in the DNA Bank of the Canarian Flora at the Jardín Botánico Canario 'Viera y Clavijo'-Unidad Asociada CSIC of the Cabildo de Gran Canaria (JBCVCSIC hence forward). Whereas the PCR conditions and primers used in all the taxa assessed were the same for *rbcL*, they were case dependent for *matK* (Table 2). PCRs were performed in a 25 μL total volume containing 2 μL of extracted DNA; 20 μL ReddyMix™ PCR Master Mix (ThermoScientific, Abgene, UK); 0.5 μL of each primer (20 μM); and 2 μL of bovine serum albumin (20 mg/mL

Table 2 Primer pairs and annealing temperatures used to amplify the DNA barcode sequences

Region	Primer	Sequence (5′-3′)	AT	Reference
rbcL	*rbcL*-F	ATGTCACCACAAACAGAGACTAAAGC	55°C	CBOL Plant Working Group (2009)
	rbcL-R	GAAACGGTCTCTCCAACGCAT		
matK	*matK*-F	AATTTACGATCHATTCATTCMATWTTT	48°C	Schaefer *et al*. (2011)
	matK-R	AGTTYTARCAAGAAAGTCGAARTATATA		
matK	AST-F	CCTTACCCAGCTCATCTGGAAAT	48°C	Dunning & Savolainen (2010)
	AST-R	CAAATAATATCCAAATACCAA		
matK	*matK* -XF	TAATTTACGATCAATTCATTC	46°C	Ford *et al*. (2009)
	matK -5R	GTTCTAGCACAAGAAAGTCG		
matK	3F_KIM	ACCCAGTCCATCTGGAAATCTTGGTTC	58°C	K. J. Kim (unpublished)
	1R_KIM	CGTACAGTACTTTTGTGTTTACGAG		
matK	*matk*-1F	ACTGTATCGCACTATGTATCA	53°C	Sang *et al*. (1997)
	matk-1R	GAACTAGTCGGATGGAGTAG		
matK	390-F	CGATCTATTCATTCAATATTTC	48°C	Cuénod *et al*. (2002)
	1326-R	TCTAGCACACGAAAGTCGAAGT		

AT, annealing temperature.

BSA, Sigma). The PCR products that provided a single band of sufficient intensity after running a 1.8% agarose gel were sent to Macrogen Inc. in Korea for bidirectional sequencing on an ABI 3730XL (Applied Biosystems, Foster City, CA, USA). In all cases, multiple individuals were sequenced in each species analysed to assess polymorphism. Previous to any further analysis, we used a fragment of each obtained sequence as a query in GenBank's BLAST algorithm to check that we retrieved congeneric sequences.

The GenBank codes of the individual sequences are given in Table S1 (Supporting information), and the aligned matrices with the sequences generated *de novo* are deposited as attachments in the genetic diversity digest D-DNASE-89 (for *matK*) and D-DNASE-90 (for *rbcL*) in the *Demiurge* information system (URLs http://www.demiurge-project.org/matrix_digests/D-DNASE-89 and http://www.demiurge-project.org/matrix_digests/D-DNASE-90).

Sequence alignment and data analyses

The consensus DNA sequences (contigs) obtained with paired forward and reverse sequences for each region were aligned using eight iterations of the option 'Muscle alignment' implemented in GENEIOUS version 5.4 (Drummond *et al*. 2011). To assess the interspecific discriminatory power of the two-locus combination, the alignments for *rbcL* and *matK* were assembled in a combined matrix.

The *p*-distances were calculated using DNADIST version 3.5c (Felsenstein 1993) as implemented in BioEdit (Hall 2007). Diagnostic characters (see Table 1 for definition) in the alignments for congeneric taxa were scored to assess their power of discrimination relative to *p*-distances (see below). The performance of the two plant barcode sequences under (i) the two discrimination criteria assessed; and (ii) different biological variables were tested using Kruskal–Wallis test (Kruskal & Wallis 1952) as implemented in XLSTAT version 2009.1.01 (Addinsoft 2009). This is a nonparametric statistical procedure that allows the comparison between average values from more than two groups of independent samples that can be of unequal size.

Discrimination criteria

For both *matK* and *rbcL*, we scored (i) the intra- and interspecific *p*-distances (see definition in Table 1) and (ii) the number of diagnostic positions (Table 1) including indels (i.e. those that allow the differentiation between individuals belonging to either of two congeneric species within a given sample; note that such positions can be monomorphic or polymorphic). The distribution of inter- vs. intraspecific *p*-distances was compared to assess the existence of a DNA barcoding gap (see Table 1 for definition). For each tested species pair, we diagnosed discrimination as successful if (i) the minimum uncorrected interspecific *p*-distance was higher than the maximum intraspecific *p*-distance (criterion-1, following CBOL Plant Working Group 2009) or (ii) there was at least one diagnostic position in our sample of individuals (criterion-2).

Results

We generated 438 (*rbcL+matK*) sequences (GenBank Accession numbers and other details in Table S1, Supporting information) that were analysed together with 52 sequences downloaded from GenBank (corresponding to 26 taxa, Table S2, Supporting information). PCR success

levels for *rbcL* and *matK* were 99% and 85%, respectively, and sequencing success rates were 98% in *rbcL* and close to 90% with *matK*. The sequence quality for the two markers was generally high for both sequencing directions, although we encountered problems to obtain high-quality reverse sequences for *matK* in some taxonomic groups.

We carried out 1314 pairwise comparisons (657 for each *rbcL* and *matK*) to evaluate the discriminatory power of each DNA barcode region for 147 nonconflictive cases and 67 conflictive cases. Also, we calculated intra- and interspecific *p*-distances for, respectively, all species and species pairs analysed. We scored diagnostic and polymorphic characters (see definitions on Table 1) for congeneric taxa only. The results of Kruskal–Wallis tests are shown in Table 3.

Distance analyses

At the intraspecific level, *rbcL* was nonsignificantly more variable than *matK*, with *p*-distances ranging from 0 to 0.047 and from 0 to 0.027, respectively. At the interspecific level, *matK* was significantly ($P < 0.0001$) more variable than *rbcL* (*p*-distances from 0 to 0.080 and from 0 to 0.068, respectively). The highest average interspecific *p*-distance found with *matK* within a genus [0.0800, in *Silene* (Caryophyllaceae)] was much higher than that found with *rbcL* [0.0624, in *Erica* (Ericaceae)]. The largest overlap between intra- and interspecific distances was found in the family Ericaceae. Although the average interspecific *p*-distance was significantly higher than the average intraspecific *p*-distance for both DNA barcode regions ($P < 0.0001$), we did not detect a barcoding gap because intra- and interspecific *p*-distances overlapped (see Fig. 1).

Character analyses

We detected over 2.5 times more diagnostic characters with *matK* than with *rbcL* for nonconflictive cases (respectively 9.156 vs. 3.381, $P < 0.0001$), and more than twice as many as with *rbcL* for the conflictive cases (respectively 1.179 vs. 0.493, $P < 0.0001$; Table 3). The average number of polymorphic nondiagnostic positions detected with *matK* was significantly higher ($P < 0.006$) for the conflictive cases (0.896) than for the nonconflictive cases (0.469), while for *rbcL*, the differences were not significant (3.164 and 1.980, respectively). Similar figures were obtained when we analysed only insular taxa (Table 3).

Distances vs. characters (rbcL+matK)

When comparing all the taxa assessed (147 nonconflictive and 67 conflictive cases), criterion-1 was less discriminating than criterion-2 (see the methods for the description of discrimination criteria). Also, when comparing only insular taxa (96 nonconflictive and 67 conflictive cases), *p*-distances resolved a lower number of cases than diagnostic characters.

matK vs. rbcL

Within all comparisons (taxonomically nonconflictive and conflictive cases), *matK* showed a much higher proportion of resolution success under either criterion than *rbcL* (Table 4), and each individual region provided a much higher proportion of discrimination with criterion-2 than with criterion-1 (Table 4). Similarly, within only insular taxa (Canarian and Macaronesian), *matK*

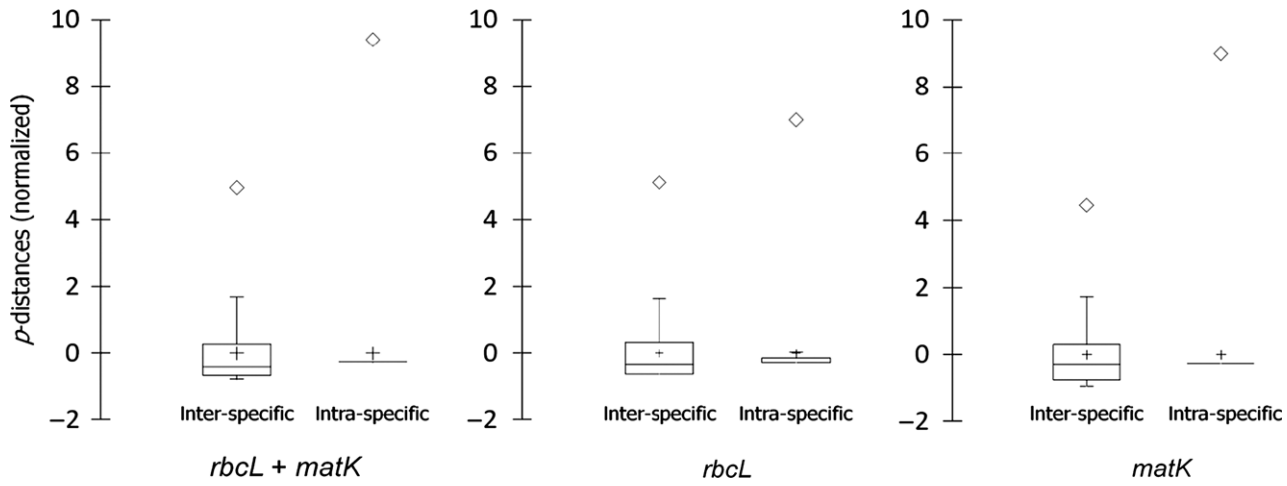

Fig. 1 Boxplots of the normalized inter- and intraspecific *p*-distances for *rbcL+matK*, *rbcL* and *matK* in all taxonomically clear cases compared in this study. The crosses correspond to the averages, and the solid lines within the box plots to the medians. Outliers are marked with a diamond.

Table 3 Average values of the parameters used to assess the taxonomic discrimination of the barcode sequences for different grouping categories under Kruskal–Wallis tests. The averages of groups supported significantly by the tests are separated by brackets and parentheses (note that, in some cases, the averages of certain categories fall in more than one statistical group). The integer numbers in brackets after the group names correspond to the sample sizes of each group

	matK			*rbcL*		
Categories tested	*p*-distances	Diagnostic characters	Polymorphic characters	*p*-distances	Diagnostic characters	Polymorphic characters
Only insular taxa						
By cases						
Nonconflictive (96) vs. conflictive (67)	(0.012) (0.003)***	(8.250) (1.179)***	(0.344) (0.896)**	(0.007) (0.004)***	(3.00) (0.493)***	(1.531, 3.164)^{NS}
By taxonomic level						
Genera (12) vs. species (97) vs. subspecies (10) vs. conspecific populations (44)	(0.015, 0.012) (0.001, 0.001)***	(11.500) (7.340) (0.200, 0.432)***	(0.000, 0.495, 1.100, 0.77)^{NS}	(0.011) (0.006 [0.005) 0.001]***	(6.333) (2.299) (0.000, 0.500)***	(0.000) (1.649, 1.600) (4.159)**
By Island distribution						
Same island (26) vs. different islands (137)	(0.008, 0.008)^{NS}	(5.500, 5.314)^{NS}	(0.500, 0.584)^{NS}	(0.005, 0.006)^{NS}	(1.577, 2.044)^{NS}	(2.423, 2.161)^{NS}
Gran Canaria (7) vs. La Gomera (15) vs. Tenerife (4)	(0.002 [0.009) 0.018]*	(0.429 [6.000) 12.500]*	(0.286, 0.600, 0.500)^{NS}	(0.000 [0.005) 0.013]**	(0.000 [1.400) 5.000]**	(0.000, 2.800, 5.250)^{NS}
By growth-habit						
Trees (49) vs. other (114)	(0.012, 0.007)^{NS}	(8.501, 3.965)^{NS}	(0.776, 0.482)^{NS}	(0.013) (0.003)***	(3.510) (1.307)**	(6.837) (0.211)***
All taxa (insular and noninsular)						
By cases						
Nonconflictive (147) vs. conflictive (67)	(0.014) (0.003)***	(9.156) (1.179)***	(0.469) (0.896)**	(0.008) (0.004)***	(3.381) (0.493)***	(1.980, 3.164)^{NS}
By taxonomic level						
Genera (22) vs. species (138) vs. subspecies (10) vs. conspecific populations (44)	(0.018, 0.013) (0.001, 0.001)***	(10.500) (8.500) (0.200, 0.432)***	(0.182 [0.580, 1.100) 0.773]*	(0.008, 0.008) (0.001, 0.005)***	(4.545) (2.957) (0.000, 0.500)***	(0.000 [2.203, 1.600) 4.159]**
By growth-habit						
Trees (64) vs. other (150)	(0.013, 0.010)^{NS}	(8.625, 5.820)^{NS}	(0.750, 0.540)^{NS}	(0.013) (0.004)***	(3.500) (2.040)**	(7.344) (0.220)***

NS, nonsignificant.
*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

Table 4 Summary of the percentage of discrimination success (all taxa vs. only insular taxa) provided by *matK* and *rbcL* under a distance-based (CR-1) or a character-based (CR-2) criterion, between conspecific species pairs for taxonomically nonconflictive cases ('Nonconflictive') and for taxonomically conflictive cases ('Conflictive')

	All taxa			Only insular taxa		
Cases/criterion	*matK* (%)	*rbcL* (%)	*rbcL+matK* (%)	*matK* (%)	*rbcL* (%)	*rbcL+matK* (%)
'Nonconflictive'						
CR-1	85.03	65.77	86.39	79.16	45.16	80.20
CR-2	85.71	66.44	87.07	81.25	53.22	82.29
'Conflictive'						
CR-1	23.88	7.46	24.18	21.39	8.69	21.60
CR-2	32.85	11.94	33.29	36.95	10.86	37.31

The results for only Canarian taxa are not shown because the percentage of discrimination success was the same (75.38%) under either criterion for *matK* and for *rbcL+matK*, although for *rbcL*, criterion-2 performed better than criterion-1 (45.16% and 38.70%, respectively).

showed again a much higher proportion of resolution than *rbcL*, and each region provided a much higher proportion of discrimination with criterion-2 than with criterion-1 (Table 4). While the character-based diagnosis significantly distinguished between conspecific and congeneric comparisons in our sample of the Canarian Flora (Table 3), the distance-based criterion could not discriminate between both taxonomic levels. Under criterion-2, *matK* resolved more nonconflictive cases than *rbcL* for all 214 comparisons carried out (126 vs. 102, respectively) and also more conflictive cases (22 vs. nine, respectively). Furthermore, *matK* resolved more cases than *rbcL* based on *p*-distances (140 vs. 106, respectively). A total of 42 cases (26 nonconflictive and 16 conflictive) were resolved only by *matK* and not by *rbcL*, while only four cases (two nonconflictive and two conflictive) were resolved by *rbcL* and not by *matK*. However, as shown in Table 4, the highest resolution for the taxonomically nonconflictive cases tested is provided by the combination of the two barcode sequences using a discrimination criterion based on diagnostic characters (criterion-2): 87.07% for all taxa examined and 82.29% for only insular taxa.

Possible cryptic species overlooked by morphology

Although DNA barcode sequences generally distinguished well between nonconflictive species, they provided no discrimination in other such cases (for instance the species in the genus *Parolinia*, see Table 5 and Discussion).

On the other hand, unexpectedly, we detected a number of diagnostic molecular characters that allowed us to differentiate among several morphologically very similar conspecific populations that may be indicating potentially cryptic taxa (see definition in Table 1) in *Erica canariensis* (Ericaceae); *Argyranthemum broussonetii* ssp. *broussonetii* and ssp. *gomerensis* (Asteraceae); *Festuca agustinii* (Poaceae); *Lolium canariense* (Poaceae); *Galium scabrum* (Rubiaceae); *Picconia excelsa* (Oleaceae), and *Viola riviniana* (Violaceae). Table 5 summarizes different aspects regarding these and other cases and provides some future recommendations for each concerned taxon. Although these considerations should be taken with precaution and be supported by further investigation, they are necessary to assess the taxonomic status of these especially divergent populations.

Discussion

The radiations that characterize many oceanic island floras entail that conspicuous morphological differences between congeneric species are generally paralleled by only low (if at all) sequence polymorphism levels. The Canaries are no exception to this general pattern (e.g. Carine *et al.* 2004; Barber *et al.* 2007; Kim *et al.* 2007; or García-Maroto *et al.* 2009), thus posing a challenge for 'DNA barcoding', whose usefulness is contingent upon the ability of the proposed sequence regions to discriminate between congeneric species which are readily diagnosable by taxonomy. Cogent with this fact, low levels of species discrimination have been reported when barcoding plant and animal species on islands, for instance, in several genera of Asteraceae from Galápagos (Seberg & Petersen 2009), of Leguminosae from Macaronesia (Ojeda *et al.* 2008), or in the arthropod genus *Copelatus* (Dytiscidae) from the Fijian archipelago (Monaghan *et al.* 2005).

Despite these potentially challenging issues inherent to island taxa, our proportions of success in the identification of either the Canarian and Macaronesian endemic angiosperm taxa with no taxonomic conflict (75.38% and 82.29%, respectively) are both above the 72% obtained in floras of nonoceanic regions of the world (CBOL Plant Working Group 2009; Hollingsworth *et al.* 2011). This and previous results (Jaén-Moli-

Table 5 Conflictive congeneric and infra-specific taxa assessed in this study and nonconflictive populations which are morphologically indistinguishable but where at least one of the two barcode regions suggests that there could be possible cryptic taxa overlooked by taxonomy

Level of comparison/taxa	Case description	*matK*	*rbcL*	Recommendation
Congeneric species				
Apollonias ceballosii (G) *Apollonias barbujana* (G, M)	Morphologically very similar (Mesa Coello *et al.* 2003)	0	0	Their current taxonomic separation is not supported by the two barcode sequences
Euphorbia bourgeauana (T) *Euphorbia lambii* (G)	Morphologically very similar (Santos 1988; Molero *et al.* 2002)	0	0	Their current taxonomic separation is not supported by the two barcode sequences
Laurus novocanariensis (G, M) *Laurus azorica* (A)	Although the recognition of *L. azorica* as the only *Laurus* endemic to Mac and southern Morocco (Barbero *et al.* 1981; Jalas & Suominen 1991) is supported by molecular data (Arroyo-García *et al.* 2001; Rodríguez-Sanchez *et al.* 2009), in 2002 Rivas-Martínez proposed a new name (*L. novo-canariensis*) for the Madeiran and Canarian populations	1 (G-A) 1 (M-A)	0	Keep separation & revise
Lolium canariense (C, T, G) *Lolium edwardii* (P)	Morphologically very similar, and some intergrades have been observed (unpublished data)	1 (C-P) 1 (T-P) 2 (G-P)	0	Keep separation & revise
Morella rivas-martinezii (G) *Morella faya* (G)	Morphologically and genetically very similar (Bañares 1990; González-Pérez *et al.* 2009)	0	0	Their current taxonomic separation is not supported by the two barcode sequences. Revise, including populations from H (the *locus classicus* of *M. rivas-martinezii*) and P
Parolinia filifolia (C) *Parolinia ornata* (C)	Morphologically very similar, and 'some populations of *P. ornata* (Veneguera & Agaete) have been integrated based on their morphology into the populational complex of *P. filifolia*' (Fernández-Palacios 2009)	0	0	Their current taxonomic separation is not supported by the two barcode sequences
Picconia excelsa (G, M) *Picconia azorica* (A)	Morphologically very similar and their taxonomy has long been discussed (Seubert & Hochstetter 1844; Tutin 1933)	1 (G-A) 1 (M-A)	13 (G-A) 2 (M-A)	Keep separation
Polycarpaea divaricata (C) *Polycarpaea latifolia* (C)	Morphologically very similar where they share ecozones, and some authors have questioned their status as different species (Poiret 1816; Pitard & Proust 1909)	1	0	Keep separation
Silene bourgeaui (G) *Silene lagunensis* (T)	Morphologically very similar (Kunkel 1992)	27	6	Keep separation

Table 5 (Continued)

Level of comparison/taxa	Case description	*matK*	*rbcL*	Recommendation
Infra-specific taxa				
Argyranthemum broussonetii ssp. *broussonetii* (G) *Argyranthemum broussonetii* ssp. *gomerensis* (G)	Different morphologically and genetically (Humphries 1976; Francisco-Ortega *et al.* 1996)	3	0	Keep separation & revise (their current taxonomic status should probably be hoisted).
Cistus chinamadensis ssp. *gomerae* (G) *Cistus chinamadensis* ssp. *chinamadensis* (G)	Morphologically quite distinct, showing a geographic disjunction	0	0	Keep separation
Euphorbia mellifera var. *canariensis* (G, T) *Euphorbia mellifera* var. *mellifera* (M)	The populations from G and M are morphologically different to the population from T (Molero & Rovira 2005), and we suspected that there could be differences in T worth of taxonomic consideration	0	0	Keep separation
Ilex perado ssp. *lopezlilloi* (G) *Ilex perado* ssp. *platyphylla* (G)	Morphologically distinct, but not always clear (Kunkel 1977; Andrews 1983) and low genetic differentiation (Werner *et al.* 2007; González-González 2011)	0	0	Keep separation & revise
Micromeria varia ssp. *gomerensis* (G) *Micromeria varia* ssp. *canariensis* (C) *Micromeria varia* ssp. *rupestris* (F) *Micromeria varia* ssp. *thymoides* (M)	*M. varia* is morphologically highly variable in the Canary Islands and Madeira (Pérez de Paz 1978). Some taxonomists consider that the differences of some current subspecies may deserve a specific status (unpubl.)	0	0	Keep separation
Teline stenopetala ssp. *pauciovulata* (G) *Teline stenopetala* ssp. *microphylla* (G)	Morphologically very different and substantial molecular diversity between these subspecies (Percy & Cronq 2002) may question their current taxonomic status	0	0	Keep separation & revise (include other subspecies of *T. stenopetala*, and other species of *Teline*)
Conspecific populations				
Erica canariensis (G) *Erica canariensis* (T) *Erica canariensis* (M)	Désamoré *et al.* (2011) report a low level of haplotype diversity in Macaronesia for *E. canariensis* (earlier known as *E. arborea*, Rivas-Martínez 2011)	3 (G-M) 3 (T-M)	1 (G-M) 8 (T-M)	Revise
Festuca agustinii (C) *Festuca agustinii* (T) *Festuca agustinii* (G)	Morphologically very similar and the latest molecular study does not support differences between the populations of this species within the Canaries (Díaz-Pérez *et al.* 2012)	1 (C-T) 1 (C-G)	2 (C-T) 2 (C-G)	Revise

Table 5 (Continued)

Level of comparison/taxa	Case description	*matK*	*rbcL*	Recommendation
Galium scabrum (G) *Galium scabrum* (T)	It shows a high morphological variability in the Canary Islands, and all plants from G are glabrescent, unlike those from T (Bornmüller 1904)	1	0	Revise
Lolium canariense (G) *Lolium canariense* (T) *Lolium canariense* (C)	Morphologically very similar, although, Scholz *et al.* (2000) contend that some plants from H previously considered *L. canariense* should be included within *L. edwardii*	1 (G-T) 1 (G-C)	0	Revise (further studies need to include populations from H)
Picconia excelsa (G) *Picconia excelsa* (M)	A high sequence variation was detected within *P. excelsa* in a phylogenetic study with the *cp*DNA region *trnH-psbA* (Ferreira *et al.* 2011)	0	8	Revise
Viola riviniana (G) *Viola riviniana* (M)	*Viola riviniana* is taxonomically complex, and the definition of best morphological diagnostic characters remains obscure (Muñoz Garmendia 1993)	3	0	Revise

The column 'Case description' summarizes relevant aspects related to the taxonomic identification of the taxa compared, based on available evidence. The abbreviations correspond to G, La Gomera; P, La Palma; T, Tenerife; C, Gran Canaria; F, Fuerteventura; M, Madeira; A, Azores; Mac, Macaronesia. Columns *matK* and *rbcL* are the number of diagnostic characters detected with either region. 'Recommendation' summarizes the decision on the taxonomic status of the concerned taxa if the barcodes were the only variable available; the meaning of the different indications is as follows. 'Keep separation': The DNA barcodes support the current separation of the species or infra-specific taxa compared (note that because DNA barcoding only applies to species, we keep separation in the infra-specific taxa where no diagnostic characters were detected); 'Revise': the number of diagnostic differences detected with the core barcode sequences suggests the need of further taxonomic studies in the compared taxa. *Ad hoc* considerations are given when needed.

na *et al.* 2010) are in line with the finding that discrimination success rates are considerably higher when studies are focused on geographically defined species (Li *et al.* 2011; de Vere *et al.* 2012), thus agreeing with the suggestion of Hollingsworth *et al.* (2011) that geographically focused studies may be the best use of DNA barcoding.

Our tests also compellingly indicate that only our character-based criterion-2 permits to distinguish comparisons between congeneric species (the basic target of DNA barcoding) from those involving conspecific populations (Table 4). Furthermore, criterion-2 resolved nine cases that criterion-1 did not, whereas no case was resolved only by criterion-1. Thus, a character-based discrimination criterion seems to be the best choice for maximizing the success rate of molecular identification in the Canarian Flora using *rbcL* and *matK* sequences. This does not mean, however, that distance-based approaches could not be viable in other areas or that character-based alternatives to criterion-2 may not be developed (e.g. the BLOG software system, Weitschek *et al.* 2013).

As expected, the highest percentage of resolution (*rbcL+matK*) for the insular taxa analysed with no taxonomic conflict is much higher than that for conflictive cases (respectively 82.29% and 37.31%). However, discrimination for the conflictive cases is again much higher with criterion-2 (32.85% with *matK* and 11.94% with *rbcL*) than with criterion-1 (23.88% with *matK* and 7.46% with *rbcL*). Thus, taxonomic conflict in our sample is closely associated with a much lower number of diagnostic characters and a higher number of polymorphic characters with respect to nonconflictive cases (see Table 3). Notably, the much higher discriminating success of criterion-2 with respect to criterion-1 in taxonomically conflictive cases further highlights the potential of character-based diagnosis in the molecular identification of species belonging to the recent radiations that characterize the floras of many oceanic insular hotspots like the Canaries.

Although the situation of 'taxonomic conflict' prevents us from discussing discrimination success further in these cases (i.e. we cannot decide which of these cases are indeed conspecific and which ones are not), we give in Table 5 some indications based on the results with the barcode sequences. In general, because *matK* and *rbcL* often show no variation between closely related species, additional DNA regions and a more intensive sampling of these conflictive cases would be necessary to provide a better molecular basis to substantiate any taxonomic reinterpretation. The only conflictive infraspecific taxa between which we detected diagnostic characters (three characters, with *matK* only) were the two subspecies of *Argyranthemum broussonetii* (Table 5).

We could not retrieve the taxonomic separation among some of the nonconflictive species pairs analysed, most notably in the case involving some of the currently recognized species of the Canarian endemic genus *Parolinia* (Brassicaceae). Overall, these cases underscore that DNA barcoding should be used as a complement to morphological data and never as a surrogate. In the cases where the sequence variation detected with *matK* and *rbcL* might suggest the existence of cryptic taxa (see Results and Table 5), further studies with an increased sampling and/or other DNA markers may be needed to confirm the existence of diagnostic molecular characters. If that was the case, it would be necessary to investigate the possible existence of morphological characters previously disregarded.

A plethora of approaches has been used to analyse plant DNA barcoding data, but there is still a burgeoning debate about which method is the best (DeSalle *et al.* 2005; Meier *et al.* 2006; Little & Stevenson 2007; Meier 2008; Fazekas *et al.* 2009; CBOL Plant Working Group 2009; Lou & Golding 2010). It is important to note that phylogenetic or distance trees are highly informative about the genetic affinities between sequences from different species in a sample, but they are built based either on distances or on synapomorphies (i.e. derived shared nucleotide changes), and they may be misleading to discuss DNA barcoding (which relies on exclusive differences between the taxa or autapomorphies). By contrast, diagnostic character methods like our criterion-2 ignore within-species variation that may obscure discrimination and use the differentiating nucleotides only (see also DeSalle *et al.* 2005; Reid *et al.* 2011). Therefore, although distance-based criteria can be applied when there is a barcoding gap (e.g. Hebert *et al.* 2004; Barrett & Hebert 2005; de Vere *et al.* 2012), their usefulness probably decreases in contexts where the gap does not exist (Lahaye *et al.* 2008; Jiang *et al.* 2011), like in the Canarian endemic flora.

Plant DNA barcoding has proven to be very important in the conservation and management of biodiversity (Lahaye *et al.* 2008; Kress *et al.* 2009; de Vere *et al.* 2012). Notably, oceanic islands harbour around one quarter of all known extant vascular plant species worldwide (Kreft *et al.* 2008; Kier *et al.* 2009); the unique characteristics of these enclaves (mostly derived from their volcanic origin, their isolation and geographical ruggedness) make their plant endemics much more sensitive to rapid environmental changes than those from other enclaves (Caujapé-Castells *et al.* 2010). Our tests provide evidence for the applicability of the two core DNA barcoding sequences as identification tools for the Canarian endemic flora (see also Jaén-Molina *et al.* 2010). Taking into account that we have combined the creation of a reference database of DNA barcode sequences with reliable taxonomic information, these data are of wide applicability to conserva-

tion and management efforts encompassing any element of this highly diverse and distinctive flora.

Concluding remarks

Our results indicate that the combination of *matK* and *rbcL* sequence data under a character-based discrimination criterion provides the most useful approach to complement morphology in the taxonomical identification of the Canarian endemic angiosperms, for which a DNA 'barcoding gap' was not detected. Cogent with the high discrimination levels obtained using diagnostic characters for the nonconflictive taxa both in our sample of Macaronesian angiosperms (82.29%) and in all comparisons conducted (87.07%), we can conclude that a barcoding gap is not a necessary prerequisite for the correct molecular identification of the Canarian flora, similarly as found in other organisms and geographical contexts (Wiemers & Fiedler 2007; Ross *et al.* 2008; Lou & Golding 2010; Pettengill & Neel 2010; Zou *et al.* 2011). However, our estimate of success has to be considered tentative; the addition of new samples to our data set may either reduce the number of diagnostic characters per species (ultimately leading to a decrease in our success percentage in some comparisons) or, in other cases, reveal putative cryptic species. Our results also suggest that combining traditional morphological taxonomy with DNA barcoding is useful to identify possible cryptic species in plants (Table 5) and thereby reveal previously hidden biodiversity (Blaxter 2004).

The technical difficulties found in the amplification of *matK* represent a glaring breach of the universality requisite underlying 'DNA barcoding' (Fazekas *et al.* 2008, 2009; Dunning & Savolainen 2010; Hollingsworth *et al.* 2011). However, *matK* clearly resolved many more cases and offered many more diagnostic characters than *rbcL* (Table 3). Therefore, we believe that every effort toward the improvement of *matK*'s amplification universality is highly needed.

Despite the high discrimination success rate obtained with only these two sequences, additional regions are needed to resolve the approximately 20% of cases for which we did not find diagnostic characters. Being aware of this limitation in many floras of the world, the PWG has been exploring new markers to complement the official barcode core. For instance, Kress *et al.* (2005) suggest that the nuclear *ITS* and the plastid *trnH-psbA* intergenic spacer are more suitable regions for use in the DNA barcoding of land plants (Kress & Erickson 2007), and Yao *et al.* (2010) and Chen *et al.* (2010) suggested that *ITS2* should be used as a DNA barcode for plant identification. Similarly, Li *et al.* (2011) recommended that the nuclear *ITS* region should be incorporated into the core barcode for seed plants (see also Hollingsworth 2011a).

Acknowledgements

We thank Dr. Ángel Fernández and all his technical staff at Garajonay National Park (La Gomera), and Pedro Romero (at the Unidad Técnica Medio Ambiental of the Cabildo de La Gomera) for their invaluable help and feedback; Rubén Barone for his helpful comments about *Heberdenia excelsa*, and his generosity collecting samples and sharing photographs and vouchers of *H. excelsa* from Madeira; all the collaborators of the DNA Bank of the Canarian Flora at the JBCVCSIC for help in sample collection; Dr. Miguel Ángel González-Pérez (formerly at the Department of Biology, ULPGC), who gave us DNA aliquots of *Ilex perado* ssp. *platyphylla*, *Morella faya* and *Sambucus nigra* ssp. *palmensis*; the staff at the JBCVCSIC for unrestricted collaboration; the Consortium for the Barcoding of Life, members of the Plant Working Group, and the delegates of Tree-BOL provided helpful scientific interaction; the Cabildo de Gran Canaria for institutional support to all our research lines. This work was supported by projects 129/2006 of the 'Organismo Autónomo de Parques Nacionales' and 'arBOLcan' (funded by the ACIISI, Gobierno de Canarias). It also received complementary support through projects 'Demiurgo' (MAC/1/C20), 'Bioclimac' (MAC/1/C067) and 'Tree-BOL' (Alfred P. Sloan Foundation and NY Botanical Garden at Bronx). The 'Fundación Canaria Amurga Maspalomas' also furnished ancillary support and collaboration.

References

Addinsoft (2009) XLSTAT 2009.1.01. http://www.xlstat.com.

Alice LA, Eriksson T, Eriksen B, Campbell CS (2001) Hybridization and gene flow between distantly related species of *Rubus* (Rosaceae): evidence from nuclear ribosomal DNA internal transcribed spacer region sequences. *Systematic Botany*, **26**, 769–778.

Andrews S (1983) A Reappraisals of *Ilex aquifolium* and *I. perado* (Aquifoliaceae). *Kew Bulletin*, **39**, 141–155.

Arnold ML (2006) Evolution through genetic exchange. Oxford series. *Ecology and Evolution*. Oxford University Press, Oxford, UK.

Arroyo-García R, Martínez-Zapater JM, Fernández Prieto JA *et al.* (2001) AFLP evaluation of genetic similarity among laurel populations (*Laurus* L.). *Euphytica*, **122**, 155–164.

Ashfaq M, Asif M, Anjum ZI *et al.* (2013) Evaluating the capacity of plant DNA barcodes to discriminate species of cotton (*Gossypium:* Malvaceae). *Molecular Ecology Resources*, **13**, 573–582.

Assunção P, Jaén-Molina R, Caujapé-Castells J *et al.* (2012) Molecular taxonomy of *Dunaliella* (Chlorophyceae), with a special focus on *D. salina*: *ITS2* sequences revisited with an extensive geographical sampling. *Aquatic Biosystems*, **8**, 1–32.

Bañares Á (1990) Híbridos de la familia Crassulaceae en las Islas Canarias. Novedades y datos corológicos, II. *Vieraea*, **18**, 65–85.

Barber JC, Finch CC, Francisco-Ortega J *et al.* (2007) Hybridization in Macaronesian *Sideritis* (Lamiaceae): evidence from incongruence of multiple independent nuclear and chloroplast sequence datasets. *Taxon*, **56**, 74–88.

Barbero M, Benabid A, Peyre C *et al.* (1981) Sur la présence au Maroc de *Laurus azorica* (Seub.) Franco. -Act. III Congr. OPTIMA. *Anales del Jardín Botánico de Madrid*, **37**, 467–472.

Barrett RDH, Hebert PDN (2005) Identifying spiders through DNA barcodes. *Canadian Journal of Zoology*, **83**, 481–491.

Blaxter ML (2004) The promise of a DNA taxonomy. *Proceedings of the Royal Society B: Biological Sciences*, **359**, 669–679.

Bornmüller J (1904) Ergebnisse zweier botanischer nach Madeira und der Canarischen Inseln. *Botanische Jahrbücher für Systematik, Pflanzengeschichte und Pflanzengeographie*, **33**, 387–492.

Bramwell D (1975) Some morphological aspects of the adaptive radiation of Canary Islands *Echium* species. *Anales del Instituto Botánico A. J. Cavanilles*, **32**, 241–254.

Brochmann C (1984) Hybridization and distribution of *Argyranthemum coronopifolium* (Asteraceae—Anthemideae) in the Canary Islands. *Nordic Journal of Botany*, **4**, 729–736.

Bruni I, de Mattia F, Martellos S, Galimberti A, Savadori P *et al.* (2012) DNA barcoding as an effective tool in improving a digital plant identification system: a case study for the area of Mt. Valerio, Trieste (NE Italy). *PLoS One*, **7**, e43256.

Buchheim MA, Keller A, Koetschan C *et al.* (2011) Internal transcribed spacer 2 (nu *ITS*2 *rRNA*) sequence-structure phylogenetics: towards an automated reconstruction of the green algal tree of life. *PLoS One*, **6**, e16931.

Burgess KS, Fazekas AJ, Kesanakurti PR *et al.* (2011) Discriminating plant species in a local temperate flora using the *rbcL+matK* DNA barcode. *Methods in Ecology and Evolution*, **2**, 333–340.

Carine MA, Russell SJ, Santos-Guerra A *et al.* (2004) Relationships of the Macaronesian and Mediterranean floras: molecular evidence for multiple colonizations into Macaronesia and back-colonization of the continent in *Convolvulus* (Convolvulaceae). *American Journal of Botany*, **91**, 1070–1085.

Caujapé-Castells J (2011) Jesters, red queens, boomerangs and surfers: a molecular outlook on the Canarian endemic flora. In: *The Biology of Islands Floras* (eds Bramwell D & Caujapé-Castells J), Chapter 12, pp. 284–324. Cambridge University Press, UK.

Caujapé-Castells J, Tye A, Crawford DJ *et al.* (2010) Conservation of oceanic island floras: present and future global challenges. *Perspectives in Plant Ecology, Evolution and Systematics*, **12**, 107–129.

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

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

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

Cuénod P, Savolainen V, Chatrou LW *et al.* (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.

DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**, 1905–1916.

Désamoré A, Laenen B, Devos N *et al.* (2011) Out of Africa: Northwestwards Pleistocene expansions of the heather *Erica arborea*. *Journal of Biogeography*, **38**, 164–176.

Díaz-Pérez AJ, Sequeira M, Santos-Guerra A *et al.* (2012) Divergence and biogeography of the recently evolved Macaronesian red *Festuca* (Gramineae) species inferred from coalescence-based analyses. *Molecular Ecology*, **21**, 1702–1726.

Doebley J, Dummin ML, Goiezrc EM *et al.* (1990) Evolutionary analysis of the large subunit of carboxylase (*rbcL*) nucleotide sequence among the grasses (Gramineae). *Evolution*, **44**, 1097–1108.

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**, 11–15.

Drummond AJ, Ashton B, Buxton S *et al.* (2011) *Geneious Pro 5.6.2*. Available from http://www.geneious.com/.

Dunning LT, Savolainen V (2010) Broad-scale amplification of *matK* for DNA barcoding plants, a technical note. *Botanical Journal of the Linnean Society*, **164**, 1–9.

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

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

Fazekas AJ, Steeves R, Newmaster SG (2010) Improving sequencing quality from PCR products containing long mononucleotide repeats. *BioTechniques*, **48**, 277–285.

Felsenstein J (1993) PHYLIP—Phylogeny Inference Package (Version 3.2). *Cladistics*, **5**, 164–166.

Fernández-Palacios O (2009) *Biodiversidad morfológico-reproductiva y genética del género endémico canario* Parolinia *Webb (Brassicaceae)*. PhD Thesis, Universidad de Las Palmas de Gran Canaria (http://acceda.ulpgc.es/handle/10553/7922).

Fernández-Palacios JM, Martín-Esquivel JL eds (2001) Naturaleza de las Islas Canarias, Turquesa, Santa Cruz de Tenerife.

Ferreira R, Piredda R, Bagnoli F *et al.* (2011) Phylogeography and conservation perspectives of an endangered Macaronesian endemic: *Picconia azorica* (Tutin) Knobl. (Oleaceae). *European Journal of Forest Research*, **130**, 181–195.

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

Francisco-Ortega J, Santos-Guerra A, Hines A *et al.* (1996) Molecular evidence for a Mediterranean origin of the Macaronesian endemic genus *Argyranthemum* (Asteraceae). *American Journal of Botany*, **84**, 1595–1613.

Francisco-Ortega J, Barber JC, Santos-Guerra A *et al.* (2001) Origin and evolution of the endemic genera of Gonosperminae (Asteraceae: Anthemideae) from the Canary Islands: evidence from nucleotide sequences of the internal transcribed spacers of the nuclear ribosomal DNA. *American Journal of Botany*, **88**, 161–169.

Gao T, Sun Z, Yao H *et al.* (2011) Identification of Fabaceae plants using the DNA barcode *matK*. *Planta Medica*, **77**, 92–94.

García-Maroto F, Mañas-Fernández A, Garrido-Cárdenas JA *et al.* (2009) Δ^6-Desaturase sequence evidence for explosive Pliocene radiations within the adaptive radiation of Macaronesian *Echium* (Boraginaceae). *Molecular Phylogenetic and Evolution*, **52**, 563–574.

González-González EA (2011) *Análisis de la diversidad genética en poblaciones naturales de especies vegetales amenazadas:* Ilex perado *ssp*. lopezlilloi *(Aquifoliaceae),* Silene nocteolens *(Caryophyllaceae) y* Sorbus aria *(Rosaceae)*. PhD Thesis, Universidad de Las Palmas de Gran Canaria (http://acceda.ulpgc.es/handle/10553/6709).

González-Pérez MA, Sosa PA, Rivero E *et al.* (2009) Molecular markers reveal no genetic differentiation between *Myrica rivas-martinezii* and *M. faya* (Myricaceae). *Annals of Botany*, **103**, 79–86.

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

Hall T (2007) BioEdit, version 7.0.9. Carlsbad, CA: Computer Program and documentation. *Ibis Biosciences*, http://www.mbio.ncsu.edu/bioedit/bioedit.html.

Hansen LB, Siegismund HR, Jorgensen RB (2003) Progressive introgression between *Brassica napus* (oilseed rape) and *B. rapa*. *Heredity*, **91**, 276–283.

Harrison RG (1990) Hybrid zones: windows on evolutionary process. *Oxford Surveys in Evolutionary Biology*, **7**, 69–128.

Hebert PDN, Cywinska A, Ball SL *et al.* (2003) Biological identification through DNA barcodes. *Proceedings of the Royal Society of Biological Sciences Series B*, **270**, 313–321.

Hebert PDN, Penton EH, Burns JM *et al.* (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences USA*, **101**, 14812–14817.

Hebert PDN, de Waard JR, Landry JF (2010) DNA barcodes for 1/1000 of the animal kingdom. *Biology Letters*, **6**, 359–362.

Humphries CJ (1976) A revision of the Macaronesian genus Argyranthemum Webb ex Schultz-Bip. (Compositae-Anthemideae). Bulletin of the British Museum (Natural History) Botany 5, 147–240.

Hoban SM, McCleary TS, Schlarbaum SE *et al.* (2009) Geographically extensive hybridization between the forest trees American butternut and Japanese walnut. *Biology Letters*, **5**, 324–327.

Hollingsworth PM (2008) DNA barcoding plants in biodiversity hot spots: progress and outstanding questions. *Heredity*, **101**, 1–2.

Hollingsworth PM (2011a) Refining the DNA barcode for land plants. *Proceedings of the National Academy of Sciences USA*, **108**, 19451–19452.

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

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

Jaén-Molina R, Marrero Rodríguez A, Reyes Betancourt JA *et al*. (2010) La Flora Endémica del Parque Nacional de Garajonay bajo la Perspectiva Molecular: Las secuencias de ADN como herramienta en la identificación taxonómica. In: *Proyectos de Investigación en Parques Nacionales: 2006–2009* (eds Ramírez L & Asensio B), pp. 249–273. Organismos de Parques Nacionales, España.

Jalas J, Suominen J (1991) *Atlas Florae Europaeae (Paeoniaceae to Capparaceae)*. Helsinki University Printing House, Helsinki, Finland.

Jiang Y, Ding C, Zhang L *et al*. (2011) Identification of the genus *Epimedium* with DNA barcodes. *Journal of Medicinal Plants Research*, **5**, 6413–6417.

Kier G, Kreft H, Ming Lee T *et al*. (2009) A global assessment of endemism and species richness across island and mainland regions. *Proceedings of the National Academy of Sciences USA*, **106**, 9322–9327.51.

Kim SC, Crawford DJ, Francisco-Ortega J *et al*. (1996) A common origin for woody *Sonchus* and five related genera in the Macaronesian islands: molecular evidence for extensive radiation. *Proceedings of the National Academy of Sciences USA*, **93**, 7743–7748.

Kim SC, Crawford DJ, Francisco-Ortega J *et al*. (1999) Adaptative radiation and genetic differentiation in the woody *Sonchus* alliance (Asteraceae: Sonchinae) in the Canary Islands. *Plant Systematics and Evolution*, **215**, 101–118.

Kim SC, Lee C, Mejías JA (2007) Phylogenetic analysis of chloroplast DNA *matK* gene and *ITS* of nrDNA sequences reveals polyphyly of the genus *Sonchus* and new relationships among the subtribe Sonchinae (Asteraceae: Cichorieae). *Molecular Phylogenetics and Evolution*, **44**, 578–597.

Kreft H, Jetz W, Mutke J *et al*. (2008) Global diversity of island floras from a macroecological perspective. *Ecology Letters*, **11**, 116–127.

Kress WJ, Erickson DL (2007) A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *PLoS One*, **2**, e508.

Kress WJ, Wurdack KJ, Zimmer EA *et al*. (2005) Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences USA*, **102**, 8369–8374.

Kress WJ, Erickson DL, Jones FA *et al*. (2009) Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proceedings of the National Academy of Sciences USA*, **106**, 18621–18626.

Kruskal WH, Wallis WA (1952) Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, **47**, 583–621.

Kunkel G (1992) Flora y Vegetación del Archipiélago Canario. Tratado Florístico 2ª parte. Editorial Edirca. Las Palmas de Gran Canaria.

Kunkel G (1977) The *Ilex*: complex in the Canary Islands and Madeira. *Cuaderno de Botánica Canaria*, **28**, 17–19.

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

Leavitt DS, Esslinger TL, Hansen ES *et al*. (2013) DNA barcoding of brown *Parmeliae* (Parmeliaceae) species: a molecular approach for accurate specimen identification, emphasizing species in Greenland. *Organisms Diversity & Evolution*, **14**, 11–20.

Li DZ, Gao LM, Li HT *et al*. (2011) Comparative analysis of a large dataset indicates that internal transcribed spacer (*ITS*) should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences USA*, **108**, 19641–19646.

Little DP, Stevenson DW (2007) A comparison of algorithms for the identification of species using DNA barcodes: examples for gymnosperms. *Cladistics*, **23**, 1–21.

Losos JB, Ricklefs RE (2009) Adaptation and diversification on islands. *Nature*, **457**, 830–836.

Lou M, Golding BG (2010) Assigning sequences to species in the absence of large interspecific differences. *Molecular Phylogenetics and Evolution*, **56**, 187–194.

Marrero A (1992) Chromosomal evolutionary trends in the genus *Sideritis* subgenus *Marrubiastrum*. In: *Advances in Labiate Science* (eds Harley RM & Reynolds T), pp. 247–256. Royal Botanic Gardens, Kew.

Martín-Esquivel JL, Marrero N, Zurita N *et al*. (2005) Biodiversidad en gráficas. *Especies Silvestres de las Islas Canarias*, pp. 56. Consejería de Medio Ambiente y Ordenación Territorial, Gobierno de Canarias, Santa Cruz de Tenerife.

de Mattia F, Bruni I, Galimberti A *et al*. (2011) A comparative study of different DNA barcoding markers for the identification of some members of Lamiacaea. *Food Research International*, **44**, 693–702.

Meier R (2008) DNA sequences in taxonomy: opportunities and challenges. In: *The New Taxonomy* (ed. Wheeler QD), Chapter 7, pp. 95–127. CRC Press, New York, New York.

Meier R, Shiyang K, Vaidya G *et al*. (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology*, **55**, 715–728.

Mesa Coello R, Marrero Gómez MV, Romero Manrique P, Oval JP (2003) *Apollonias barbujana subsp. ceballosi* (Svent.) G. Kunkel. In: *Atlas y Libro Rojo de la Flora Vascular Amenazada de España* (eds Bañares Á, Blanca G, Güemes J, Moreno JC & Ortiz S), pp. 1072. Dirección General de Conservación de la Naturaleza, Madrid.

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

Molero J, Rovira AM (2005) Typification of some Macaronesian and Mediterranean dendroid spurges. *Taxon*, **54**, 472–474.

Molero J, Garnatje T, Rovira A *et al*. (2002) Karyological evolution and molecular phylogeny in Macaronesian dendroid spurges (*Euphorbia* subsect. *Pachycladae*). *Plant Systematics and Evolution*, **231**, 109–132.

Monaghan MT, Balke M, Gregory TR *et al*. (2005) DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers. *Philosophical Transactions of the Royal Society Biological Sciences*, **360**, 1925–1933.

Moreno Saiz JC (2011) La diversidad florística vascular española. *Memorias de la Real Sociedad Española de Historia Natural*, vol. **9**, pp. 75–107. Facultades de Biología y Geología. Universidad Complutense de Madrid, Madrid.

Mort ME, Soltis DE, Soltis PS *et al*. (2002) phylogenetics and evolution of the Macaronesian clade of Crassulaceae inferred from nuclear and chloroplast sequence data. *Systematic Botany*, **27**, 271–288.

Muñoz Garmendia F *et al*. (1993) Viola L. In: *Flora Ibérica: Plantas vasculares de la Península Ibérica e Islas Baleares* (eds Castroviejo S *et al*.), vol. **III**, pp. 276–317. Real Jardín Botánico de Madrid-CSIC, Madrid.

Newmaster SG, Fazekas AJ, Steeves RAD *et al*. (2008) Testing candidate plant barcode regions with species of recent origin in Myristicaceae. *Molecular Ecology Notes*, **8**, 480–490.

Ojeda I, Santos-Guerra A, Oliva-Tejera F *et al*. (2008) *Barcoding plant island radiation, what works and what does not. A case study in* Lotus *(Leguminosae) in the Macaronesia*. Abstract ID. 669, *Botany* 2008, Vancouver, British Columbia, Canada.

Palmer JD, Jansen RK, Michaels HJ *et al*. (1988) Chloroplast DNA variation and plant phylogeny. *Annals of the Missouri Botanical Garden*, **75**, 1180–1206.

Percy DM, Cronq Q (2002) Different fates of island brooms: contrasting evolution in *Adenocarpus*, *Genista* and *Teline* (Genisteae, Fabaceae) in the Canary Islands and Madeira. *American Journal of Botany*, **89**, 854–864.

Pérez de Paz PL (1978) Revisión del género Micromeria Bentham (Lamiaceae-Stachyoideae) en la región Macaronésica. In *Monografías del Instituto Estudios Canarios* (eds Cabildo Insular, Aula de Cultura), vol. **16**, 1–308 + XXXII láminas. Instituto de Estudios Canarios, La Laguna, Tenerife.

Pettengill JB, Neel MC (2010) An evaluation of candidate plant DNA barcodes and assignment methods in diagnosing 29 species in the

genus *Agalinis* (Orobanchaceae). *American Journal of Botany*, **97**, 1391–1406.
Pitard J, Proust L (1909) Les Iles Canaries. Flore de l'Archipel. *Librairie des sciences naturelles*. Paul Klincksieck, Paris.
Poiret JLM (1816) Encyclopédie méthodique. *Botanique* (Suppl. 4), pp. 472–473. Chez Mme. veuve Agasse, Imprimeur-Libraire, Paris.
Ragupathy S, Newmaster SG, Murugesan M *et al.* (2009) DNA barcoding discriminates a new cryptic grass species revealed in an ethnobotany study by the hill tribes of the Western Ghats in southern India. *Molecular Ecology Resources*, **9**(Suppl. 1), 164–171.
Ran JH, Wang PP, Zhao H-J *et al.* (2010) A test of seven candidate barcode regions from the plastome in *Picea* (Pinaceae). *Journal of Integrative Plant Biology*, **52**, 1109–1126.
Reid BN, Le M, McCord WP *et al.* (2011) Comparing and combining distance-based and character-based approaches for barcoding turtles. *Molecular Ecology Resources*, **11**, 956–967.
Reyes-Betancort JA, Santos Guerra A, Guma IR *et al.* (2008) Diversity, rarity and the evolution and conservation of the Canary Islands endemic flora. *Anales Jardín. Botánico de Madrid*, **65**, 25–45.
Rivas-Martínez S *et al.* (2002) Vascular plant communities of Spain and Portugal. *Itinera Geobotanica*, **15**, 433–922.
Rivas-Martínez S (2011) Mapa de series, geoseries y geopermaseries de vegetación de España: [Memoria del mapa de vegetación potencial de España]. Parte II. *Itinera Geobotanica*, **18**, 425–800.
Rodríguez-Sanchez F, Guzman B, Valido A *et al.* (2009) Late Neogene history of the laurel tree (Laurus L., Lauraceae) based on phylogeographical analyses of Mediterranean and Macaronesian populations. *Journal of Biogeography*, **36**, 1270–1281.
Ross HA, Murugan S, Sibon Li WL (2008) Testing the reliability of genetic methods of species identification via simulation. *Systematic Biology*, **57**, 216–230.
Sang T, Crawford DJ, Stuessy TF (1997) Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany*, **84**, 1120–1136.
Santos A (1988) Notas sobre *Euphorbia bourgeauana* Gay ex Boiss. In DC. y especies afines. *Botánica Macaronésica*, **16**, 29–36.
Savolainen V, Cowan RS, Vogler AP *et al.* (2005) Towards writing the encyclopedia of life: an introduction to DNA barcoding. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, **360**, 1811–1850.
Schaefer H, Moura M, Belo Maciel MG *et al.* (2011) The Linnean shortfall in oceanic island biogeography: a case study in Azores. *Journal of Biogeography*, **38**, 1345–1355.
Scholz H, Stierstorfer C, Gaisberg M (2000) *Lolium edwardii* sp. nova (Gramineae) and its relationship with Schedonorus sect. Plantynia Dumort. *Feddes Repertorium*, **111**, 561–565.
Seberg O, Petersen G (2009) How many loci does it take to DNA barcode a crocus? *PLoS One*, **4**, e4598.
Seubert M, Hochstetter C (1844) *Flora Azorica*, pp. 50. Bonnae : Apud A. Marcum, Bonn.
Smith MA, Poyarkov NA Jr, Hebert PDN (2008) *CO1* DNA barcoding amphibians: take the chance, meet the challenge. *Molecular Ecology Resources*, **8**, 235–246.
Soltis DE, Morgan DR, Grable A *et al.* (1993) Molecular systematics of Saxifragaceae *sensu stricto*. *American Journal of Botany*, **80**, 1056–1081.
Tutin TG (1933) Notes from the university herbarium, Cambridge. *Journal of Botany*, **71**, 99–101.
Van Hengstum T, Lachmuth S, Oostermeijer GB *et al.* (2012) Human-induced hybridization among congeneric endemic plants on Tenerife, Canary Islands. *Plant Systematics and Evolution*, **298**, 1119–1131.
de Vere N, Rich TCG, Ford CR *et al.* (2012) DNA barcoding the native flowering plants and conifers of Wales. *PLoS One*, **7**, e37945.
Weitschek E, Van Velzen R, Felici G *et al.* (2013) BLOG 2.0: a software system for character-based species classification with DNA Barcode sequences. What it does, how to use it. *Molecular Ecology Resources*, **13**, 1043–1046.
Werner O, Ros RM, Fernández Á (2007) Caracterización Genética de Poblaciones de Varias Especies Amenazadas en el Parque Nacional de Garajonay. *III Congreso de Biología de la Conservación de Plantas*. Puerto de la Cruz, España.
Whittaker RJ, Fernández-Palacios JM (2007) *Island Biogeography. Ecology, Evolution and Conservation*, 2nd edn. Oxford University Press, Oxford, UK.
Wiemers M, Fiedler K (2007) Does the DNA barcoding gap exist?—a case study in blue butterflies (Lepidoptera: Lycaenidae). *Frontiers in Zoology*, **4**, 8.
Yang J-B, Wang Y-P, Moller M *et al.* (2012) Applying plant DNA barcodes to identify species of Parnassia (Parnassiaceae). *Molecular Ecology Resources*, **12**, 267–275.
Yao H, Song J, Liu C *et al.* (2010) Use of *ITS2* Region as the universal DNA barcode for plants and animals. *PLoS One*, **5**, e13102.
Zhang W, Fan X, Zhu S *et al.* (2013) Species-specific identification from incomplete sampling: applying DNA barcodes to monitoring invasive *Solanum* plants. *PLoS One*, **8**, e55927.
Zou S, Li Q, Kong L *et al.* (2011) Comparing the usefulness of distance, monophyly and character-based DNA barcoding methods in species identification: a case study of *Neogastropoda*. *PLoS One*, **6**, e26619.

R.J.M. performed research, analysed data, wrote the paper. A.M.R. performed research, wrote the paper. J.A.R.B. performed research, wrote the paper. A.S.G. performed research, wrote the paper. J.N.S. performed research. J.C.C. designed research, analysed data, performed research, wrote the paper.

Data accessibility

GenBank Accessions nos: Table S1 (Supporting information). Aligned sequence matrices and relevant ancillary information: *Demiurge* information system (URLs http://www.demiurge-project.org/matrix_digests/D-DNASE-89 and http://www.demiurge-project.org/matrix_digests/D-DNASE-90).

Supporting Information

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

Table S1 Taxa sampled for this study and sequences generated *de novo* (see Demiurge, D-DNASE 89 AND 90) with Genbank accession numbers and codes for the samples deposited in the DNA Bank of the Canarian Flora (JBCVCSIC) and localities where the plant material was collected (the island abbreviations correspond to G, La Gomera; P, La Palma; T, Tenerife; C, Gran Canaria; F, Fuerteventura; M, Madeira; A, Azores)

Table S2 Taxa (Macaronesian population or not) for which sequences were downloaded from Genbank (Accession numbers for each barcode region are shown)

Table S3 Taxa per genus included in this study, indicating the total number of the endemic and native species in the Canary Islands with respect to the taxa analyzed (following Arechavaleta M, Rodríguez S, Zurita N, García A (eds) (2010) *Lista de especies silvestres de Canarias. Hongos, plantas y animales terrestres. 2009*, 579 pp. Gobierno de Canarias.)