*Botanical Journal of the Linnean Society*, 2019, **190**, 374–388. With 4 figures.

# DNA barcodes fail to accurately differentiate species in Hawaiian plant lineages

JEFFERY K. STALLMAN[1,*], VICKI A. FUNK[2], JONATHAN P. PRICE[3] and MATTHEW L. KNOPE[1]

[1]*Department of Biology and Tropical Conservation Biology and Environmental Science Graduate Program, University of Hawai'i Hilo, Hilo, Hawai'i 96720, USA*
[2]*Department of Botany, Smithsonian Institution, Washington DC 20013–7012, USA*
[3]*Department of Geography and Environmental Science and Tropical Conservation Biology and Environmental Science Graduate Program, University of Hawai'i Hilo, Hilo, Hawai'i 96720, USA*

*Received 26 February 2019; revised 22 April 2019; accepted for publication 3 June 2019*

DNA barcoding has been largely successful in differentiating animal species, but the most effective loci and evaluative methods for plants are still debated. Floras of young, oceanic islands are a challenging test of DNA barcodes, because of rapid speciation, high incidence of hybridization and polyploidy. We used character-based, tree-based and genetic distance-based methods to test DNA barcoding of 385 species of native Hawaiian plants constituting 20 lineages at the nuclear ITS(2) locus, nine lineages at each of the plastid loci *trnH-psbA* and *rbcL*, eight lineages at the plastid locus *matK* and four lineages with concatenated data. We also incorporated geographical range information and tested if varying sample sizes within a lineage influenced identification success. Average discrimination success was low (22% maximum) with all methods of analysis across all loci. The character-based method generally provided the highest identification success, there were limited benefits from incorporating geographical data and no relationship between number of species sampled in a lineage and identification success was found. Percentages of identification success are the lowest reported in a DNA barcoding study of comparable scale, and multi-species groups that radiated in the Hawaiian archipelago probably cannot be identified based on current DNA barcoding loci and methodologies.

KEYWORDS: adaptive radiation – biogeography – DNA fingerprinting – endemic – island biology.

## INTRODUCTION

Since the development of the use of molecular data for species identification (DeSalle & Birstein, 1996) and the advent of the term 'DNA barcoding' (Hebert, Cywinska & deWaard, 2003a), single-locus barcodes have been proposed and widely adopted for animals (Hebert, Ratnasingham & deWaard, 2003b) and fungi (Schoch *et al.*, 2012); however, the best loci and methods for identifying land plants are still vigorously debated. Commonly used loci for DNA barcoding of plants include the Consortium for the Barcode of Life (CBOL) plant working group recommendation of the plastid coding regions *matK* and *rbcL* (CBOL, 2009), the most popular non-coding plastid region *trnH-psbA* [proposed by Kress & Erickson (2007) as a barcode with *rbcL*], and the nuclear locus ITS (Chen *et al.*, 2010; China Plant BOL Group, 2011), although other regions have also been proposed (e.g. Dong *et al.*, 2015). Sequences from these loci are usually incorporated into character, tree (monophyly) or distance-based identification schemes (with variation within each method). Without a universally agreed barcode(s) or identification method(s), studies have generally focused on testing specific taxonomic groups and/or geographical regions, while comparing different loci and/or methodologies (e.g. Zou *et al.*, 2011; Li, Tong & Xing, 2016; Liu, Yan & Ge, 2016) to determine a 'best fit' for a particular geographical region or taxonomic group of interest. Results have been highly variable, with analyses across geographical regions and taxonomic groups ranging from 56% identification success in a wide range of taxonomic groups in the Canadian Arctic (Saarela *et al.*, 2013) to 93% in medicinal plants from around the globe (Chen *et al.*, 2010); more targeted geographical studies have achieved up to 97% identification success

*Corresponding author. E-mail: jstall@hawaii.edu

 

from broad-taxonomic sampling in a nature reserve in Ontario, Canada (Burgess *et al.*, 2011). However, in contrast to these broad-scale taxonomic studies, more targeted studies in richly sampled, closely related groups have found species identification success rates as low as essentially 0% in *Solanum* L. section *Petota* Dumort. from the Americas (Spooner, 2009), 1.4% in *Salix* L. across the Holarctic (Percy *et al.*, 2014) and 30.8% in Indian *Berberis* L. (Roy *et al.*, 2010).

Many variables appear to have predictable effects on the outcome of DNA barcoding studies. For example, it has been demonstrated that restricting the geographical range of a DNA barcoding study generally increases identification success (Hollingsworth *et. al.*, 2016) as sample size decreases and species comparisons are limited. However, decreased identification success is expected in groups that have recently speciated, where many closely related species co-occur, that hybridize or where polyploidy is common (Cowan & Fay, 2012). Recently speciated taxa and closely related species may not have had time for diagnostic mutations to arise or may not exhibit polymorphisms in the DNA barcoding loci commonly examined. Hybrids present challenges where (usually) maternally inherited plastid DNA would not provide any evidence of a different paternal species, and polyploids can present difficult-to-interpret nuclear DNA sequences if many alleles of the region examined are present. All four of these challenges to DNA barcoding occur with high frequency in the Hawaiian flora (Keeley & Funk, 2011).

Hawai'i is the most isolated archipelago on earth, > 3500 km from the next-nearest landmass, and all indigenous organisms successfully arrived and colonized the islands via wind, ocean travel or transport on other organisms (e.g. birds) (Ziegler, 2002). The native Hawaiian angiosperm flora consists of a disharmonic community of *c.* 1000 species (Wagner, Herbst & Sohmer, 1999; Price & Wagner, 2018), with many common (sub)tropical groups being poorly represented (e.g. no native species of Zingiberaceae), and other families over-represented (e.g. 124 species of Campanulaceae) due to *in situ* adaptive radiation of a small number of successful lineages (for a review see Keeley & Funk, 2011). The well-known geological history of Hawai'i, a linear chronosequence of oceanic islands of volcanic origin with the current high islands all under *c.* 5.1 Myr old (Clague, 1996; Price & Clague, 2002), has made it an excellent location for studies on speciation, adaptive radiation and historical biogeography. The biogeographies and phylogenies of many well-known Hawaiian plant radiations [e.g. Hawaiian Madiinae (silversword alliance), lobelioids and *Bidens* L.] have been explored, first through morphology (e.g. Carlquist, 1976) and, beginning in the late 1980s, through DNA sequence data (Witter & Carr, 1988). Molecular studies on the Hawaiian flora have continued to the present day, even incorporating entire plastomes in attempts to resolve species relationships within the archipelago and identify the origin of the ancestors that established certain Hawaiian plant lineages (e.g. Welch *et al.*, 2016). Because of these studies, molecular data from loci commonly used in DNA barcoding analyses exist for many Hawaiian plant lineages, but only one study (Pillon *et al.*, 2013) attempted to identify the Hawaiian flora through DNA barcodes, finding success rates up to only 20% in two genera using a method requiring unique haplotypes for identification success. In contrast, a broader taxonomic study of the flora of the Canary Islands (Jaén-Molina *et al.*, 2015), also a relatively young volcanic archipelago with high incidence of adaptive radiation (Bramwell, 1975), found a high rate of identification success (82%) using a character-based identification method.

The DNA barcoding of plants has much-discussed potential and realized uses (e.g. Kress *et al.*, 2005) and may be helpful in plant conservation in Hawai'i where 9% of the native flora is already extinct and at least 53% is at risk (Sakai, Wagner & Mehrhoff, 2002). The Hawaiian Islands are botanically well-explored, but it is not uncommon that populations of species thought extinct or extirpated are rediscovered in inaccessible or remote areas (e.g. Douglas & Shaw, 1989; Wood, 2007). Rediscovery of a species, or identification of new populations of species of concern may aid in their recovery, but in some cases identification of plants with overlapping morphologies and geographical ranges is not possible if plants are immature, or diagnostic features such as flowers or fruits are not present. To illustrate one example, *Melanthera venosa* (Sherff) W.L.Wagner & H.Rob., an endangered species, and *M. subcordata* (A.Gray) W.L.Wagner & H.Rob., a species of least concern, co-occur on Hawai'i Island and are morphologically similar to one another (Wagner *et al.*, 1999). DNA barcoding could potentially provide a more rapid identification of the species with a small piece of tissue without the requirement of a trained taxonomist examining a mature, flowering plant.

To test the identification success of DNA barcoding across the Hawaiian flora, we used the four most commonly employed loci [ITS(2), *trnH-psbA*, *matK* and *rbcL*] and the three most common identification methods (based on molecular characters, phylogenetic trees or genetic distances) on 1315 DNA sequences from 26 native Hawaiian plant lineages across the eight main Hawaiian Islands. New sequences (Supporting Information, Table S1) were generated from the four Hawai'i Island lineages of Asteraceae: *Bidens* [small trees to herbaceous annuals; estimated *c.* 2 Myr old

(Knope *et al.*, 2012)]; Hawaiian Madiinae [small trees to cushion plants; estimated 5.2 Myr old (Baldwin & Sanderson, 1998)]; *Melanthera* Rohr (subshrubs; no estimated lineage age); and *Tetramolopium* Nees [shrubs to subshrubs; estimated > 1 Myr (Lowrey, Whitkus & Sykes, 2005)]. Other examined lineages are variable in habit, habitat and lineage size, from *Plantago* L. (three species; herbs to subshrubs; no estimated lineage age) to *Melicope* J.R.Forst. & G.Forst. (54 species; shrubs to trees; no estimated lineage age).

Because the distribution of Hawaiian plant taxa is well-documented, we also incorporated geography at a finer scale in our best-sampled lineage at each locus, limiting species comparisons to only those taxa found on the same island (eliminating extraneous comparisons from species that cannot co-occur) to determine how an island-by-island analysis would affect identification success. We hypothesized that: (1) overall identification success across loci and methods would be low due to the high incidence of recently diverged species, hybridization and polyploidy; (2) the character-based method would outperform the tree and distance-based methods due to its higher success in the flora of the Canary Islands; (3) lineages with smaller sample sizes, and therefore fewer species comparisons, would have higher percentages of identification success and (4) that incorporating geographical range information on an island-by-island basis would increase identification success as species comparisons are limited to only taxa occurring on the same island.

## MATERIAL AND METHODS

### Sampling strategy and molecular protocols

We sampled 40 individuals representing 19 species of all four multi-species Asteraceae lineages on Hawai'i Island (*Bidens*, the Hawaiian Madiinae, *Melanthera* and *Tetramolopium*) between February 2017 and February 2018. We collected two samples from every taxon except for *Melanthera lavarum* (Gaudich.) W.L.Wagner & H.Rob., which could not be located, *Bidens hillebrandiana* (Drake) O.Deg. ex Sherff, for which only one individual was accessible for sampling, and *Dubautia ciliolata* (DC.) D.D.Keck, for which four samples were collected. We also collected two samples of an undescribed *Tetramolopium* sp. (Schnell *et al.*, 2003) that was included in our analysis. Samples were collected by the protocol in Funk *et al.* (2017) and consisted of a cutting of a plant that was pressed, dried and mounted, and was deposited at the National Tropical Botanical Garden Herbarium (PTBG) on Kaua'i, or in the case of endangered species, a photographic voucher that is available by searching the Smithsonian National Museum of Natural History Department of Botany website by biorepository number (see Supporting Information, Table S1 for species sampled, PTBG accession numbers and biorepository numbers). Also collected was a sample of fresh leaf tissue that was desiccated in silica beads and sent to the Department of Botany and Laboratory for Analytical Biology, Smithsonian Institution, for DNA extraction and DNA barcode sequencing. For 18 of the 19 species, at least two unique individuals were sampled. Genes and gene regions sequenced include: ITS2, *trnH-psbA*, *matK* and *rbcL*. Samples were PCR amplified and sequenced on an ABI 3700 96-well capillary sequencer, totalling 160 new sequences (see Supporting Information, Table S1 for GenBank accession numbers and Supporting Information, Table S2 for primers used). Sequences were generated in both the 5′ to 3′ and 3′ to 5′ directions and were trimmed, edited and aligned into consensus contigs in Geneious v.9.1.8 (Kearse *et al.*, 2012).

In addition to these newly collected samples, we also downloaded all native Hawaiian Plant DNA sequences available in GenBank as of January 2018 at the ITS, *trnH-psbA*, *matK* and *rbcL* loci (see Supporting Information, Table S3 for taxa and accession numbers). For *Bidens*, *Melanthera, Tetramolopium* and the Hawaiian Madiinae lineages, only the ITS2 region was used for consistency with our newly generated sequences; we analysed the entire ITS region in all other lineages. Currently recognized species names, lineages and distributions were taken from the Smithsonian Botany Website (Wagner, Herbst & Lorence, 2005–continuously updated) as of January 2018, except in the cases of the Hawaiian lobelioids (Givnish *et al.*, 2009), *Cyrtandra* J.R.Forst. & G.Forst. (Johnson *et al.*, 2017), *Plantago* (Dunbar-Co, Wieczorek & Morden, 2008), *Hesperomannia* A.Gray (Morden & Harbin, 2013) and *Silene* L. (Eggens *et al.*, 2007), in which updated lineage information was taken directly from these studies.

Sequences from plant lineages lacking at least two species represented by at least two unique accessions per locus for comparison could not be examined with our methods (see 'Analysing Barcodes' section) and were excluded from analysis. We use the term lineage instead of genus to refer to our analysis groups because in multiple cases a single colonizer speciated into several genera (Hawaiian lobeloids, Hawaiian Madiinae and Hawaiian Lamioideae lineages), and in some cases, species from the same genus colonized Hawai'i in separate events (*Euphorbia* L., *Scaevola* L.). Species derived from the same lineage (colonization event) are analysed together irrespective of their generic placement [e.g. all species in different genera composing the Hawaiian Madiinae represent a single lineage (Baldwin & Sanderson, 1998) and are analysed

together, whereas *Scaevola taccada* (Gaertn.) Roxb. and *S. glabra* Hook. & Arn. are excluded from our analysis of *Scaevola* because they represent separate, non-radiating colonization events (Howarth *et al.*, 2003)]. Alignments of sequences within lineages were made using the default settings of Multiple Alignment using Fast Fourier Transform (MAAFT; Katoh & Standley, 2013) in Geneious. Alignments were unambiguous and manual editing beyond trimming ends to ensure elimination of primers and full coverage of all taxa with two accessions was not required.

In total, 1315 sequences (1155 from GenBank and 160 newly generated) representing 24 unique lineages, 18 families, 33 genera and 385 species were included in 46 analysis groups (e.g. the lineage *Cyrtandra* was analysed separately at the ITS and *trnH-psbA* loci). Using information on total lineage numbers from Price & Wagner (2004), the 24 unique lineages represent 26% of the total multi-species lineages reported (91 lineages), and the 385 unique species represent 46% of native Hawaiian plants that are members of multi-species lineages (832 species).

### Analysing barcodes

We used character-based, tree-based and genetic distance methods to determine whether species could be accurately identified within each lineage at each individual locus and with the concatenated data set of all four loci for the newly generated sequences from the Asteraceae of Hawai'i Island (the species for which we had sequence data for the same individuals at all four loci). Ambiguity codons were allowed in the analysis at the nuclear locus, ITS(2), to account for heterozygotes, but excluded for the haploid plastid loci. Uncorrected *P* distances were used for the distance and tree analyses, as the often utilized K2P model (Kimura, 1980) of molecular evolution is generally no longer considered to be warranted in DNA barcoding studies (Amrita & Rudolph, 2012; Collins *et al.*, 2012). In addition, we also incorporated species geographical range data for the best-sampled lineage at each locus to test if adding known geographical constraints (island occurrences) among species helps DNA barcoding identification success.

For the character-based analysis, we looked for diagnostic characters or positions in the alignment (nucleotide bases and/or indels) that were shared by all accessions (individuals) of a single species, but no other species in that lineage. If one or more diagnostic character(s) were found, we considered that species correctly identified (following Jaén-Molina *et al.*, 2015). We also counted the number of diagnostic characters at each locus, normalized these counts by average sequence length and number of species with two accessions present at that locus, tested for normality and ran a Kruskal–Wallis test in R v.3.4.3 (R Core Development Team, 2017) to test for significant differences in median number of diagnostic characters among loci.

For the tree-based method, we created a consensus tree from our lineage alignments in Geneious using the neighbor-joining method (Saitou & Nei, 1987), with a Jukes and Cantor distance model (Jukes & Cantor, 1969), a random seed generator, 1000 replicates and an appropriate outgroup species (see Supporting Information, Table S4 for outgroup taxa and GenBank accession numbers). If all conspecific sequences formed a monophyletic clade with bootstrap support of 70% or greater, usually corresponding to a 95% probability (Hillis & Bull, 1993), we considered that species to be correctly identified. Although there are valid critiques of the use of neighbor-joining trees in DNA barcoding studies (e.g. Collins & Cruikshank, 2012), it is used here for comparative purposes as the majority of plant DNA barcoding studies from the earliest (Hebert *et al.*, 2003a) to the most recent (Gong *et al.*, 2018) often use this method.

For the distance-based method, we determined whether there was a 'barcoding gap' (Meyer & Paulay, 2005), a difference between the greatest intraspecific distance and closest interspecific distance among species within a lineage using the 'extreme pairwise comparison' tool in Species Identifier v.1.8 (Meier *et al.*, 2006). If a species maximum intraspecific *P* distance was lower than the minimum interspecific *P* distance compared to all other individuals within its lineage, we considered that species to be correctly identified (Hebert *et al.*, 2004).

We calculated Pearson's correlation coefficient on the relationship between percent identification success in each lineage based on: (1) the total number of species sampled in a lineage and (2) the total number of species sampled in a lineage with two or more accessions. To see if identification success differed in better-sampled lineages, we analysed a subset of the data, including only lineages with five or more species with at least two accessions per species, or lineages with 50% or more of the species in that lineage with at least two accessions.

To incorporate geography, we used range information from the Smithsonian website (Wagner *et al.*, 2005–continuously updated) for the best-sampled lineage at each locus: *Scaevola* (excluding two, non-radiating colonizations) at ITS, *Melicope* at *trnH-psbA*, *Pritchardia* Seem. & H.Wendl. at *matK* and *Geranium* L. at *rbcL*, respectively. To analyse each island individually, a new alignment of accessions was created for each island, only including species occurring on that island, so our three methods of analysis could be completed without the influence of species not known to co-occur on the same island. The total number of successful

identifications across all islands were tallied and visualized in ArcGIS v.10.4.1 (ESRI, 2012). See Supplementary Information, Table S5 for taxa used in the geographical analysis, number of accessions used and distribution by island.

For each analysis method, at least two accessions representing two individuals per taxon were required for a positive identification, whether to calculate an intraspecific genetic distance, calculate a bootstrap support value or to consider a position in an alignment as a diagnostic character. Species represented by only one accession were included in each analysis method because excluding them may have inflated our identification success (excluded species may have a low interspecific distance, cluster in the same terminal branch of a tree or share characters with species represented by two or more accessions), but when calculating the percentage of positive identification, only species that could be identified based on two accessions (individuals) were considered.

## RESULTS

### Hawai‘i Island Asteraceae data

The four lineages of Asteraceae collected contain 77 species throughout the archipelago and 20 species on Hawai‘i Island, 19 (25% from the archipelago, 95% from Hawai‘i Island) of which were sampled, with at least two accessions for 18 (23% from the archipelago, 90% from Hawai‘i Island). Three species (17%) were correctly identified using character, one species (6%) using tree and three species (17%) using the genetic distance-based method and concatenated sequences (see Fig. 1 for overall percentages; Fig. 2E for results by lineage). On average across all lineages with the concatenated dataset, the character-based method correctly identified 16% of species in each lineage, tree method 4% and distance method 16% (absolute species numbers not applicable to averages across lineages).

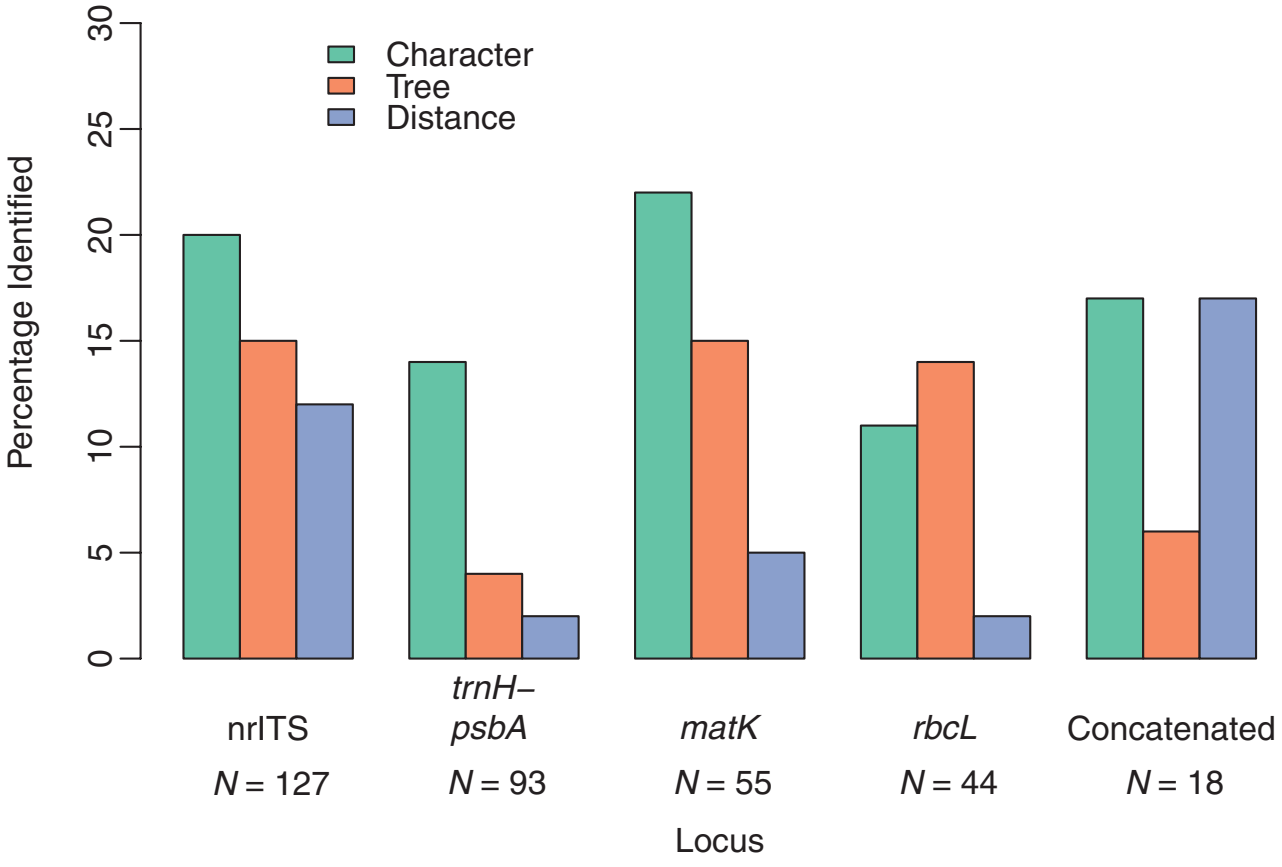


**Figure 1.** Overall percentage identification success by locus and barcoding method used. Sample sizes are the total number of species with at least two accessions at that locus. Overall, the ITS locus is represented by 20 lineages, *trnH-psbA* by nine, *matK* by eight, *rbcL* by nine and concatenated loci by four. The ITS and concatenated columns contain four Hawai‘i Island lineages of Asteraceae that only include the ITS2 region.

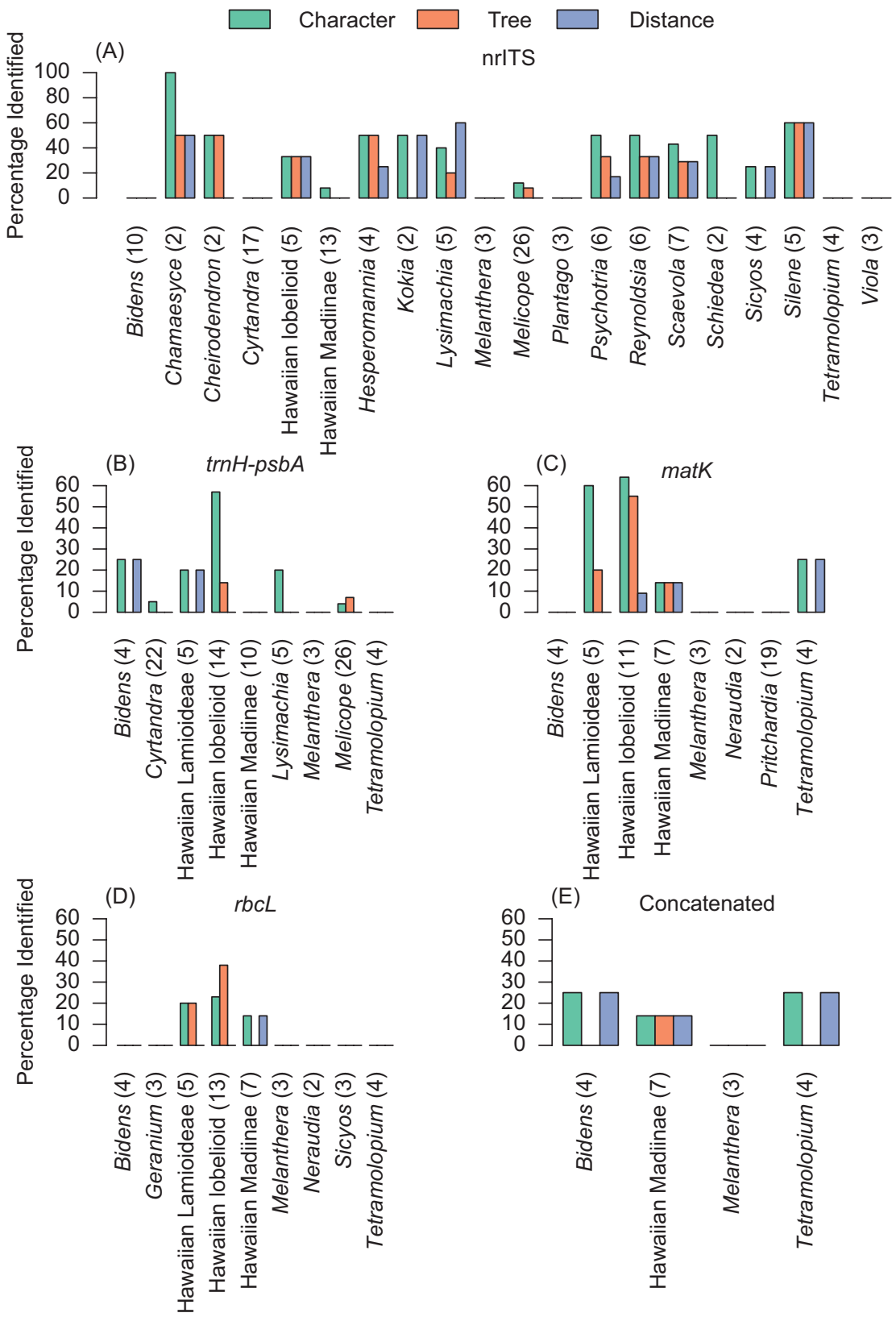


**Figure 2.** Percentage identification success at each lineage at: A, the ITS; B, *trnH-psbA*; C, *matK*; D, *rbcL* and E, the concatenated Hawai'i Island Asteraceae. Numbers in parentheses indicate the total number of species with at least two accessions in that lineage. At the ITS locus, the lineages *Bidens*, Hawaiian Madiinae, *Melanthera* and *Tetramolopium* are represented by the ITS2 region only.

### GenBank barcode data

In addition to the barcode data generated for native Hawai‘i Island Asteraceae, 1155 sequences from 24 unique lineages (20 ITS, nine *trnH-psbA*, eight *matK*, nine *rbcL*) of native Hawaiian plants were downloaded from GenBank (Clark *et al.*, 2015). The 20 lineages at the ITS(2) locus comprise a total of 468 species throughout the Hawaiian archipelago, 277 (59%) of which were sampled, with at least two accessions for 129 species (28%). In total, 26 species (20%) were identified using character, 19 species (15%) using tree and 16 species (12%) using genetic distance-based methods (Figs 1, 2A). On average across all lineages at the ITS(2) locus, the character-based, tree and distance methods identified 31, 19 and 19% of plant species in each lineage, respectively.

The nine lineages at the *trnH-psbA* locus contain 404 species, 154 (38%) of which were sampled, with at least two accessions for 95 species (24%). In total, 13 species (14%), four (4%) and two (2%) were correctly identified using character, tree and genetic distance-based methods, respectively (Figs 1, 2B). On average across all lineages at the *trnH-psbA* locus, the character-based, tree and genetic distance methods correctly identified 14, 2 and 5% of plant species in each individual lineage, respectively (higher than the percentage of total species identified since these identification successes were in lineages with low sample size).

The eight lineages at the *matK* locus contain a total of 305 species, 76 (25%) of which were sampled, with at least two accessions for 55 species (18%). In total, 12 (22%), eight (15%) and three species (5%) using character, tree and genetic distance-based methods, respectively, were correctly identified (Figs 1, 2C). On average across all lineages at the *matK* locus, the character-based, tree and genetic distance-based methods correctly identified 20, 11 and 6% of plant species in each lineage, respectively.

The nine lineages at the *rbcL* locus contain a total of 303 species, 85 (28%) of which were sampled, with at least two accessions for 44 species (15%). In total, five species (11%) were correctly identified using character, six (14%) using tree and one (2%) using distance (Figs 1, 2B). On average across all lineages at the *rbcL* locus, the character-based, tree and genetic distance-based methods correctly identified 6, 7 and 2% of plant species in each individual lineage, respectively.

Our diagnostic character count data was not normally distributed; therefore the untransformed data (normalized by sequence length and number of species with two or more sequences present) was analysed with the non-parametric Kruskal–Wallis test. Results showed that the number of normalized diagnostic characters was not significantly different across the four loci tested ($H = 6.368$, d.f. = 3, $P = 0.095$), although on average, the ITS locus displayed the greatest number of diagnostic characters (Fig. 3).

We found no relationship between percentage identification success and the total number of species sampled in a lineage (Pearson correlation coefficient $r = 0.058$, $R^2 = 0.003$, $N = 46$, $P = 0.702$), nor in the relationship between percentage identification success and the total number of species with two or more accessions sampled in a lineage (Pearson correlation coefficient $r = -0.132$, $R^2 = 0.017$, $N = 46$, $P = 0.384$). In addition, limiting our analysis to lineages with at least five species with two accessions, or > 50% of species with two accessions in a lineage had minimal effect on overall identification success (Supporting Information, Fig. S1). Considering the average percent identification success using the best method at each locus, success with ITS decreased by 1% in the better-sampled groups, *trnH-psbA* remained the same and *matK* and *rbcL* increased identification success by 4 and 7%, respectively.

### Incorporating geographical range data

We used the *Scaevola* lineage, of which we had samples representing all nine species (100%) and at least two accessions for eight of the nine species (89%) at the ITS locus, to incorporate geographical analyses throughout six of the main Hawaiian Islands (excluding Ni‘ihau and Kaho‘olawe). Counting individual species occurrence on each island separately, there are 25 total occurrences, of which we have sequences for 25 (100%) and at least two accessions for 24 (96%). In total, considering each occurrence separately, 17 (71%), 13 (54%) and 11 (46%) species occurrences were identified using character, tree and genetic distance-based methods, respectively (Fig. 4B). On average, the character-based method identified 72% of *Scaevola* occurrences on each island, compared with 58% (tree) and 46% (genetic distance).

We used the *Melicope* lineage, of which we had samples representing 38 of 54 species (70%) and at least two accessions for 28 (52%) at the *trnH-psbA* locus to incorporate geographical analyses throughout six of the main Hawaiian Islands (excluding Ni‘ihau and Kaho‘olawe). Counting individual species occurrence on each island separately, there are 83 total occurrences, of which we have sequences for 62 (75%) and at least two accessions for 50 (60%). In total, considering each occurrence separately, ten occurrences (20%) were identified using character, four (8%) using the tree and one (2%) using the genetic distance-based method (Fig. 4D). On average, the character-based method identified 28% of *Melicope* occurrences on each island, compared with 8% (tree) and 1% (genetic distance).

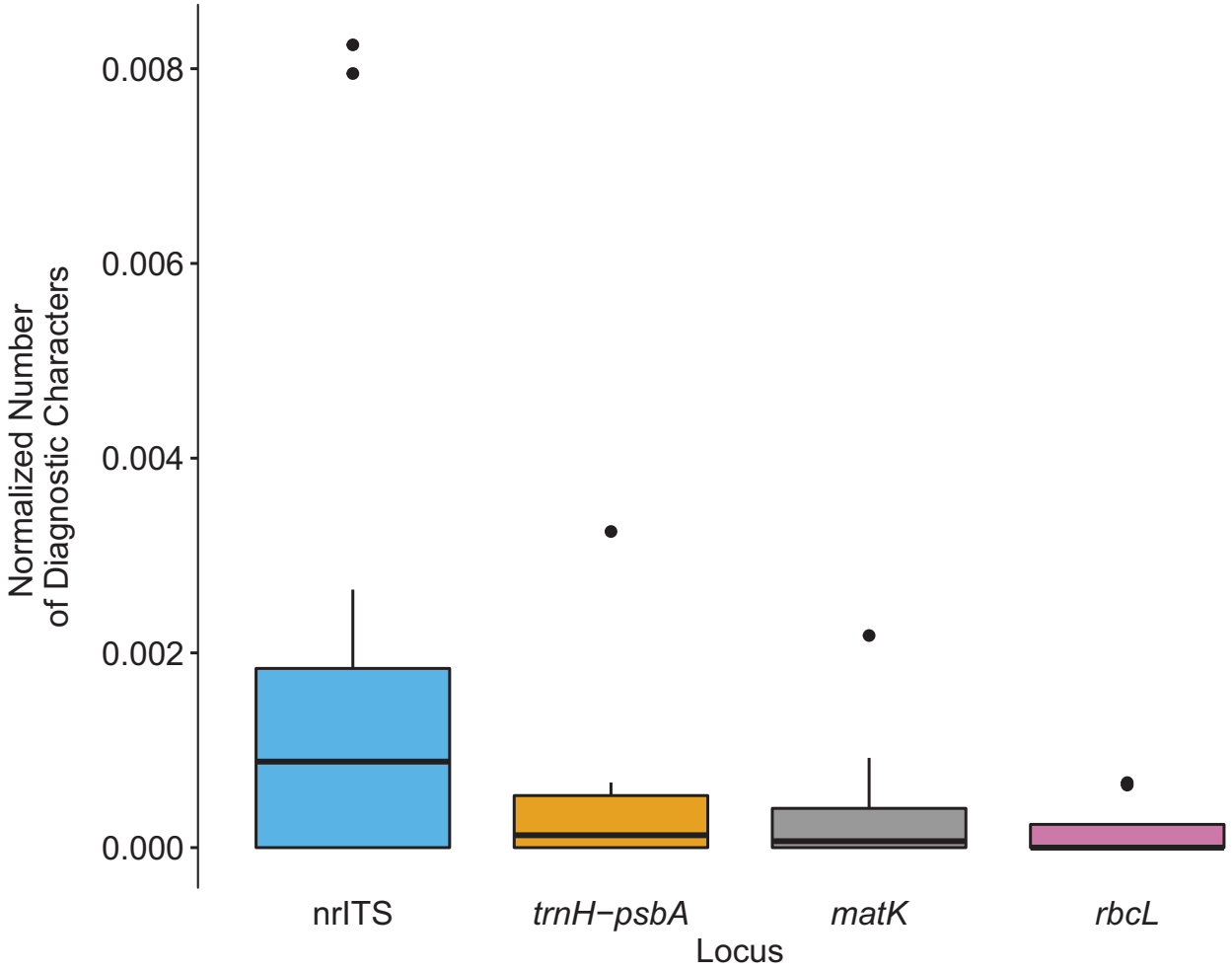


**Figure 3.** Boxplot of diagnostic characters per lineage at each locus normalized by sequence length and number of individuals with two accessions at each lineage; Kruskal–Wallis ($H$ = 6.368, d.f. = 3, $P$ = 0.095).

We used the *Pritchardia* lineage, of which we had samples representing 19 of 22 species (86%) and at least two accessions for 19 (86%) at the *matK* locus to incorporate geographical analyses across five of the main Hawaiian Islands (excluding Ni'ihau, Lanai and Kaho'olawe). Counting individual species occurrence on each island separately, there are 23 total occurrences, of which we have sequences for 19 (87%), and at least two accessions for 19 (87%). In total, considering each occurrence separately, one (5%), zero (0%) and zero (0%) occurrence were identified using character, tree and genetic distance-based methods, respectively (Fig. 4C). On average, the character-based method identified 5% of *Pritchardia* occurrences on each island, compared with 0%, (tree) and 0% (genetic distance).

We used the *Geranium* lineage, of which we had samples representing five of six species (83%) and at least two accessions for three (50%) at the *rbcL* locus to incorporate geographical analyses on Maui (the only island multiple species in this lineage are found). Four species are found on Maui, of which we had samples for four (100%), and at least two accessions for three (75%). Zero (0%) *Geranium* spp. were identified using any of the methods (Fig. 4A).

## DISCUSSION

As hypothesized (1), our average identification success across all loci and methods was low: no method achieved > 22% average identification success. This is the lowest result we are aware of in a DNA barcoding study of this broad taxonomic scale in land plants, substantially lower than the 82% identification success for the Canary Islands and much lower than the next-largest study of this size we are aware of at 56% identification success in the Canadian Arctic (Saarela *et al.*, 2013). We hypothesize lower success than the Canary Island study because our sampling

in large, radiating groups was much more extensive, with 33 groups that had five or more species in our analysis versus two groups with five or more species in the Canary Island study (Jaén-Molina *et al.*, 2015). Additionally, despite the similarity of being a young, volcanic archipelago with many radiating lineages, the Canary Islands have current high islands that are up to 23 Myr old (van den Bogaard, 2013), four times as old as the eight main Hawaiian Islands (Price and Clague, 2002), and only one of our sampled lineages, the Hawaiian lobelioid group, is known to be older than 5 Myr (Givnish *et al.*, 2009).

As hypothesized (2), the character-based method outperformed tree and distance in averages across all loci (8% higher than the next highest method at ITS, 7% higher at *matK*, 10% higher at *trnH-psbA*), except at *rbcL* where it was 3% below the tree-based method and the concatenated dataset where it was equal to the distance-based method. This finding agrees with studies showing character-based methods outperforming distance or tree-based methods for insects or other animals (e.g. Rach *et al.*, 2008; Bergman *et al.*, 2009; Yassin *et al.*, 2010), but results are mixed with plants. The Canary Island study suggested some promise for character-based methods if coupled with appropriate taxa and molecular loci, or at least better results than other methods. However, Roy *et al.* (2010) did not find an increase in identification success with the character-based method in species-rich sampling of Indian *Berberis*. Due to the scarcity of studies using character-based methods (Taylor & Harris, 2012), particularly in plants, it is premature to draw general conclusions about the effectiveness of this method. However, with its potential promise and the option to use automated tools for analysis (Sarkar, Planet & Desalle, 2008; Weitschek *et al.*, 2013), we suggest future studies should test its utility in other groups.

In contradiction to our hypothesis (3), our analysis of a subset of better-sampled lineages and our Pearson's correlations between sample sizes and identification success show that identification success is not higher in lineages with smaller sample sizes. This contrasts with other studies showing a clear negative relationship between number of taxa in a group sampled and percent identification success (e.g. Parmentier *et al.*, 2013; Saarela *et al.*, 2013; Liu *et al.*, 2015). Our result does not lend support to the explanation that our study had a lower percent identification success than the Canary Island study due to higher sampling in large, radiating groups, but our analysis does contain several examples where higher sample size did reduce identification success. For example, *Dubautia plantaginea* Gaudich. (Hawaiian Madiinae) is successfully identified with the *rbcL*, *matK* and concatenated loci where sample size is low, but not at the better-sampled *trnH-psbA* or ITS loci where one of its sister species, *D. laxa* Hook. & Arn. (Baldwin, 1997), is included in analysis.

With regards to hypothesis (4), incorporating geography into our best-sampled lineages resulted in either no improvement or modest improvement in identification success. The character-based method in combination with geographical range information provided the best results at each lineage. Identification success in *Scaevola* at the ITS locus increased from 43 to 71% after incorporating geography, *Melicope* (*trnH-psbA*) increased from 4 to 20%, *Pritchardia* (*matK*) increased from 0 to 5% and *Geranium* (*rbcL*) remained at 0% after incorporating geography. The lineage *Pritchardia*, with 16 single-island endemic species, had little improvement despite greatly limiting species comparisons, whereas both *Scaevola* and *Melicope* exhibited the largest improvement but had the most overlapping species among islands (Fig. 4), suggesting that they would not benefit as much from this analysis. Additional testing is needed on more species groups from different regions of the world to determine whether incorporating geographical data at finer scales would be a useful addition to DNA barcoding studies in general. In our study, increase in identification success is modest to moderate compared to the additional data and analysis required, and was primarily effective in groups that already had some success without its inclusion (*Scaevola* and *Melicope*), whereas groups without any identifications before inclusions (*Pritchardia* and *Geranium*) were not improved at all or only minimally. Although limiting overall species comparisons increases the chances that species not identified in an archipelago-wide lineage analysis could be identified when sample size is restricted to a single island, it is likely that species co-occurring on the same island are on average more closely related, so the closest genetic relatives may remain in the analysis group, and this may explain why improvement in identification success was modest. Additionally, incorporating geography requires strictly defining the ranges of plants that are capable of long-distance dispersal and introduces an additional possibility of error that is difficult to quantify.

To place our study into context with the other DNA barcoding studies from the Hawaiian Islands, Pillon *et al.* (2013) provided the best comparison, finding similar results in percent identification success using a modified character-based analysis that considered species identified if they had a unique haplotype. Pillon *et al.* (2013) found the highest identification success for Hawaiian *Cyrtandra* was 20% with the plastid marker *trnH-psbA* and a sample size of 20 species, compared with our 5% identification success at the *trnH-psbA* locus with a sample size of 37 species. Their highest identification success, with *Clermontia* Gaudich., was

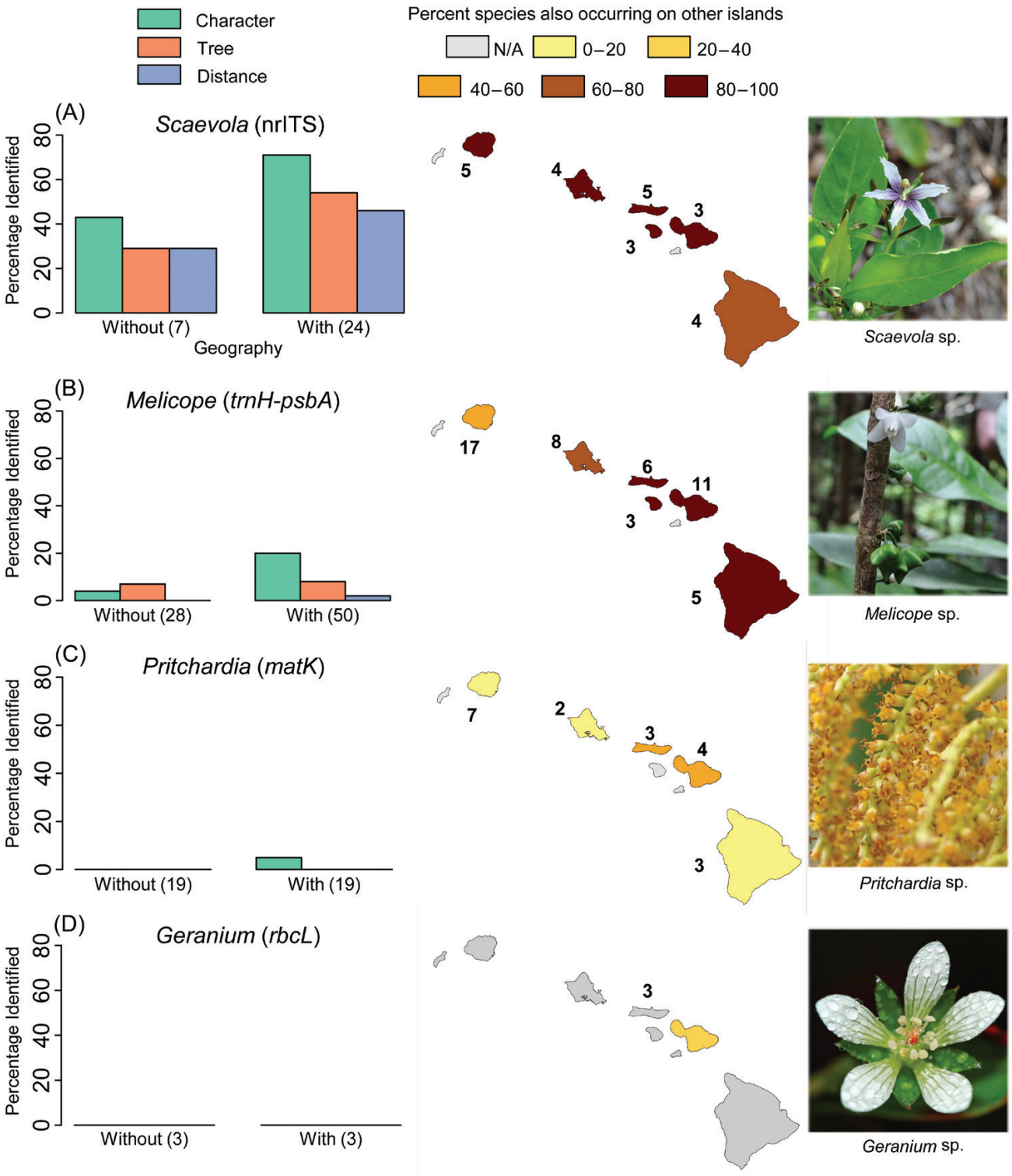


**Figure 4.** Identification success when including geographical information for the lineages: A, *Scaevola*; B, *Melicope*; C, *Pritchardia* and D, *Geranium*. Bar graphs indicate percentage identification success without (before) and with (after) incorporating geographical information; numbers in parentheses indicate the total number of species with two accessions considered in each analysis. On the maps of the Hawaiian Islands, darker colours indicate a higher percentage of species also occurring on other islands, and lighter colours indicate more single-island endemics (limiting the numbers of species comparisons and theoretically increasing identification success). Numbers adjacent to islands indicate the total number of species present on that island for which we have at least two accessions.

18% with concatenated *rbcL*, *matK* and *trnH-psbA* and a sample size of 17 species, compared with our 33% at *trnH-psbA* ($N$ = 17), 25% at *matK* ($N$ = 17) and 20% at *rbcL* ($N$ = 20) after we separated out *Clermontia* from the Hawaiian lobelioid lineage for a direct comparison and used the best-performing method (character-based). We attribute our lower identification success in *Cyrtandra* to our higher sample size, which is more likely to eliminate diagnostic characters or haplotypes (as discussed for *D. plantaginea*). Our *Clermontia* results are not drastically different, and our sample size of species with more than two accessions was small when breaking out only this genus from the larger Hawaiian lobelioid lineage ($N$ = 2 at ITS, $N$ = 4 at *matK*, $N$ = 6 at *trnH-psbA*, $N$ = 5 at *rbcL*), whereas Pillon *et al.* (2013) did not require multiple sequences from a single species to confirm diagnostic haplotypes. The other DNA barcoding studies at the species level completed in the Hawaiian Islands illustrate the discrepancy in DNA barcoding success between plants and animals, even when considered for the same archipelago. Magnacca & Brown (2010) found 71–100% identification success using a tree-based method and mtDNA in native *Hylaeus* bees depending on methods used and whether individuals exhibited mitochondrial heteroplasmy, even though the lineage is only 0.4–0.7 Myr old (Magnacca & Danforth, 2006), and Krehenwinkel *et al.* (2019) found 100% identification success (although some clades had lower support) using a tree-based method and long ribosomal DNA (including the ITS) in native *Tetragnatha* spiders.

Our results show that DNA barcoding using the currently available loci and methods cannot be suggested for any multi-species lineages in the Hawaiian flora. Our only lineage with 100% identification success, *Chamaesyce* Gray (now synonymized with *Euphorbia*; Horn *et al.*, 2012), was poorly sampled (50% total species sampled and 13% of species with two or more sequences), and our next highest results were 64% identification success in the Hawaiian lobelioids at the *matK* locus. Although a relationship between sample size and identification success was not found, based on other studies (e.g. Parmentier *et al.*, 2013; Saarela *et al.*, 2013; Liu *et al.*, 2015) and specific examples in our lineages (e.g. *Dubautia plantaginea*), it is unlikely identification success will remain high once sampling in these lineages increases. Our 24 lineages sampled represent 26% of the 91 multi-species lineages in the archipelago, but they comprise 385 species, or 46% of 832 species in multi-species lineages in Hawai'i (Price & Wagner, 2004). We note that for the four lineages of Asteraceae for which concatenated loci were available, no lineages had identification success higher than that which could be obtained from a single locus and the concatenated dataset among *Clermontia* from Pillon *et al.* (2013) had similar to lower identification success compared to our single-locus analyses in this group. Although many lineages remain to be sampled, and most lineages have not been tested with concatenated barcode regions, the evidence to date suggests lineages of native Hawaiian plants cannot be identified with levels of success comparable to other broad-taxonomic scale, but geographically-limited DNA barcoding studies, and use of barcodes with currently employed methodologies would be an unreliable tool to identify native taxa in multi-species lineages in Hawaii.

Pillon *et al.* (2013) suggested a two-million-year minimum before DNA barcodes become effective at identifying closely related species. Our study finds this estimate is too recent for the Hawaiian flora, and suggests that all Hawaiian plant lineages consisting of multi-species groups that radiated in the Hawaiian Islands, even extending to beyond the age of the main Hawaiian Islands (*c.* 5 Myr old) in the Hawaiian lobelioid lineage (*c.* 13 Myr old; Givnish *et al.*, 2009), cannot be successfully barcoded with the most commonly utilized loci and methodologies. The majority of Hawaiian plant lineages with estimated ages are found to have descended from a species colonizing the Hawaiian Islands after Kaua'i emerged *c.* 5 Mya (Keeley & Funk, 2011), after a period of *c.* 5–8 Myr when there were no islands > 1000 m high (Price & Clague, 2002). Even for those species that colonized the archipelago at an earlier time, such as the Hawaiian lobelioids, identification success is too low for reliable species identification at 44–64% success across all loci with the best-performing method and relatively low sample sizes. If a species colonized the archipelago before the emergence of the main Hawaiian Islands 5 Mya, it would also need to have speciated at an earlier date and have multiple species succeed in colonizing Kaua'i (difficult due to ecological and geological constraints) to have a time period > 5 Myr for potential diagnostic mutations to accumulate. Of all single-island endemics with two accessions that were sampled, only 4% from Hawai'i Island (the youngest island at *c.* 0.5 Myr old) were identified, 20% from Maui, 29% from Molokai, 9% on Oahu and 36% of those on Kaua'i (the oldest main Hawaiian Island at *c.* 5 Myr old).

The results of this study underscore the importance of taxonomy based on morphology when identifying the Hawaiian flora. To continue exploring the possibility of identifying Hawaiian land plants through molecular tools, sequencing entire plastomes shows promise in some recently radiated groups such as Hawaiian *Bidens* (Knope *et al.*, unpubl. data) and are useful in polyploid individuals, although data from complete plastome sequences was not sufficient to differentiate many species in a phylogenetic analysis (analogous to tree-based methodology) of the Hawaiian Lamioidae (Welch *et al.*, 2016). Finding additional single or

low-copy rapidly evolving nuclear loci is another option, although Pillon *et al.* (2013) had lower success with their four novel low-copy nuclear regions than with commonly employed plastid loci. Krehenwinkel *et al.* (2019) showed the utility of long ribosomal DNA in a variety of groups, but the effectiveness of this region in closely related groups of plants with potentially high ploidies is unknown. Incorporating simple ecological, morphological and geographical data to assist identification after using a DNA barcode to first narrow down a sample to a handful of possible species could potentially increase identification success, although the limits of incorporating geographical information are shown in this study, and the requirement for additional information beyond a small piece of tissue limits the overall utility of a DNA barcode. While the Hawaiian flora is thus far resistant to DNA barcoding, our results suggest that future studies may benefit from testing additional molecular loci, incorporating character-based methods and striving to sample closely related groups of species thoroughly to further examine the relationship between sample size and identification success to ensure results are not artificially inflated by omission of closely related taxa.

## ACKNOWLEDGEMENTS

We would like to thank Geneviève Blanchet, Heather Coad, Steven Evans, Pomai Lyman, Sierra McDaniel, Tara Myers, Lyman Perry, Jeremy Schrader, Joshua VanDeMark and the University of Hawai'i Hilo undergraduate students in Biology 481 course in Spring 2017 and 2018 who all assisted collecting specimens in the field. Thank you to Erin Datlof, who provided photographs of native Hawaiian plants for this manuscript and also collected specimens in the field. Thank you to Carol Kelloff, who coordinated supplies and collection logistics, and Jose Zúñiga for extraction, amplification, sequencing, annotation and submission to GenBank of DNA samples, both at the Smithsonian Institution. Thank you to the Hawai'i State Natural Area Reserve System, Hawai'i Volcanoes National Park, Kaloko-Honokōhau National Historical Park, Department of Forestry and Wildlife, US Fish & Wildlife Service, National Park Service, Plant Extinction Prevention Program and the Natural Resources Office at Pōhakuloa Training Area for securing access and collection permits. This material is based on work supported by the Global Genome Initiative under Grant Number GGI-Rolling-2016–077. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Global Genome Initiative.

## REFERENCES

**Amrita S**, **Rudolf M. 2012.** On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature. *Cladistics* **28:** 190–194.

**Baldwin BG. 1997.** Adaptive radiation of the Hawaiian silversword alliance: congruence and conflict of phylogenetic evidence from molecular and non-molecular investigations. In: Givnish TJ, Sytsma KJ, eds. *Molecular evolution and adaptive radiation*. Cambridge: Cambridge University Press, 103–127.

**Baldwin BG**, **Sanderson MJ. 1998.** Age and rate of diversification of the Hawaiian silversword alliance (Compositae). *Proceedings of the National Academy of Sciences, U.S.A.* **95:** 9402.

**Bergmann T**, **Hadrys H**, **Breves G**, **Schierwater B. 2009.** Character-based DNA barcoding: a superior tool for species classification. *Berl Munch Tierarztl Wochenschr* **122:** 446–450.

**van den Bogaard P. 2013.** The origin of the Canary Island Seamount Province – new ages of old seamounts. *Scientific Reports* **3:** 2107.

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

**Burgess KS**, **Fazekas AJ**, **Kesanakurti PR**, **Graham SW**, **Husband BC**, **Newmaster SG**, **Percy DM**, **Hajibabaei M**, **Barrett SCH. 2011.** Discriminating plant species in a local temperate flora using the *rbcL+matK* DNA barcode. *Methods in Ecology and Evolution* **2:** 333–340.

**Carlquist S. 1976.** Tribal interrelationships and phylogeny of the Asteraceae. *Aliso* **8:** 465–492.

**CBOL Plant Working Group**, **Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, van der Bank M, Chase MW, Cowan RS, Erickson DL, Fazekas AJ, Graham SW, James KE, Kim KJ, Kress WJ, Schneider H, van AlphenStahl J, Barrett SC, van den Berg C, Bogarin D, Burgess KS, Cameron KM, Carine M, Chacón J, Clark A, Clarkson JJ, Conrad F, Devey DS, Ford CS, Hedderson TA, Hollingsworth ML, Husband BC, Kelly LJ, Kesanakurti PR, Kim JS, Kim YD, Lahaye R, Lee HL, Long DG, Madriñán S, Maurin O, Meusnier I, Newmaster SG, Park CW, Percy DM, Petersen G, Richardson JE, Salazar GA, Savolainen V, Seberg O, Wilkinson MJ, Yi DK, Litt DP. 2009.** A DNA barcode for land plants. *Proceedings of the National Academy of Sciences, U.S.A.* **106:** 12794–12797.

**Chen S**, **Yao H**, **Han J**, **Liu C**, **Song J**, **Shi L**, **Zhu Y**, **Ma X**, **Gao T**, **Pang X**, **Luo K**, **Li Y**, **Li X**, **Jia X**, **Lin Y**, **Leon C. 2010.** Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS One* **5:** e8613.

**China Plant BOL Group**, **Li DZ**, **Gao LM, Li HT, Wang H, Ge XJ, Liu JQ, Chen ZD, Zhou SL, Chen SL, Yang JB, Fu CX, Zeng CX, Yan HF, Zhu YJ, Sun YS, Chen SY, Zhao L, Wang K, Yang T, Duan GW. 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, U.S.A.* **108:** 19641–19646.

**Clague DA. 1996.** Growth and subsidence of the Hawaiian-emperor volcanic chain. In: Keast A, Miller SE, eds. *The origin and evolution of Pacific Island biotas, New Guinea to eastern Polynesia: patterns and processes*. Amsterdam: SPB Academic Press, 35–50.

**Clark K**, **Karsch-Mizrachi I**, **Lipman DJ**, **Ostell J**, **Sayers EW. 2015.** GenBank. *Nucleic Acids Research* **44:** D67–D72.

**Collins RA**, **Cruickshank RH. 2012.** The seven deadly sins of DNA barcoding. *Molecular Ecology Resources* **13:** 969–975.

**Cowan RS**, **Fay MF. 2012.** Challenges in the DNA barcoding of plant material. In: Sucher NJ, Hennell JR, Carles MC, eds. *Plant DNA fingerprinting and barcoding: methods and protocols*. Totowa: Humana Press, 23–33.

**DeSalle R**, **Birstein VJ. 1996.** PCR identification of black caviar. *Nature* **381:** 197.

**Dong W**, **Xu C**, **Li C**, **Sun J**, **Zuo Y**, **Shi S**, **Cheng T**, **Guo J**, **Zhou S. 2015.** *ycf1*, the most promising plastid DNA barcode of land plants. *Scientific Reports* **5:** 8348.

**Douglas PP**, **Shaw RB. 1989.** Rediscovery of *Tetramolopium arenarium* subsp. *arenarium* var. *arenarium* (Asteraceae: Astereae) on the Pohakuloa Training Area, Hawaii. *Annals of the Missouri Botanical Garden* **76:** 1182–1185.

**Dunbar-Co S**, **Wieczorek AM**, **Morden CW. 2008.** Molecular phylogeny and adaptive radiation of the endemic Hawaiian *Plantago* species (Plantaginaceae). *American Journal of Botany* **95:** 1177–1188.

**Eggens F**, **Popp M**, **Nepokroeff M**, **Wagner WL**, **Oxelman B. 2007.** The origin and number of introductions of the Hawaiian endemic *Silene* species (Caryophyllaceae). *American Journal of Botany* **94:** 210–218.

**ESRI**. **2012**. *ArcGIS Release 10.1*. Environmental Systems Research Institute, Redlands, CA.

**Funk VA**, **Gostel M**, **Devine A**, **Kelloff CL**, **Wurdack K**, **Tuccinardi C**, **Radosavljevic A**, **Peters M**, **Coddington J. 2017.** Guidelines for collecting vouchers and tissues intended for genomic work (Smithsonian Institution): botany best practices. *Biodiversity Data Journal* **5:** e11625.

**Givnish TJ**, **Millam KC**, **Mast AR**, **Paterson TB**, **Theim TJ**, **Hipp AL**, **Henss JM**, **Smith JF**, **Wood KR**, **Sytsma KJ. 2009.** Origin, adaptive radiation and diversification of the Hawaiian lobeliads (Asterales: Campanulaceae). *Proceedings of the Royal Society of London. Series B: Biological Sciences* **276:** 407–416.

**Gong L**, **Qiu XH**, **Huang J**, **Xu W**, **Bai JQ**, **Zhang J**, **Su H**, **Xu CM**, **Huang ZH. 2018.** Constructing a DNA barcode reference library for southern herbs in China: a resource for authentication of southern Chinese medicine. *PLoS One* **13:** e0201240.

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

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

**Hebert PDN**, **Stoeckle MY**, **Zemlak TS**, **Francis CM. 2004.** Identification of birds through DNA barcodes. *PLoS Biology* **2:** e312.

**Hillis DM**, **Bull JJ. 1993.** An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* **42:** 182–192.

**Hollingsworth PM**, **Li DZ**, **van der Bank, M**, **Twyford AD. 2016.** Telling plant species apart with DNA: from barcodes to genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371:** 20150338.

**Horn JW**, **van Ee BW**, **Morawetz JJ**, **Riina R**, **Steinmann VW**, **Berry PE**, **Wurdack KJ. 2012.** Phylogenetics and the evolution of major structural characters in the giant genus *Euphorbia* L. (Euphorbiaceae). *Molecular Phylogenetics and Evolution* **63:** 305–326.

**Howarth DG**, **Gustafsson MHG**, **Baum DA**, **Motley TJ. 2003.** Phylogenetics of the genus *Scaevola* (Goodeniaceae): implication for dispersal patterns across the Pacific Basin and colonization of the Hawaiian Islands. *American Journal of Botany* **90:** 915–923.

**Jaén-Molina R**, **Marrero-Rodríguez Á**, **Reyes-Betancort J**, **Santos-Guerra A**, **Naranjo-Suárez J**, **Caujapé-Castells J. 2015.** Molecular taxonomic identification in the absence of a 'barcoding gap': a test with the endemic flora of the Canarian oceanic hotspot. *Molecular Ecology Resources* **15:** 42–56.

**Johnson MA**, **Clark JR**, **Wagner WL**, **McDade LA. 2017.** A molecular phylogeny of the Pacific clade of *Cyrtandra* (Gesneriaceae) reveals a Fijian origin, recent diversification, and the importance of founder events. *Molecular Phylogenetics and Evolution* **116**:30–48.

**Jukes TH**, **Cantor CR. 1969.** Chapter 24 – evolution of protein molecules. In: Munro HN, ed. *Mammalian protein metabolism*. New York: Academic Press, 21–132.

**Katoh K**, **Standley DM. 2013.** MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* **30:** 772–780.

**Kearse M**, **Moir R**, **Wilson A**, **Stones-Havas S**, **Cheung M**, **Sturrock S**, **Buxton S**, **Cooper A**, **Markowitz S**, **Duran C**, **Thierer T**, **Ashton B**, **Meintjes P**, **Drummond A. 2012.** Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28:** 1647–1649.

**Keeley SC**, **Funk VA. 2011.** Origin and evolution of Hawaiian endemics: new patterns revealed by molecular phylogenetic studies. In: Bramwell D, Caujapé-Castells J, eds. *The biology of island floras*. Cambridge: Cambridge University Press, 57–87.

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

**Knope ML**, **Morden CW**, **Funk VA**, **Fukami T. 2012.** Area and the rapid radiation of Hawaiian *Bidens* (Asteraceae). *Journal of Biogeography* **39:** 1206–1216.

**Krehenwinkel H**, **Pomerantz A**, **Henderson JB**, **Kennedy SR**, **Lim JY**, **Swamy V**, **Shoobridge JD**,

**Graham N**, **Patel NH**, **Gillespie RG**, **Prost S. 2019.** Nanopore sequencing of long ribosomal DNA amplicons enables portable and simple biodiversity assessments with high phylogenetic resolution across broad taxonomic scale. *GigaScience* **8:** 1–16.

**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**, **Weigt LA**, **Janzen DH. 2005.** Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences, U.S.A.* **102:** 8369–8374.

**Li Y**, **Tong Y**, **Xing F. 2016.** DNA barcoding evaluation and its taxonomic implications in the recently evolved genus *Oberonia* Lindl. (Orchidaceae) in China. *Frontiers in Plant Science* **7:** 1791.

**Liu J**, **Yan H**, **Ge X. 2016.** The use of DNA barcoding on recently diverged species in the genus *Gentiana* (Gentianaceae) in China. *PLoS One* **11:** e0153008.

**Liu J**, **Yan H**, **Newmaster SG**, **Pei N**, **Ragupathy S**, **Ge X. 2015.** The use of DNA barcoding as a tool for the conservation biogeography of subtropical forests in China. *Diversity and Distributions* **21:** 188–199.

**Lowrey TK**, **Whitkus R**, **Sykes WR. 2005.** A new species of *Tetramolopium* (Asteraceae) from Mitiaro, Cook Islands: biogeography, phylogenetic relationships, and dispersal. *Systematic Botany* **30:** 448–456.

**Magnacca KN**, **Brown MJF. 2010.** Mitochondrial heteroplasmy and DNA barcoding in Hawaiian *Hylaeus* (*Nesoprosopis*) bees (Hymenoptera: Colletidae). *BMC Evolutionary Biology* **10:** 174.

**Magnacca KN**, **Danforth BN. 2006.** Evolution and biogeography of native Hawaiian *Hylaeus* bees (Hymenoptera: Colletidae). *Cladistics* **22:** 393–411.

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

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

**Morden CW**, **Harbin SC. 2013.** Evolution and biogeographic origins of the endemic Hawaiian genus *Hesperomannia* (Asteraceae). *Pacific Science* **67:** 219–235.

**Parmentier I**, **Duminil J**, **Kuzmina M**, **Philippe M**, **Thomas DW**, **Kenfack D**, **Chuyong GB**, **Cruaud C**, **Hardy OJ. 2013.** How effective are DNA barcodes in the identification of African rainforest trees? *PLoS One* **8:** e54921.

**Percy DM**, **Argus GW**, **Cronk QC**, **Fazekas AJ**, **Kesanakurti PR**, **Burgess KS**, **Husband BC**, **Newmaster SG**, **Barrett SCH**, **Graham SW. 2014.** Understanding the spectacular failure of DNA barcoding in willows (*Salix*): does this result from a trans-specific selective sweep? *Molecular Ecology* **23:** 4737–4756.

**Pillon Y**, **Johansen J**, **Sakishima T**, **Chamala S**, **Barbazuk WB**, **Roalson EH**, **Price DK**, **Stacy EA. 2013.** Potential use of low-copy nuclear genes in DNA barcoding: a comparison with plastid genes in two Hawaiian plant radiations. *BMC Evolutionary Biology* **13:** 35.

**Price JP**, **Clague DA. 2002.** How old is the Hawaiian biota? Geology and phylogeny suggest recent divergence. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **269:** 2429–2435.

**Price JP**, **Wagner WL. 2004.** Speciation in Hawaiian angiosperm lineages: cause, consequence, and mode. *Evolution* **58:** 2185–2200.

**Price JP**, **Wagner WL. 2018.** Origins of the Hawaiian flora: phylogenies and biogeography reveal patterns of long-distance dispersal. *Journal of Systematics and Evolution* **56:** 600–620.

**R Core Team**. **2017.** *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

**Rach J**, **DeSalle R**, **Sarkar IN**, **Schierwater B**, **Hadrys H. 2008.** Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. *Proceedings of the Royal Society of London, Series B: Biological Sciences* **275:** 237.

**Roy S**, **Tyagi A**, **Shukla V**, **Kumar A**, **Singh UM**, **Chaudhary LB**, **Datt B**, **Bag SK**, **Singh PK**, **Nair NK**, **Husain T**, **Tuli R. 2010.** Universal plant DNA barcode loci may not work in complex groups: a case study with Indian *Berberis* species. *PLoS One* **5:** e13674.

**Saarela JM**, **Sokoloff PC**, **Gillespie LJ**, **Consaul LL**, **Bull RD. 2013.** DNA barcoding the Canadian arctic flora: core plastid barcodes (*rbcL* + *matK*) for 490 vascular plant species. *PLoS One* **8:** e77982.

**Saitou N**, **Nei M. 1987.** The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4:** 406–425.

**Sakai AK**, **Wagner WL**, **Mehrhoff LA. 2002.** Patterns of endangerment in the Hawaiian flora. *Systematic Biology* **51:** 276–302.

**Sarkar IN**, **Planet PJ**, **Desalle R. 2008.** CAOS software for use in character-based DNA barcoding. *Molecular Ecology Resources* **8:** 1256–1259.

**Schnell L**, **Evans S**, **Kawakami K**, **Aubrey K**, **Trainer J. 2003.** *Annual Report for the Ecosystem Management Program Pohakuloa Training Area, Island of Hawaii*. Honolulu: Pacific Cooperative Studies Unit.

**Schoch CL**, **Seifert KA**, **Huhndorf S**, **Robert V**, **Spouge JL**, **Levesque CA**, **Chen W**, **Fungal Barcoding Consortium**. **2012.** Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. *Proceedings of the National Academy of Sciences, U.S.A.* **109:** 6241–6246.

**Spooner DM. 2009.** DNA barcoding will frequently fail in complicated groups: an example in wild potatoes. *American Journal of Botany* **96:** 1177–1189.

**Taylor HR**, **Harris WE. 2012.** An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. *Molecular Ecology Resources* **12:** 377–388.

**Wagner WL**, **Herbst DR**, **Lorence DH. 2005. continuously updated**. *Flora of the Hawaiian Islands*. [Online] Available at: http://botany.si.edu/pacificislandbiodiversity/hawaiianflora/index.htm. accessed January 2018.

**Wagner WL**, **Herbst DR**, **Sohmer SH. 1999.** *Manual of the flowering plants of Hawai'i, revised edition with supplement by W. L. Wagner and D. R. Herbst*. Honolulu: Bishop Museum Press.

**Weitschek E**, **Velzen R**, **Felici G**, **Bertolazzi P. 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.

**Welch AJ**, **Collins K**, **Ratan A**, **Drautz-Moses DI**, **Schuster SC**, **Lindqvist C. 2016.** The quest to resolve recent radiations: plastid phylogenomics of extinct and endangered Hawaiian endemic mints (Lamiaceae). *Molecular Phylogenetics and Evolution* **99**:16–33.

**Witter MS**, **Carr GD. 1988.** Adaptive radiation and genetic differentiation in the Hawaiian silversword alliance (Compositae: Madiinae). *Evolution* **42:** 1278–1287.

**Wood K. 2007.** New plant records, rediscoveries, range extensions, and possible extinctions within the Hawaiian Islands. *Bishop Museum Occasional Papers* **96:** 18–22.

**Yassin A**, **Markow TA**, **Narechania A**, **O'Grady PM**, **DeSalle R. 2010.** The genus *Drosophila* as a model for testing tree- and character-based methods of species identification using DNA barcoding. *Molecular Phylogenetics and Evolution* **57:** 509–517.

**Ziegler A. 2002.** *Hawaiian Natural History, Ecology, and Evolution*. Honolulu: University of Hawai'i Press.

**Zou S**, **Li Q**, **Kong L**, **Yu H**, **Zheng X. 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.

## SUPPORTING INFORMATION

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

**Figure S1.** Overall identification success by locus and barcoding method used for better-sampled (at least five species with two accessions, or > 50% of species with two accessions) lineages. Sample sizes are the total number of species with at least two accessions at that locus.

**Table S1.** Asteraceae taxa collected for this project with GenBank accession numbers for each locus, National Tropical Botanical Garden Herbarium (PTBG) accession numbers for taxa with physical specimens, and Global Genome Biodiversity Network (GGBN) biorepository numbers for taxa with photographic vouchers.

**Table S2.** Primers used for PCR and sequence generation.

**Table S3.** GenBank taxa used in study by locus, then alphabetically by family, lineage, genus and species with accession numbers.

**Table S4.** Outgroup taxa used for tree-based method and GenBank accession numbers.

**Table S5.** *Scaevola*, *Melicope*, *Pritchardia* and *Geranium* taxa used for geographic analysis, number of accessions and range (K for Kaua'i, O for O'ahu, Mo for Molokai, L for Lana'i, M for Maui and H for Hawai'i Island). See Table S3 for accession numbers.