METHODS AND TECHNIQUES

A large scale plant survey: efficient vouchering with identification through morphology and DNA analysis

Laura E. Dugan[1,3], Martin F. Wojciechowski[2] & Leslie R. Landrum[2]

[1] *Global Institute of Sustainability, Arizona State University, Tempe, Arizona 85287, U.S.A.*
[2] *School of Life Sciences, Arizona State University, Tempe, Arizona 85287, U.S.A. les.landrum@asu.edu (author for correspondence)*
[3] *Current address: Integrative Biology, University of Texas at Austin, 1 University Station C0930, Austin, Texas 78712, U.S.A.*

A large-scale survey of the Central Arizona-Phoenix Long-Term Ecological Research study area was conducted in which many kinds of samples were collected (e.g., plant, insect, soil). The plant samples were of a smaller size than typical herbarium accessions due to limitations in personnel, storage space, available plant materials, and time. The vouchers were identified in the field if possible. The vast majority of the rest were identified in the herbarium morphologically after the survey; less than 5% of the vouchers remained unidentified. Of these, a subset were subjected to molecular analysis using sequences of the nuclear ribosomal ITS region and to a BLAST search of sequences in GenBank to find closely related taxa. The close relatives along with the unknown specimens were subjected to maximum parsimony analysis to identify the potential phylogenetic relationships of the unknown specimens. Through this analysis nine unknown plant accessions were identified to family, tribe, genus, and species. Once this level of identification was established, some specimens were re-examined morphologically and compared with potentially related taxa in the herbarium to confirm or improve identifications. We outline an efficient method for surveying and collecting plant vouchers for a large sampling area. We also demonstrate that one can combine morphological and molecular data in the identification process to produce more complete datasets.

KEYWORDS: Arizona, nrDNA ITS sequences, plant identification, survey, voucher specimens

INTRODUCTION

Voucher specimens are fundamental in the biological sciences (Boom, 1996). Documenting what organisms are being studied, whether it is for molecular systematics, taxonomic monographs, ethnobotanical studies or ecological studies, links names of organisms to specific specimens that can be re-examined by future workers. The identifications can be verified or corrected, greatly enhancing the value of the study when voucher specimens are obtained. Studies without voucher specimens are sometimes without value, while incorrect determinations have a negative effect by providing false information that may be perpetuated in subsequent publications.

Collecting voucher specimens is crucial to obtaining accurate identifications of plants present in a study area. Morgan & Overholt (2005) document a situation in St. Lucie County, Florida in which very poor records had been maintained of the vegetation in the county, and the result is a gap in the knowledge of the community structure of plants in that area. In one survey they found an additional 18 category one invasive plant species (community-altering species) that had not been known to exist in the county before. It is highly likely that many had been present in the county for years, but had not been documented. Voucher specimens help to create a measure of a particular area's biodiversity and vegetation structure at a given point in time and provide documented proof that a plant did, in fact, exist in a given place at a particular time. In the Sonoran Desert about half of the species of plants are ephemerals that respond to seasonal rainfall and do not appear every year. Based on records provided by vouchers, we know that some plants have disappeared from the Phoenix area as wild plants (e.g., *Cephalanthus occidentalis* L., *Prosopis pubescens* Benth.) and that others have only appeared in the last few years (e.g., *Oncosiphon piluliferum* (L. f.) Källersjö). Thus, the visible flora changes according to rainfall patterns, and the total flora changes because of local extinctions and introductions.

The use of DNA sequence analysis using molecular phylogenetic methods can further enhance our knowledge of the flora of an area by aiding in the identification of unknown specimens and biodiversity (e.g., Kress & al., 2005; Cowan & al., 2006). Specimens may be unidentifiable morphologically because they do not have mature

vegetative or reproductive structures, or do not have an essential identifying structure present (e.g., leaf, flower, fruit). A specimen may remain unknown because the species has not been known to exist in an area previously, or is extremely rare in an area, and thus there is little or no previous knowledge of it existing in a particular locality. Also, as in the survey described here, many plants were left unidentified ("unknown") because they were rarely used or are non-native horticultural plants (found at urban sites). By analyzing the sequence of a variable, nuclear spacer region that is commonly used for phylogenetic studies in plants, we were able to determine the identity of some otherwise unidentifiable plants collected during the course of this biodiversity survey.

METHODS

The CAP LTER 200 Point Survey. — From mid-February to early June of 2005 a large-scale survey of the Phoenix, Arizona, metropolitan area and outlying agricultural and desert areas was conducted as part of the interdisciplinary Central Arizona-Phoenix Long-Term Ecological Research (CAP LTER) study at Arizona State University (ASU) funded by the National Science Foundation. This survey is performed once every five years and the 2005 survey was the second such survey conducted. One of the purposes of this survey is to create a "snapshot" of the biodiversity present in the study region at that point in time, which can then be analyzed alone and can also be compared to the "snapshots" of the area in other years (e.g., the 2000 survey as well as those surveys yet to be conducted).

An area of 6,387 km^2 was surveyed by collecting data from 204 randomly selected sites, each measuring 30 × 30 m, over a range of land-use types from heavily urbanized commercial and residential sites within the city, to agricultural land on the outskirts of the city, to undeveloped desert surrounding the city. Several values were measured and many kinds of specimens collected (e.g., soil, plant, insect). We anticipated that a few thousand plants would be encountered in this survey. Vouchering each plant on a standard herbarium sheet at each site seemed impractical for various reasons: (1) adequate material for a full-sized sheet was often not available; (2) sufficient personnel were not available to collect complete, full-sized specimens; (3) storage of the anticipated number of specimens would have required about five full-sized herbarium cabinets. An alternative method that produced vouchers but that required less time and expense was devised. Because the survey lasted from early spring to early summer, specimens were collected in whatever stage they happened to be at when the survey point was visited, and therefore ranged from emerging seedlings to desiccated mature plants.

Method of collection. — Instead of standard-sized herbarium sheets (30.5 × 43.0 cm) specimens were collected in the field into herbarium packets 10 cm by 10.5 cm in size. The design of the ASU packet is such that they do not easily open (see ASU website for instructions at http://lifesciences.asu.edu/herbarium/packet.html). The packets each had a pre-printed label that could easily be filled-in in the field, with information about the particular survey point locality, field identification if any, and a collection number for that point. Each collection was given a unique identification code that is a combination of the collection point and a separate specimen number. The "collector" is CAP LTER survey crew for 2005. Succulent plants that could not be easily collected or plants in urban settings where sampling would have disrupted a landscape design were photographed only. The packets were placed into a standard plant press and placed in a drier within two days. When the survey was completed the identified specimens were organized alphabetically by family, genus, and species. Five thousand two hundred and twenty-five voucher specimens collected were stored in 28 CD storage boxes. The boxes required ca. one-half of one herbarium cabinet for storage. Their label data are available at http://seinet.asu.edu/collections/selection2.jsp, and the collection is stored in the ASU herbarium.

Taking photographs of every plant collected was part of the plan at the beginning of the survey, but after one week photography was discarded in most cases because of the amount of time it required. Field crews usually consisted of three or four members. Sites contained up to 60 plant species, and it was quickly realized that photographing every individual species was not practical for the field crew available. An additional member whose job it was to photograph specimens would have been desirable and a more complete record of each site could have been made.

Plants were collected, when possible, that were either in fruit or flower, and as much of the plant was collected as (1) could easily be removed from the ground and (2) fit into a collection packet. When the plants were neither flowering nor fruiting, as much of the plant (e.g., basal and stem leaves, stems showing branching, roots if possible) was collected as feasible. For every plant that was not identifiable in the field, notes were generally taken on the collection packet regarding architecture of the plant, microhabitat, and other pertinent characteristics to aid in later identification.

The large majority of the specimens collected in the 2005 survey were identified in the field and only required verification by comparison to existing specimens in the herbarium. The bulk of those remaining were subsequently identified in the herbarium using the small vouchers available (Fig. 1). This high level of identification was possible because the flora of the Phoenix area has been well studied and is well documented with specimens in

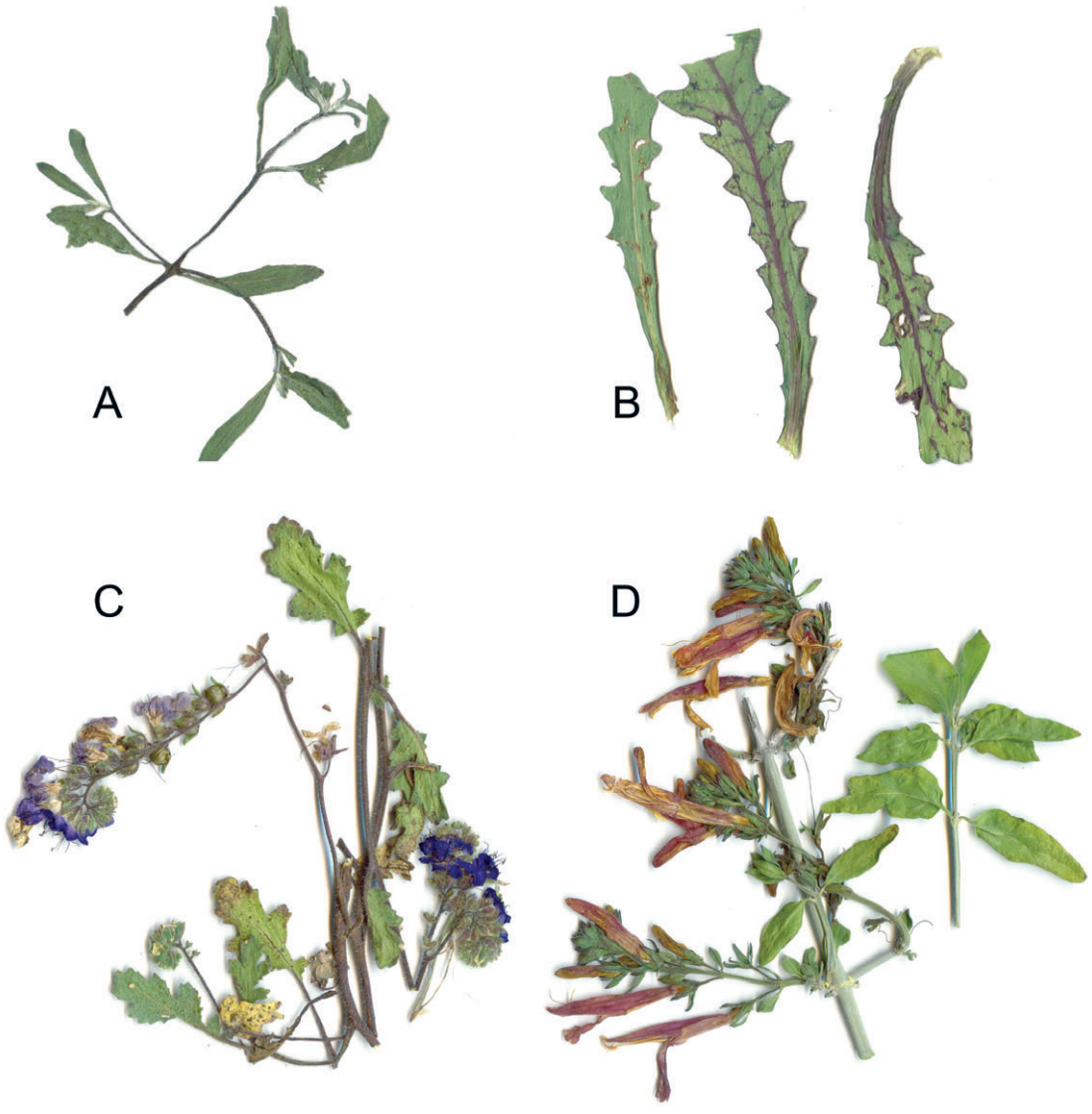

Fig. 1. Four fragment vouchers of the CAP LTER 200 point survey for 2005. A, AA191-31, *Eclipta prostrata*; B, AA181-5, *Hedypnois cretica*; C, V201-12, *Phacelia ambigua*; D, Z101-27, *Justicia californica*. These samples are typical of the specimen voucher size. Over 95% of specimens could be identified to species by morphological examination in the field and/or herbarium; A and B in this figure were among the few that could not be. They were identified first by DNA analysis to a likely group of taxa. Then they were compared morphologically with the same or related species known to grow in the Phoenix area and specific identifications could be made.

the ASU herbarium. Our method would not have worked well for a poorly known flora.

Identification by DNA sequence analysis. — The ITS region of nuclear ribosomal DNA is the most commonly sequenced genetic locus used in plant molecular systematic investigations at the generic and infrageneric levels and shows high levels of interspecific variation (Baldwin & al., 1995; Alvarez & Wendel, 2003). As of November 2006, nrDNA ITS is the most widely used locus in angiosperms with more than 60,000 sequences published in sequence databases such as GenBank in comparison to the *trnL* intron (15,900), *rbcL* gene (18,750), *matK* gene (17,550), *ndhF* gene (7,740), and the *trnH-psbA* spacer (3,830). Recent studies have demonstrated nrDNA ITS' potential for use in plant identifications, either alone or in combination with a plastid DNA region such as the *trnH-psbA* spacer (Kress & al., 2005; Gemeinholzer & al., 2006). Although problems with intraspecific sequence variation and paralogues in nrDNA ITS do occur (e.g., Alvarez & Wendel, 2003), such problems are often well documented and appear limited to certain taxa, and can be resolved by additional analyses.

DNA preparation. — Of the total specimens collected in the 2005 CAP LTER survey, there were 242

remaining unidentified plant specimens. As a test of our ability to use a standardized molecular identification (DNA sequencing and phylogenetic analyses) as a tool to aid in their identification, genomic DNAs were extracted from leaf material from 25 of these unknowns using standard protocols (Qiagen DNeasy Plant Minikits, Qiagen, Valencia, California). Nine of these produced a sufficiently clean sequence (with 100% sequence overlap, both strands) and were chosen for further examination by standard parsimony analysis. Polymerase chain reaction (PCR) amplifications of the nuclear ribosomal internal transcribed spacer region (nrDNA ITS region, including ITS1, 5.8S rRNA gene, and ITS2) were performed on diluted DNA samples using the primers ITS 18ML and ITS 26ML (Beyra-Matos & Lavin, 1999) as described by Wojciechowski & al. (1999). Amplified products were purified by ultrafiltration and then sequenced using these same primers. DNA sequencing was performed on an Applied Biosystems 3100 Sequencer at the Arizona State University DNA Laboratory. Sequencer output files were assembled into contigs and edited using Sequencher 4.1 (GeneCodes, Ann Arbor, Michigan). Final sequences from the nine unidentified plant specimens with good sequence data were subjected to BLAST analysis ("blastn" at http://www.ncbi.nlm.nih.gov/BLAST/) to determine potentially homologous sequences in GenBank to aid in identification.

Phylogenetic analysis. — The nrDNA ITS sequences from nine unknowns (AA181-5, AA191-31, AE91-27, L141-12, S131-13, U151-17, W181-23, X171-17, Z111-39) were subjected to further phylogenetic analysis. Selected sequences in GenBank with the highest similarity (i.e., lowest e-scores) to the unknown query sequences based on the BLASTn results were downloaded and aligned using ClustalX (Thompson & al., 1997) by using standard pairwise and multiple alignment parameter settings (default gap opening and penalty parameters). The resulting datasets were manually readjusted as necessary to maximize consistency in the placement of gaps in the sequences (due to insertions or deletions, "indels").

Phylogenetic analyses were performed using maximum parsimony using PAUP* (version 4.0b10; Swofford, 2002). Multiple tree searches were conducted using heuristic search options that included SIMPLE, CLOSEST, and RANDOM (1,000 replicates) addition sequences and tree bisection-reconnection branch swapping, with retention of multiple parsimonious trees (MAXTREES set to 10,000). Clade support was determined using nonparametric bootstrap resampling (Felsenstein, 1985), estimated from 500 bootstrap replicates that incorporated heuristic searches using addition sequences and branch swapping options as in our standard parsimony analyses. In all analyses, ambiguously aligned and gapped positions (indels) were excluded.

RESULTS

Identification.— Based on phylogenetic analysis of the nine unknowns for which good nrDNA ITS sequences could be obtained, three unknowns were identified to tribe, four were identified to genus, and two were identified to species (Table 1).

The phylogenetic tree produced for unknown AA181-5 shows it to be nested within a *Leontodon* and *Hedypnois* clade, both of the tribe Lactuceae; hence we can determine that this unknown is a member of the Lactuceae tribe (Fig. 2). However, it is not clear based on molecular evidence alone what the genus of the unknown is. The same situation occurs with unknown AA191-31, which rests within a clade containing *Kingianthus*, *Eclipta*, and *Helianthus*, all of which are in the tribe Heliantheae (Fig. 3). It is reasonable to assume that AA191-31 is a member of the Heliantheae, but it is not clear to what genus it belongs. The root of this tree was determined using information about the subtribe Ecliptinae from Panero & al. (1999).

Therefore, using the molecular data and knowledge of the local flora, the specimens were re-examined morphologically and compared to identified accessions in the herbarium for verification. More specific identities were

Table 1. Identification of some unknowns based on nrDNA ITS sequences alone and identification of the same specimens upon morphological re-examination.

Accession #	Taxonomic identity as determined by phylogenetic analysis	Taxonomic identity after re-examination using molecular and morphological data
AA181-5	Tribe Lactuceae (Asteraceae)	*Hedypnois cretica*
AA191-31	Tribe Heliantheae (Asteraceae)	*Eclipta prostrata*
AE91-27	*Baccharis* sp. (Asteraceae)	*Baccharis salicifolia*
L141-12	Tribe Astereae (Asteraceae)	*Heterotheca* cf. *subaxillaris*
S131-13	*Ligustrum* spp. (Oleaceae)	*Ligustrum* cf. *japonicum*
U151-17	*Fraxinus* spp. (Oleaceae)	*Fraxinus* spp.
W181-23	*Brachychiton populneus* (Sterculiaceae)	*Brachychiton populneus*
X171-17	*Lantana* spp. (Verbenaceae)	*Lantana* spp.
Z111-39	*Nicotiana obtusifolia* (Solanaceae)	*Nicotiana obtusifolia*

then given to five of the sequenced specimens (Table 1). AA181-5 (tribe Lactuceae from the molecular data) was determined to be *Hedypnois cretica*, an uncommon exotic weed. AA191-31 (tribe Heliantheae from the molecular data) was identified as *Eclipta prostrata*. AE91-27 (*Baccharis* sp. from the molecular data) was determined to be *B. salicifolia*. L141-12 (tribe Astereae from the molecular data) was determined to be *Heterotheca* cf. *subaxillaris*. S131-13 (*Ligustrum* spp. from the molecular data) bears closest resemblance to *L. japonicum* and was given the identity *L.* cf. *japonicum*.

DISCUSSION

There are advantages and disadvantages to this method of sampling. The advantages are: (1) Time and materials are saved in making collections directly into packets in the field (no mounting time or additional materials are needed). In many cases the amount of plant material available on the day of the survey is unavoidably small and using a full sheet would not be sensible. (2) Space is saved in the storage of 5,000 specimens (ca. 1/2 herbarium cabinet is used instead of ca. 5 full cabinets every time the survey is repeated). (3) Field biologists who often make no

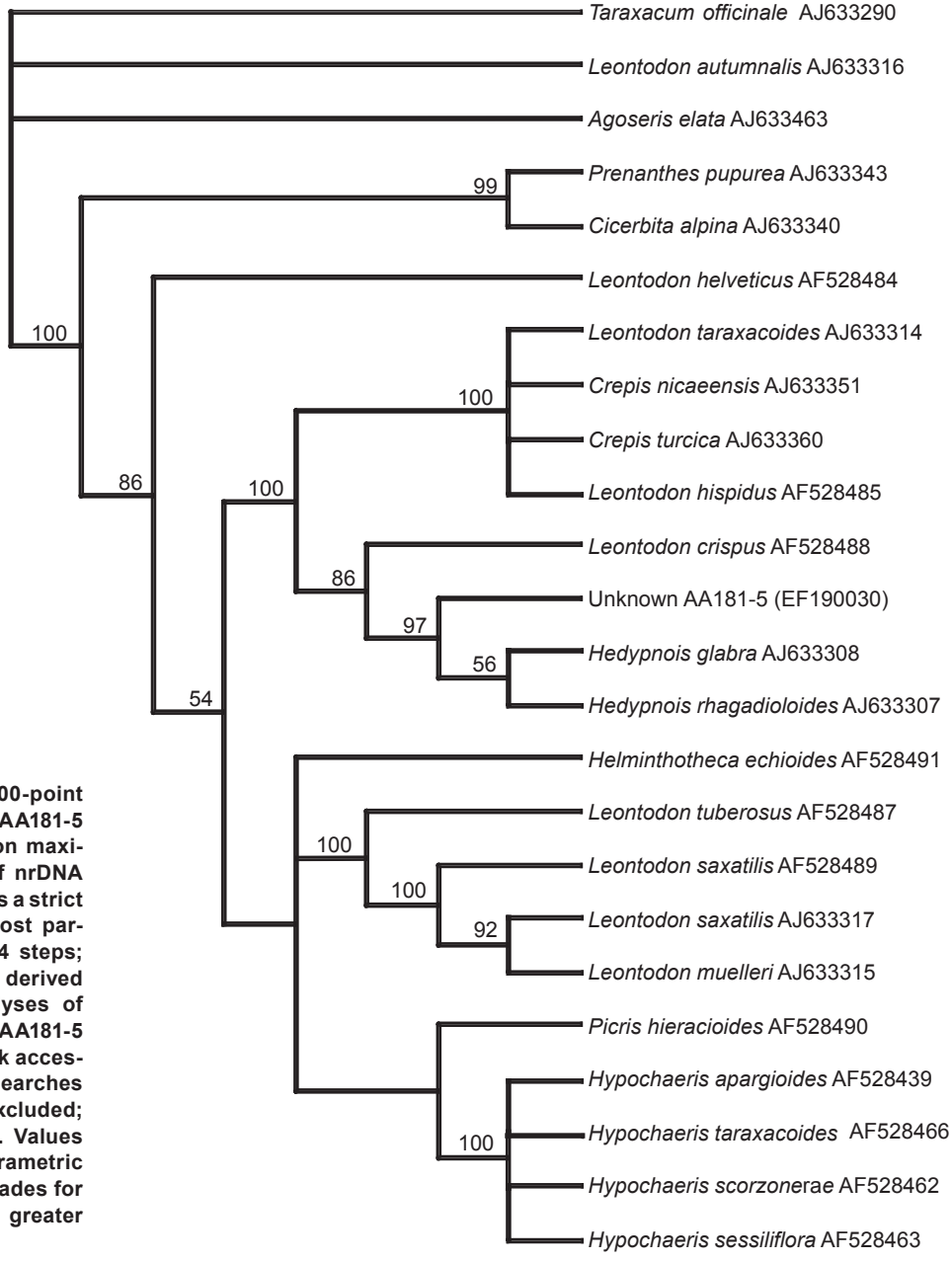

Fig. 2. Identification of 200-point survey unknown accession AA181-5 (*Hedypnois cretica*) based on maximum parsimony analysis of nrDNA ITS sequences. Tree shown is a strict consensus of 90 equally most parsimonious trees (length 404 steps; CI = 0.7228; RI = 0.8066) derived from heuristic search analyses of sequences from accession AA181-5 and 23 most similar GenBank accessions, based on BLASTn searches (863 total characters, 237 excluded; 139 parsimony informative). Values above the nodes are non-parametric bootstrap proportions for clades for which support values were greater than 50%.

voucher specimens at all are willing to make these small vouchers. (4) Packet samples can sometimes be used in molecular analysis in spite of their small size.

The disadvantages are: (1) Success with this type of survey requires that the local flora is well known. It may work especially well in desert environment where many of the plants are small anyway. In an area with a poorly known flora the method would not be advisable. In a tropical area where most plants would require large samples to be sufficient for identification, the method would not be advisable. (2) Small samples are sometimes identifiable but do not offer complete morphological information. They are not replacements for full sized specimens used in systematic morphological studies. (3) It is possible that non-taxonomists will be discouraged from making full sized herbarium specimens and choose to make packet collections instead, knowing that the latter method is available. (4) Packet specimens are stored separately from standard sheet specimens. Since the packet specimens are principally geographic or ecological vouchers (not especially useful for morphology) this is not a great problem.

In dealing with ecologists and other colleagues for years we have found that they often neglect to make adequate voucher specimens, so that many otherwise valuable

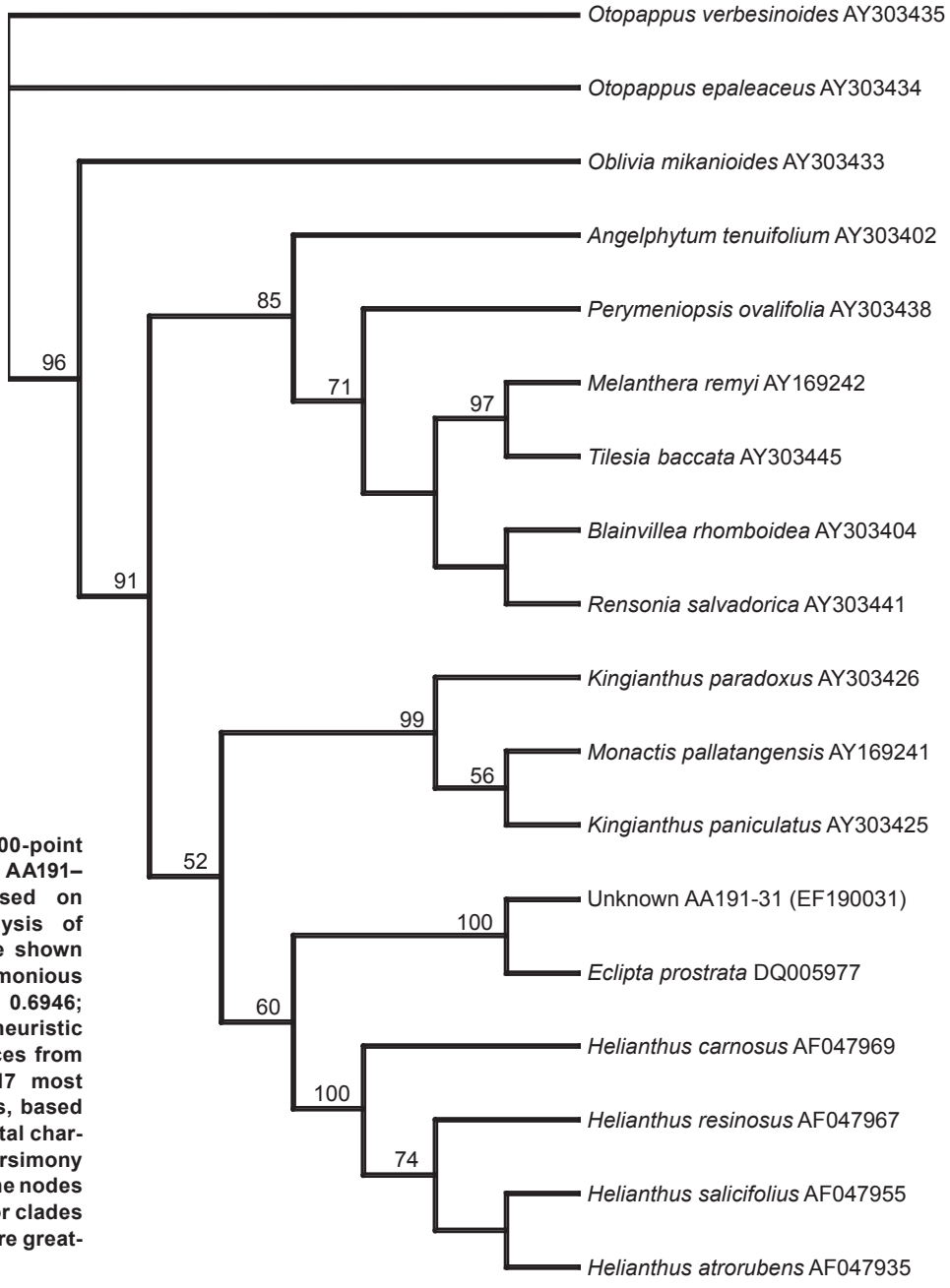

Fig. 3. Identification of 200-point survey unknown accession AA191–31 (*Eclipta prostrata*) based on maximum parsimony analysis of nrDNA ITS sequences. Tree shown is the single most parsimonious tree (length 429 steps; CI = 0.6946; RI = 0.7519) derived from heuristic search analyses of sequences from accession AA191-31 and 17 most similar GenBank accessions, based on BLASTn searches (750 total characters, 112 excluded; 152 parsimony informative). Values above the nodes are bootstrap proportions for clades for which support values were greater than 50%.

studies go poorly documented, or are entirely without voucher specimens. The method proposed here provides a solution for such studies but in no case encourages the abandonment of traditional full-sized specimens for systematic studies. Time was a limiting constraint in this study as a large amount of survey work had to be completed in a short period, about 3.5 months. Thus, carefully searching each site for the best specimens was not always possible. Also, because some of the sites were visited well after many of the ephemeral plants had flowered, fruited, and set seed, many plants were dead or dying at the time of collection allowing only for less than perfect specimens.

The use of molecular identification tools to further enhance the success of this type of floristic study is still in the early stages. While currently (November 2006) there are more than 17,568,900 nucleotide sequences for some 63,597 angiosperm taxa in GenBank, many plant genera and species have yet to be sequenced for one or another locus and/or have their sequences deposited in GenBank. This creates "gaps" in the set of sequences available for comparison. However, this will certainly be less of a limitation to studies such as ours as more and more species are sequenced for informative loci and deposited in sequence databases. Moreover, although a sequence from a particular species may not be found in the database, sequences from another member of the unknown's genus, tribe, or family, are very likely to be present in GenBank, thereby allowing one to identify potentially related taxa (based on BLASTn searches) and build a phylogeny of the unknown and its close relatives. Knowing the probable close relatives of an unknown, one may then be able to determine its identity through more detailed morphological comparison and knowledge of the local flora. Molecular data can be particularly beneficial as an identification tool in cases in which a species is new to a specific region or state (potentially invasive) and that may have otherwise remained unknown, because the new plant does not appear in local flora lists or keys (M.F. Wojciechowski, unpub. data). The advantage of molecular based identification is obvious when a particular collection is immature or propagates primarily by asexual means and thus does not bear the structures that are necessary to identify it with certainty according to morphological criteria.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. DEB-0423704, Central Arizona-Phoenix Long-Term Ecological Research (CAP LTER) to Nancy Grimm and Charles Redman. Any opinions, findings and conclusions or recommendation expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation. Participants in the field crew, data analysis, and plant identification have included Corinna Gries, Diane Hope, Rob Bills, Quincy Stewart, Eric Holmes, Marty Feldner, Xiaoding Zhuo, Daniel Gonzales, Jason Walker, Sean Witcomb, Elizabeth Makings, Shannon Doan, Donald Pinkava, and Dixie Damrel. Molecular lab work was supported by funds provided to MFW by Arizona State University. We thank Tim Lowrey, Robbert Gradstein, Elvira Hörandl, and Rosemary Wilson for their helpful reviews of our manuscript.

LITERATURE CITED

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

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

Beyra-M., A. & Lavin, M. 1999. Monograph of *Pictetia* (Leguminosae-Papilionoideae) and review of Aeschynomeneae. *Syst. Bot. Monogr.* 56: 1–93.

Boom, B.M. 1996. Societal and scientific information needs from plant collections. Pp. 16–27 in: Stuessy, T.F. & Sohmer, S.H. (eds.), *Sampling the Green World: Innovative Concepts of Collection, Preservation, and Storage of Plant Diversity.* Columbia University Press, New York.

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

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783–791.

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

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

Morgan, E.C. & Overholt, W.A. 2005. New records of invasive exotic plant species in St. Lucie County, Florida. *Castanea* 70: 59–62.

Panero, J.L., Jansen, R.K. & Clevinger, J.A. 1999. Phylogenetic relationships of subtribe Ecliptinae (Asteraceae: Heliantheae) based on chloroplast DNA restriction site data. *Amer. J. Bot.* 86: 413–427.

Swofford, D.L. 2002. *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods),* version 4. Sinauer Associates, Sunderland, Massachusetts.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 24: 4876–4882.

Wojciechowski, M.F., Sanderson, M.J. & Hu, J-M. 1999. Evidence on the monophyly of *Astragalus* (Fabaceae) and its major subgroups based on nuclear ribosomal DNA ITS and chloroplast DNA *trnL* intron data. *Syst. Bot.* 24: 409–437.

Appendix. Unknown specimens and taxa employed in molecular analyses with GenBank accession numbers.

Each entry begins with taxon identified by a combination of molecular and morphological criteria, followed by: the GenBank accession number for the nrDNA ITS sequences from that collection; the CAP LTER collection site and plant number for that site; ASU herbarium accession number; and list of taxa used in the phylogenetic analysis with their corresponding GenBank accession numbers and e-values from BLASTn searches using the unknown sequence as query sequence.

***Baccharis salicifolia* (Ruiz & Pavón) Pers. (Asteraceae), EF190032, AE91-27 (ASU 262955).** *Aster amellus* (AF046961; 0.0); *Aster glehnii* (AY722010; 0.0); *Baccharis dracunculifolia* (AF046958; 0.0); *Baccharis neglecta* (U97604; 0.0); *Baccharis* sp. (AY193800; 0.0); *Calotis cuneifolia* (AF497647; 0.0); *Conyza gouanii* (AF046948; 0.0); *Diplostephium rupestre* (AF046962; 0.0); *Doellingeria umbellate* (AF477625; 0.0); *Laennicia sophiifolia* (AF046964; 0.0); *Lagenifera pumila* (AF422124; 0.0); *Olearia arguta* (AF497661; 0.0); *Olearia elliptica* (AF497669; 0.0); *Olearia picridifolia* (AF497683; 0.0); *Olearia pimeleoides* (AF497673; 0.0); *Podocoma notobellidiastrum* (AF046963; 0.0); *Solidago juncea* (DQ005981; 0.0); *Vittadinia australis* (AF422140; 0.0).

***Brachychiton populneus* R. Br. (Sterculiaceae), EF190036, W181-23 (ASU 262952).** *Brachychiton populneus* (AJ277463; 0.0); *Brachychiton rupestris* (AY083654; 0.0).

***Eclipta prostrata* (L.) L. (Asteraceae), EF190031, AA191-31 (ASU 262954).** *Angelphytum tenuifolium* (AY303402; $4e^{-144}$); *Blainvillea rhomboidea* (AY303404; $2e^{-164}$); *Eclipta prostrata* (DQ005977; 0.0); *Helianthus atrorubens* (AF047935; $1e^{-140}$); *Helianthus carnosus* (AF047969; $>2e^{-106}$); *Helianthus resinosus* (AF047967; $>2e^{-106}$); *Helianthus salicifolius* (AF047955; $2e^{-139}$); *Kingianthus paniculatus* (AY303425; $6e^{-174}$); *Kingianthus paradoxus* (AY303426; $2e^{-179}$); *Melanthera remyi* (AY169242; $1e^{-162}$); *Monactis pallatangensis* (AY169241; 0.0); *Oblivia mikanioides* (AY303433; $5e^{-162}$); *Otopappus epaleaceus* (AY303434; $3e^{-160}$); *Otopappus verbesinoides* (AY303435; $1e^{-165}$); *Perymeniopsis ovalifolia* (AY303438; $5e^{-162}$); *Rensonia salvadorica* (AY303441; $3e^{-154}$); *Tilesia baccata* (AY303445; $7e^{-155}$).

***Fraxinus* sp. (Oleaceae), EF190035, U151-17 (ASU 262963).** *Abeliophyllum distichum* (AF534805; $4e^{-115}$); *Chionanthus retusus* (DQ120723; 0.0); *Fontanesia phillyreoides* (AF534815; $8e^{-138}$); *Forsythia mandschurica* (DQ022428; $8e^{-144}$); *Forsythia viridissima* (AF534810; $2e^{-120}$); *Fraxinus americana* (U82908; $4e^{-109}$); *Fraxinus anomala* (U82914; $1e^{-87}$); *Fraxinus biltmoreana* (U82911; $5e^{-96}$); *Fraxinus latifolia* (U82913; $6e^{-105}$); *Fraxinus pennsylvanica* (U82903; $1e^{-96}$); *Fraxinus texensis* (AF174623; 0.0); *Fraxinus tomentosa* (U82897; $3e^{-94}$); *Fraxinus velutina* (AF169855; 0.0); *Jasminum humile* (AF534818; $3e^{-116}$); *Ligustrum japonicum* (AF361299; $1e^{-145}$); *Osmanthus fragrans* (AF135190; $2e^{-156}$); *Philadelphus incanus* (DQ248970; $7e^{-135}$); *Syringa emodi* (AF277762; $2e^{-166}$); *Syringa meyeri* (AF277750; $9e^{-162}$); *Syringa microphylla* (AF277754; $2e^{-166}$); *Syringa patula* (AF277759; $2e^{-166}$); *Syringa reticulata* (AF297080; $6e^{-169}$); *Syringa wolfii* (DQ022423; $5e^{-139}$).

***Hedypnois cretica* (L.) Dum.-Cours. (Asteraceae), EF190030, AA181-5 (ASU 262951).** *Agoseris elata* (AJ633463; $2e^{-180}$); *Cicerbita alpina* (AJ633340; $1e^{-153}$); *Crepis nicaeensis* (AJ633351; 0.0); *Crepis turcica* (AJ633360; 0.0); *Hedypnois glabra* (AJ633308; 0.0); *Hedypnois rhagadioloides* (AJ633307; 0.0); *Helminthotheca echioides* (AF528491; 0.0); *Hypochaeris apargioides* (AF528439; 0.0); *Hypochaeris scorzonerae* (AF528462; 0.0); *Hypochaeris sessiliflora* (AF528463; 0.0); *Hypochaeris taraxacoides* (AF528466; 0.0); *Leontodon autumnalis* (AJ633316; $1e^{-175}$); *Leontodon crispus* (AF528488; 0.0); *Leontodon helveticus* (AF528484; $3e^{-161}$); *Leontodon hispidus* (AF528485; 0.0); *Leontodon muelleri* (AJ633315; 0.0); *Leontodon saxatilis* (AF528489; 0.0); *Leontodon saxatilis* (AJ633317; 0.0); *Leontodon taraxacoides* (AJ633314; 0.0); *Leontodon tuberosus* (AF528487; 0.0); *Picris hieracioides* (AF528490; 0.0); *Prenanthes purpurea* (AJ633343; $2e^{-153}$); *Taraxacum officinale* (AJ633290; $1e^{-175}$).

***Heterotheca* cf. *subaxillaris* (Lam.) Britt. & Rusby (Asteraceae), EF190033, L141-12 (ASU 262964).** *Brintonia discoidea* (AY523853; 0.0); *Chrysopsis gossypina* (AF046993; 0.0); *Chrysothamnus greenei* (AY171016; 0.0); *Chrysothamnus nauseosus* (U97605; 0.0); *Croptilon divaricatum* (AF251576; 0.0); *Croptilon rigidifolium* (U97606; 0.0); *Eastwoodia elegans* (AY170949; 0.0); *Ericameria cervina* (AY171008; 0.0); *Ericameria crispa* (AY171011; 0.0); *Ericameria nana* (AY171022; 0.0); *Ericameria nauseosa* (AY170952; 0.0); *Erigeron rhizomatus* (AF046992; 0.0); *Euthamia graminifolia* (AF046982; 0.0); *Heterotheca fulcrata* (U97615; 0.0); *Macronema discoidea* (U97636; 0.0); *Solidago gigantea* (DQ005979; 0.0); *Solidago simplex* (DQ005982; 0.0); *Solidago simplex* (DQ006069; 0.0); *Stenotus lanuginosus* (AY170962; 0.0); *Tonestus graniticus* (AY170968; 0.0); *Xylothamia pseudobaccharis* (AF477683; 0.0).

***Ligustrum* cf. *japonicum* Thunb. (Oleaceae), EF190034, S131-13 (ASU 262953).** *Ligustrum acutissimum* (AF361295; 0.0); *Ligustrum ibota* (AF361297; 0.0); *Ligustrum japonicum* (AF361299; 0.0); *Ligustrum massalongianum* (AF361293; 0.0); *Ligustrum obtusifolium* (AF361294; 0.0); *Ligustrum ovalifoilium* (AF361296; 0.0); *Ligustrum sempervirens* (AF361300; 0.0); *Ligustrum vulgare* (AF361298; 0.0); *Osmanthus fragrans* (AF135190; $3e^{-165}$); *Syringa amurensis* (AF297074; 0.0); *Syringa chinensis* (DQ022419; 0.0); *Syringa emodi* (AF277762; 0.0); *Syringa julianae* (AF277749; 0.0); *Syringa komarowii* (AF361286; 0.0); *Syringa meyeri* (AF277751; 0.0); *Syringa microphylla* (AF277754; 0.0); *Syringa oblata* (DQ022424; 0.0); *Syringa patula* (*A*F277759; 0.0); *Syringa pekinensis* (AF297075; 0.0); *Syringa pubescens* (AF277746; *Syringa reflexa* (AF361282; 0.0); *Syringa reticulata* (AF297078; 0.0); *Syringa reticulata* (DQ022417; 0.0); *Syringa tigerstedtii* (AF361287; 0.0); *Syringa velutina* (DQ022416; 0.0); *Syringa villosa* (*A*F277760; 0.0); *Syringa vulgaris* (DQ184479; 0.0); *Syringa wolfii* (AF361284; 0.0); *Syringa wolfii* (DQ022423; 0.0); *Syringa yunnanesis* (AF361285; 0.0).

***Lantana* sp. (Verbenaceae), EF190037, X171-17 (ASU 262962).** *Aloysia gratissima* (AY178651; $3e^{-157}$); *Aloysia macrostachya* (AY178652; $4e^{-156}$); *Aloysia wrightii* (AY178653; $4e^{-141}$); *Ceratotheca triloba* (AY178649; $6e^{-103}$); *Colpias mollis* (AJ616318; $2e^{-111}$); *Duranta erecta* (AF477781; $1e^{-159}$); *Glandularia wrightii* (AY178656; $2e^{-149}$); *Lantana camara* (AF437859; 0.0); *Lantana camara* (AF437873; 0.0); *Lantana urticoides* (AY178664; 0.0); *Phyla nodiflora* (AY178654; $1e^{-137}$); *Scrophularia peregrine* (AF375146; $6e^{-103}$); *Verbena bonariensis* (AY178661; $>5e^{-97}$); *Verbena bracteata* (AY178662; $5e^{-131}$); *Verbena officinalis* (AF477793; $3e^{-154}$); *Verbena urticifolia* (DQ006043; $2e^{-152}$).

***Nicotiana obtusifolia* Mertens & Galeotti (Solanaceae), EF190038, Z111-39 (ASU 262965).** *Anthocercis gracilis* (AJ492457; 0.0); *Nicotiana alata* (AJ492424; 0.0); *Nicotiana clevelandii* (AJ492444; 0.0); *Nicotiana glauca* (AJ492410; 0.0); *Nicotiana glutinosa* (AJ492433; 0.0); *Nicotiana nesophila* (AJ492442; 0.0); *Nicotiana obtusifolia* (AJ492430; 0.0); *Nicotiana obtusifolia* (DQ272593; 0.0); *Nicotiana palmeri* (AJ492451; 0.0); *Nicotiana pauciflora* (AJ492428; 0.0); *Nicotiana stocktonii* (AJ492443; 0.0); *Nicotiana sylvestris* (AJ492423; 0.0); *Nicotiana tabacum* (AJ012364; 0.0); *Nicotiana tabacum* (AJ492448; 0.0); *Nicotiana tomentosa* (AJ492449; 0.0).