Molecular Ecology (2000) **9**, 1549–1559

A universal molecular method for identifying underground plant parts to species

C. R. LINDER*, L. A. MOORE† and R. B. JACKSON‡
Section of Integrative Biology, University of Texas, Austin, TX 78712, USA, †Department of, Biological Sciences, Stanford University, Stanford, CA 94305, USA, ‡Department of Botany and, Nicholas School of the Environment, Duke University, Durham, NC 27708, USA

Abstract

As part of a large project to determine rooting depth and resource uptake on the Edwards Plateau of central Texas, we developed a DNA-based technique that allows the below-ground parts of all plants to be identified to the level of genus and usually to species. Identification is achieved by comparing DNA sequences of the internal transcribed spacer (ITS) region of the 18S–26S nuclear ribosomal DNA repeat, derived from below-ground plant material, with a reference ITS region database for plants at a site. The method works throughout plants because the plant ITS region can be PCR amplified using a set of universal primers. Congeneric species can usually be identified because the ITS region evolves relatively rapidly. In our study, all roots were easily identified to the level of genus; most congeneric species were identified solely by ITS sequence differences but some required a combination of ITS sequence data and above-ground surveys of species at a site. In addition to showing the feasibility and efficacy of our technique, we compare it with another DNA-based technique used to identify below-ground plant parts. Finally, we also describe a DNA extraction and purification technique that reliably provides high-quality DNA of sufficient quantity from roots so that PCR can be readily accomplished. Our technique should allow the below-ground parts of plants in any system to be identified and thereby open new possibilities for the study of below-ground plant communities.

Keywords: community composition, DNA sequence analysis, internal transcribed spacer, PCR, roots, species identification

Received 27 February 2000; revision received 22 May 2000; accepted 22 May 2000

Introduction

Plant roots influence the hydrology, biogeochemistry and primary productivity of terrestrial ecosystems (Nepstad *et al*. 1994; Jackson 1999). However, the study of below-ground ecological and evolutionary patterns and processes has been hampered by a lack of broadly applicable methods for identifying plants from their underground parts (roots, stems, buds, tubers, etc.; Caldwell *et al*. 1987; Casper & Jackson 1997). Identification of individual species within the rhizosphere would help elucidate how different species partition below-ground resources and which species influence different parts of the rhizosphere. For example, such a method could be used to assess the spatial distribution of below-ground community composition. It would also dramatically help researchers studying root interactions and allelopathy below ground (Schenk *et al*. 1999). Ideally, a method for identifying below-ground parts would be rapid, reliable and inexpensive.

As part of a project to determine the rooting depth and resource uptake of woody species on the Edwards Plateau in central Texas (Jackson *et al*. 1999), we developed a method that uses the internal transcribed spacer (ITS) region of the nuclear ribosomal DNA (rDNA) repeat to identify most underground plant parts to the level of species. The ITS region was chosen to identify roots because: (i) it can be amplified universally using the polymerase chain reaction (PCR) in all plants and fungi using primers anchored in the conserved 18S and 26S genes of the rDNA repeat (White *et al*. 1990; Baldwin *et al*. 1995); (ii) it is known to evolve rapidly in plants so that genera and, in many cases, species can be readily distinguished by sequence variation (Baldwin *et al*. 1995); (iii) it undergoes rapid concerted evolution (Zimmer *et al*. 1980; Moritz & Hillis 1996), which usually eliminates multiple alleles in a population in the absence of hybridization; and (iv) it is used extensively by

Correspondence: Dr C. R. Linder. Fax: 1-512-471-3878; E-mail: rlinder@mail.utexas.edu

plant systematists, so many ITS sequences are already available in the GenBank databank for comparison with sequences from roots. As of 10 January 2000 there were over 8800 ITS region entries for green plants in GenBank.

Here, we present full details of the method used in Jackson *et al*. (1999), including improvements in the technique and a description of potential problems other researchers may encounter. We also describe root extraction buffers and purification methods that we developed to overcome special phenological and chemical problems associated with acquiring high-quality DNA from the roots of some species. Finally, we compare our method of identifying below-ground plant parts with an independently developed method (Bobowski *et al*. 1999) which uses restriction fragment length polymorphisms (RFLPs) in the chloroplast-encoded gene *rbcL*.

Materials and methods

Development of a reference-sequence database

Our study area was the Edwards Plateau, a 100 000 km^2 karst region of central Texas. The vegetation is primarily savanna and woodland dominated by trees from the genera *Quercus*, *Juniperus*, *Ulmus*, *Celtis* and, farther to the west, *Prosopis* (Auken *et al*. 1980). To make positive identifications of woody species from DNA extracted from roots found in caves 5–65 m deep, we required a reference database of ITS region sequences from known samples. Of the woody species at our sites (Table 1), only the ITS region of *Fraxinus pennsylvanica* was in GenBank at the time of our study. For the remaining species, we obtained leaf, flower or seed samples from local plants. Because many species of *Quercus* are either known or suspected to hybridize (Van Valen 1976; Whittemore & Schaal 1991; Dumolin-Lapègue *et al*. 1997, 1999), multiple samples of three of the oak species were extracted (*Q. fusiformis*, *Q. sinuata*, *Q. buckleyi*) to test whether intraspecific polymorphisms exist that might complicate species identification using the ITS region. In addition, DNA was available for two species of *Carya* and one species of *Ulmus* (Table 1) not found at our study site. We included these species to assess more broadly the ability of the ITS region to distinguish among congeneric species in our study. DNA was extracted from above-ground material using either a standard 2× CTAB buffer (Doyle & Doyle 1987) or, if that failed, a 1-M boric acid 1× CTAB buffer described below. A standard Doyle & Doyle (1987) protocol was followed. DNA was further purified by either Elu-Quik (Schleicher & Schuell) or the DNeasy Plant Mini Kit (Qiagen).

Twenty nanograms of template DNA were used in 25-μL PCRs consisting of 30 mM tricine pH 8.4, 2 mM $MgCl_2$, 50 mM KCl, 5% acetamide (w/v), 100 μM each dNTP, 1.0 units *Taq* DNA polymerase and 30 nM of each amplification primer. The complete ITS region (ITS1, the 5.8S gene and ITS2) was amplified using primers 'ITS-4' (White *et al*. 1990) and a modification of 'ITS-5' (White *et al*. 1990) based on the sequence reported for glycine (Eckenrode *et al*. 1984; 5′-GGAAGGAGAAGTCGTAACAAGG-3′). Reactions were overlaid with mineral oil and run on an MJ Research

Table 1 Species for which reference ITS regions either were sequenced for this study (S) or were taken from GenBank (G). Species native to the Edwards Plateau are followed by 'E'

Species	Source	GenBank accession no.
Berberis trifoliolata (E)	S	AF174616
Bumelia lanuginosa (E)	S	AF174617
Carya glabra	S	AF174618
C. illinoiensis (E)	S	AF174619
C. ovata	S	AF174620
Celtis laevigata (E)	S	AF174621
Diospyros texana (E)	S	AF174622
Fraxinus texensis (E)	S	AF174623
F. pennsylvanica (E)	G	AH004982, AH004985, AH004986, U82894, U82895, U82900, U82901, U82902, U82903
Ilex decidua (E)	S	AF174624
I. vomitoria (E)	S	AF174625
Juglans nigra (E)	S	AF174626
Juniperus ashei (E)	S	AF174627
Matelea reticulata (E)	S	AF174628
Populus deltoides (E)	S	AF174629
Prosopis glandulosa (E)	S	AF174630
Quercus buckleyi (E)	S	AF174631, AF174632
Q. fusiformis (E)	S	AF174633, AF174634, AF176380
Q. sinuata var. *breviloba* (E)	S	AF174635, AF176381
Q. stellata (E)	S	AF174636
Sophora secundiflora (E)	S	AF174638
Ulmus americana (E)	S	AF174640
U. crassifolia (E)	S	AF174641
U. rubra	S	AF174642

Thermal Cycler programmed for a hot start (95 °C, 5 min; 74 °C, 7 min) and 30 cycles of 94 °C, 45 s denaturing; 50 °C, 45 s annealing and 72 °C, 40 s extension. A final extension of 72 °C for 7 min was also included. Following amplification, 3 μL of the reaction mixture was run on a 1.5% agarose gel in 0.5× TBE buffer to determine whether an appropriately sized product had been produced. Products were visualized using ethidium bromide and their sizes were determined using a standard 1-kb ladder (Gibco).

PCR products were sequenced directly following purification (QIAquick PCR Purification Kit, Qiagen). All DNA sequencing for this project was performed on an ABI Prism 377 automated sequencer using standard reagents and protocols (Perkin-Elmer, California). The staff of the Core Facility at the Institute of Cellular and Molecular Biology, University of Texas, performed the sequencing reactions. For all reference sequences, with the exception of the *Q. sinuata* sample from Powells Cave, both strands of the complete ITS region were sequenced. For most species, sequencing was accomplished with the ITS-5 and ITS-4 primers (White *et al*. 1990), but for one sample of *Q. fusiformis* and one of *Q. sinuata* a poly(G) tract prevented sequencing of most of the ITS region with the ITS-5. For these samples, the internal primers ITS-2 and ITS-3 (White *et al*. 1990) were used to provide coverage of the remaining areas. For the *Q. sinuata* sample from Powells Cave, only the sequence primed by the ITS-4 primer was used.

Complete ITS region sequences for each sample were assembled using SeqMan II (DNASTAR 1999b). All ABI traces were checked by eye for misread bases and polymorphisms before the sequences were assembled. Intra-individual polymorphisms that could be confirmed on both strands were retained. Diagnostic interspecific polymorphisms were assessed by aligning the sequences using the CLUSTAL method (Thompson *et al*. 1994) as implemented in MEGALIGN (DNASTAR 1999a).

Extensive use of the ITS region for phylogenetic reconstruction has shown that intergeneric comparisons routinely reveal multiple diagnostic differences in the form of point mutations and often insertions and deletions (reviewed in Baldwin *et al*. 1995). Intragenerically, the level of diagnostic differences between species varies from multiple differences to none. In our study, intergeneric polymorphisms were considered diagnostic for a species when multiple base substitutions or insertions and deletions were unique for the species or when combinations of changes uniquely identified a species. Intragenerically, differences were only considered diagnostic if several (>4) unique differences were found. We did not perform extensive population level sampling to determine if small numbers of intrageneric differences were diagnostic for our species because above-ground vegetation surveys, in combination with the ITS sequence data, always uniquely identified the roots in individual caves.

Sampling of roots for identification by ITS sequence analysis

We sampled 18 caves across the central and eastern Edwards Plateau (Table 2), ranging in depth from 5 to 65 m (Jackson *et al*. 1999). Roots were discovered in 14 caves, and, from these, fine roots were obtained either from soil on the floor of the caves or, in a few cases, from underground streams. Individual roots were placed in sealed plastic bags with adequate moisture to keep them hydrated. They were placed on ice if they could not be brought to the laboratory within 48 h. Once in the laboratory, roots were cleaned thoroughly of all soil by rinsing first in tap water and then distilled water.

In addition to collecting roots, vegetation surveys were conducted above each cave (Jackson *et al*. 1999) to independently verify the species which might have roots in the caves and to narrow the set of species with which to compare our root-derived ITS region sequences.

Development of root extraction buffers and a DNA purification procedure

Early in the study, it was apparent that the standard Doyle & Doyle (1987) buffer failed to extract DNA from many root samples and that, even when DNA could be extracted successfully from roots, standard purification procedures (Elu-Quik, Schleicher & Scheull, and DNeasy Plant Mini Kit, Qiagen) left contaminants that interfered with the PCR. We tested five extraction buffers to determine which could most consistently recover intact DNA in the largest quantities from root tissue. The first buffer was the Doyle & Doyle (1987) 2 × CTAB buffer, which served as a standard. The remaining buffers were: a 2 × CTAB extraction buffer in which the concentration of Tris was increased to 1 M; a 1× CTAB extraction buffer in which the concentration of Tris was 1 M (Steenkamp *et al*. 1994); and 1× CTAB buffers in which boric acid was substituted for Tris at concentrations of 0.01 M and 1 M. Once we determined that the 1 M boric acid 1× CTAB buffer was the most efficient at recovering DNA from roots (see Results), we used it consistently. Up to 0.2 g of the finest and freshest root material available was ground in liquid N_2 using a pestle and mortar. Occasionally, only larger, older roots were available for extraction. In these cases, the layer outside the xylem was peeled and extracted to reduce the proportion of woody tissue with low DNA content in the extraction. The extraction procedure followed the protocol of Doyle & Doyle (1987), but a chloroform : phenol step (Sambrook *et al*. 1989) was added before performing chloroform : isoamyl alcohol purification. The quantity and quality of DNA from extractions were assessed using a DyNAQuant 2000 DNA fluorometer (Pharmacia) and by running 3 μL of the extraction on a 0.5× TBE agarose gel. Gels were stained with EtBr and visualized using UV light.

Table 2 Caves, plant species identified by ITS sequence, GenBank accession numbers of ITS sequences, and the surface vegetation

Cave	Species identified by ITS from roots	GenBank accession nos	Woody species on the surface
District Park Cave	*Q. fusiformis*	AF176410, AF176411	*J. ashei*, *Q. fusiformis*
Seven Room Cave	*U. crassifolia*	AF176417	*C. laevigata*, *U. crassifolia*
Pearson's Lost Cave	*C. laevigata*	AF176414	*C. laevigata*, *J. ashei*
	J. ashei	AF176413	
Spyglass Cave	*Q. fusiformis*	AF176398, AF176399, AF176400	*C. laevigata*, *J. ashei*, *Q. fusiformis*
Cotterell Cave	*J. ashei*	AF231979, AF231980	*C. laevigata*, *J. ashei*, *Q. sinuata*
	Q. sinuata	AF176392, AF176393, AF176394	
Mystery Hole	*U. crassifolia*	AF176384, AF176385	*J. ashei*, *U. crassifolia*
Sour Cave	*U. americana*	AF176387, AF176388	*J. ashei*, *U. americana*
Turtle Shell Cave	*J. ashei*	AF231974, AF231975, AF231976, AF231977, AF231978	*J. ashei*, *Q. fusiformis*, *U. crassifolia*
	Q. fusiformis	AF176395	
Cave of Many Names	*U. crassifolia*	AF176386	*J. ashei*, *U. crassifolia*
Natural Bridge (south cave)	*Q. fusiformis*	AF176407, AF176408	*J. ashei*, *Q. fusiformis*, *U. crassifolia*
Sweet Cave	*Q. fusiformis*	AF176396, AF176397	*J. ashei*, *Q. fusiformis*, *U. crassifolia*
Cicurina Cave	*Q. fusiformis*	AF176412	*C. laevigata*, *C. illinoiensis*, *J. ashei*, *Q. fusiformis*, *Q. sinuata*, *Q. stellata*, *U. crassifolia*
Powells Cave	*B. lanuginosa*	AF176382, AF176383, AF176415, AF176416	*B. lanuginosa*, *C. laevigata*, *P. glandulosa*, *Q. fusiformis*, *Q. sinuata*
	Q. fusiformis	AF176401, AF176402, AF176403, AF176404, AF176405, AF176406	
Honey Creek Cave	*Q. fusiformis*	AF176409	*Q. fusiformis*

Both methods of determination were necessary because contaminants in the extracted DNA occasionally interfered with the Hoechst dye used by the fluorometer, producing artificially high readings. If a sample was determined to have sufficient DNA, the remainder was gel purified using the QIAEX II Gel Extraction Kit (Qiagen). The 0.5× TBE purification gels were run for a minimum of 1.5 h to ensure separation of contaminants from the DNA. Pigmented contaminants, which ran more rapidly than DNA, were often visible in the gel. We did not test whether they were responsible for interfering with PCR amplifications when present.

Amplification and sequencing of the ITS region from DNA extracted from roots

Gel-purified DNAs were used as templates in separate PCR amplifications of the ITS region of the rDNA repeat. Twenty-five microlitre PCR reactions, identical to those run for the reference samples, were performed to amplify the complete ITS region. PCR products were sequenced using the same procedures and facilities as the reference samples. Initially, we sequenced just the ITS1 region using the ITS-5 primer; however, a poly(G) stretch early in the ITS1 of several samples prevented reliable sequencing of downstream regions. Subsequently, we used the ITS-4 primer to sequence the ITS2 region. Sequence variation in the ITS2 alone was sufficient to distinguish among species within a site. All ABI traces were inspected by eye and corrected using SEQMAN II (DNASTAR 1999b). Ambiguous bases were retained using standard IUB symbols (A, adenosine; G, guanosine; C, cytidine; T, thymidine; M = A or C, R = A or G, W = A or T, S = C or G, Y = C or T, K = G or T, N = unreadable). Only regions in which the traces could be scored clearly were retained.

Identification of roots

To identify roots, ITS region sequences from unknown samples were aligned with reference sequences using the pairwise alignment feature in MEGALIGN (DNASTAR 1999a). An unknown sample was considered to be identified if its ITS sequence matched perfectly that of a reference species or if it differed from the reference sequence by only one to three base pairs and was not more similar to some other reference sequence. For the few congeneric species that could not be identified unambiguously by DNA sequence data alone, the above-ground surveys were used to eliminate species not growing above the cave from which the roots were collected.

Results

Reference database

ITS region sequences for all reference species collected for this study have been annotated and deposited in GenBank (Table 1). In all cases, species in different genera could be readily distinguished by multiple sequence differences

```
Q.fus.campus TCGAAACCTGCACAGCAGAACGACCCGCGAATTGGTTACAACCGACGGGGGGCGGGGGGCGNTCNTCNNCCCCTCGCCCCTTCCTG 118
U.cra        ..........C..................CAC...GTT.A-AC.........A...A.CTC.GGGCC..GAC.CT...CCGGC. 116

Q.fus.campus TGGGTGGGGACCTTGTGTCTCCTGCCCGCAAACCGAACCCCGGCGCGGAACGCGCCAAGGAAATCTAACCAAGAGAGCCATGCCGG 204
U.cra        CT..CC.C.G.ACC.CCGTG.GCCAG..TC..A............TATCT............C.A...G..CGAGCG.GC.A... 202

                                                      <--- ITS1 | 5.8S --->
Q.fus.campus AGGCCCCGGACACGGTGCGCCCCGGCGTCGGCGTCTTATGAATTAT-TCAAAACGACTCTCGGCAACGGATATCTAGGCTCTCGCA 289
U.cra        TC.G......A........-.GT...A.CT.-....G.C.TCGA...G.............................C..... 286

Q.fus.campus TCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAATCCCGCGAATCATCGAGTTTTTGAACGCAAGTTGCGCCC 375
U.cra        ..................................................T...C........C.................... 372

                                                <--- 5.8S | ITS2 --->
Q.fus.campus GAAGCCATTCGGCCGAGGGCACGTCTGCCTGGGTGTCACGCATCGTTGCCCCCCCAAACTCCGGTTCGGGCAG--------GGCGG 453
U.cra        ......T.CG.....................C.....A..C............C..AC..C..CG......AAGGGTGG..... 458

Q.fus.campus AAGTTGGCCTCCCGTGCGTGCCTGCACGCGCGGTTAGCCCAAAAGCGAGTCCTCGGCGACGAGCGCCACGACAATCGGTGGTTTTT 539
U.cra        .T.C............A.C-.YC..-.T.....C.G.......T.....AT..AC..TG......T.G..G.G..-.......G.C 541

Q.fus.campus TT-TACCCTCGTTCCTCGTCGTGCGTGCCCCGTYGCCCGAACGCGCTCCTGCGACCCTCACGCGTCGCCTCG-GTGGCGCTCCCAA 623
U.cra        GAA..A.TCG..G..C.....C.AAC..T..-CG..GT.C-..TCTCGGAA......CTG....CG......C.CC.....T.... 625
```

Fig. 1 CLUSTAL W alignment of the complete ITS regions for *Quercus fusiformis* (Q.fus.campus) and *Ulmus crassifolia* (U.cra) reference sequences. The *Q. fusiformis* sequence is used as a reference sequence for the *U. crassifolia* sequence. Agreement with the *Q. fusiformis* sequence is indicated by periods and disagreement by a nucleotide at a position. Gaps introduced to improve the alignment are indicated by dashes. Standard IUB symbols for nucleotides are used. The boundaries between ITS1, 5.8S and ITS2 are indicated above the sequences.

```
U.ame TCGAAACCTGCCCAGCAGAACGACCCGCGAACACGTTGTTAAACCGGGGGGGCGAGGGACCTCCGGGCCCCGACCCTCCCCCGGCG 117
U.cra ..................................................................................... 116
U.rub .....................................................TT.............GT..................... 113

U.ame CTGGCCGCGGCAMMGCYGYGCGCCAGCGTCAAACGAACCCCGGCGCTATCTGCGCCAAGGAAACCAAACGAACGAGCGCGCGACG 202
U.cra .............CC..C.T..................................................................... 201
U.rub .............CT..C.T..................................................................... 198

                                                        <--- ITS1 | 5.8S --->
U.ame GTCGGCCCGGAAACGGTGCCGTCGGAGCCGCGTCGTCTTCGATATGTCAAAACGACTCTCGGCAACGGATATCTCGGCTCTCGCA 287
U.cra ............................T..................................................... 286
U.rub ............................T..................................................... 283

U.ame TCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTCTTTGAACGCAAGTTGCGCC 372
U.cra ..................................................................................... 371
U.rub ..................................................................................... 368

                                                  <--- 5.8S | ITS2 --->
U.ame CGAAGCCATCCGGCCGAGGGCACGTCTGCCTGGGCGTCACACGCCGTTGCCCCCCC==AACCCCGTCGGGGRAAGCGGG--GGGC 453
U.cra .......T..G.............................A.............CA.............C.GAA...TG.... 456
U.rub .....................................A.............=A.............A.GAA...--.... 450

U.ame GGATGCTGGCCTCCCGTGAGCCTCGCCTCGCGGCTGGCCCAAATGCGAGATCTCTGCTGCGGGCGTCGCGGCGATGGTGGTTGTC 538
U.cra ......................Y..............................AC......A..................... 541
U.rub ..........................................................A..................... 535

U.ame GAATAACTCGGTGCCCCGTCGCGAGCGCCCCCGGCGTGCCGTCTCGGAACGACCCCTGCGCGCGGCCTCGCGCCGCGCTTCCAA  622
U.cra ........................A...T..................................................  625
U.rub ........................A.......A......T..................................... 619
```

Fig. 2 CLUSTAL alignment of the ITS regions from three reference *Ulmus* species: *U. americana* (U.ame), *U. crassifolia* (U.cra) and *U. rubra* (U.rub). The *U. americana* sequence is used as a reference for the other two sequences. The conventions used in Fig. 1 to represent agreement and disagreement with the reference sequence are followed, except that gaps that uniquely identify a species are represented with bold equal signs. Nucleotide differences that uniquely identify a species are indicated in bold. The boundaries between ITS1, 5.8S and ITS2 are indicated above the sequences.

(nucleotide differences or gaps). Even the most closely related pair of genera, *Quercus* and *Ulmus*, had a large number of diagnostic differences (Fig. 1). Within genera, the ability to distinguish species varied. At one extreme, *Ilex decidua* and *I. vomitoria* had identical ITS regions. At the other extreme, the three *Ulmus* species each had eight unique differences (Fig. 2). The results for the three *Carya* species were less consistent. *C. illinoiensis* had seven unique

```
C.gla TCGATACCTGCCCAGCAGAACGACCCGTGAACTTGTAATAACCTTCTGGGTTGGGGTGTCATGCCCCCTCCCAAAAACGGTTGGG 113
C.ill ..................................................................................... 116
C.ova ..................................................................................... 116

C.gla AGGGCACGTTGAAAGCTGCCCACCGCTCCTCGTGTGTGGCGGGTCAGTCTCCTCGTTCCCTTCCCAGTCGAACAATGAAACCCCG 198
C.ill ........A..........................................................G.............=..... 200
C.ova ..................................................................................... 201

C.gla GCGCGGTCTGCGCCAAGGAACTCAAACAAGGAGTAACCACGGCCGCCCCGGAAACGGTGTGCGTGCCGTTGGTGACATCTTACCA 283
C.ill ........................................T.......................................... 285
C.ova ..................................................................................... 286

      <--- ITS1 | 5.8S --->
C.gla CGATACATAACGACTCTCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTG 368
C.ill ..................................................................................... 370
C.ova ..................................................................................... 371

                                                                    <--- 5.8S | ITS2 --->
C.gla CAGAATCCCGCGAATCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCATTCGGCCGAGGGCACGTCTGCCTGGGTGTCACGC 453
C.ill ..................................................................................... 455
C.ova ..................................................................................... 456

C.gla ATCGTTGCCCCCACCCCAACACCTCGTATGATGTGTGCGGTGCGGGGCAGAAATTGGCCTCCCGTGCGCGTTGCTCGCGGTTAGC 538
C.ill ....................................A.............................T.................. 540
C.ova ..........................C.......................................................... 541

C.gla CTAAAAGCGAGTCCTAGGCGACGAGCGCCACGACAATCGGTGGTTGACAAAACCCTCGTGCCCCGTCGTGCGTTGCTCGACGCCG 623
C.ill ..............G..................................................................... 625
C.ova ..................................................................................... 626

C.gla TGAAGGCTCTCCTCGACCCTATTGCGTCGTTCTTGCGACTCTACCAT                                   670
C.ill ..........................................                                       672
C.ova ..........................................                                       673
```

Fig. 3 CLUSTAL alignment of the ITS regions from three *Carya* species: *C. glabra* (C.gla), *C. illinoiensis* (C.ill) and *C. ovata* (C.ova). The *C. glabra* sequence is used as a reference for the other two sequences. The conventions used in Fig. 2 to represent agreement and disagreement with the reference sequence are followed. Differences that uniquely identify a species are indicated in bold. The boundaries between ITS1, 5.8S and ITS2 are indicated above the sequences.

differences, *C. ovata* had one and *C. glabra* had none, although it might be distinguished from the others by a single combination of bases (Fig. 3). The *Quercus* species exhibited a high degree of polymorphism within and between individuals and species (Fig. 4). The within-individual polymorphisms are indicated by the large number of positions with two or more bases. The high degree of within-species polymorphism prevented many unique interspecific differences from occurring. Only *Q. buckleyi*, the sole red oak species, had a large number of unique differences (12). *Q. stellata* had three possible unique nucleotide differences and one unique gap. However, given the amount of within-species polymorphism in *Q. fusiformis* and *Q. sinuata*, more extensive sampling of *Q. stellata* is needed to determine if the interspecific differences seen here are consistent. *Q. fusiformis* and *Q. sinuata* could not be distinguished by unique or combinatorial differences.

Effectiveness of root extraction buffers and DNA purification methods

The standard Doyle & Doyle (1987) 2 × CTAB extraction buffer generally did a poor job of extracting DNA from root tissues. Only two samples yielded DNA that could be amplified readily using PCR. In contrast, early in the growing season, the 1 M Tris 1× CTAB extraction buffer was able to extract intact DNA from all samples, as visualized on EtBr-stained agarose gels (data not shown). As the season progressed, the 1 M Tris 1× CTAB buffer began to fail to extract DNA from roots, perhaps due to increased plant phenolics (see Discussion). By late July we could no longer extract sufficient DNA for PCR amplifications following DNA purification. The 1 M boric acid 1× CTAB buffer was able to extract root DNA even late in the season.

No matter which extraction buffer was used, Elu-Quik and DNeasy purification methods usually failed to produce

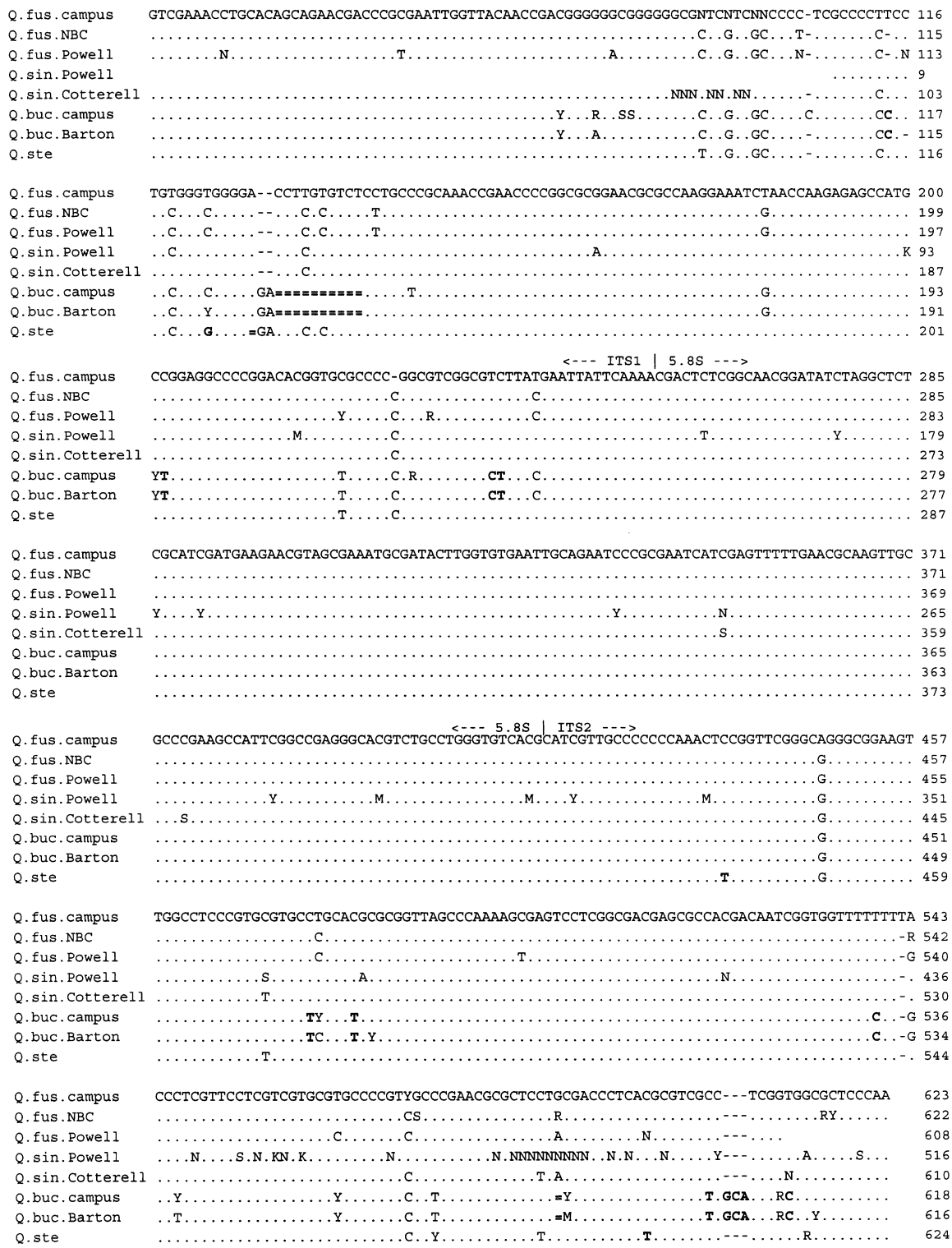

Fig. 4 CLUSTAL alignment of the ITS regions from four *Quercus* species: *Q. fusiformis* (Q.fus), *Q. sinuata* (Q.sin), *Q. buckleyi* (Q.buc) and *Q. stellata* (Q.ste). Three *Q. fusiformis*, two *Q. sinuata* and two *Q. buckleyi* sequences are shown to illustrate intraspecific polymorphisms within those species. Names of intraspecific sequences indicate the locations from which leaf material was collected: campus, University of Texas at Austin campus, NBC, Natural Bridge Cavern; Powell, Powells Cave; Cotterell, Cotterell Cave; Barton, Barton Creek. The Q.fus.campus sequence is used as a reference for the other sequences. The conventions used in Fig. 2 to represent agreement and disagreement with the reference sequence are followed. Differences that uniquely identify *Q. buckleyi* and *Q. stellata* are indicated in bold. The boundaries between ITS1, 5.8S and ITS2 are indicated above the sequences.

root DNA pure enough for PCR reactions. Early in the growing season we were occasionally able to amplify the ITS region with 10 : 1 to 50 : 1 dilutions of template DNA (undiluted concentration: 20 ng per 25 μL PCR reaction), however, even diluting the template failed as the growing season progressed. Only gel purification with the QIAEX II Gel Extraction Kit consistently yielded DNA that was free from contaminants that interfered with PCR amplification, although the yields from this purification method were low. When running gels for purification, it was evident in many samples that a pigmented fraction was being separated from the DNA. We did not perform any analyses to determine the nature of the contaminants that interfered with PCR amplification.

Identification of root samples

Roots for which the ITS region amplified well consistently produced legible DNA sequences for the ITS2 region, and all roots that yielded amplifiable DNA were ultimately identified to species using sequence data alone or sequence data in combination with our above-ground surveys of species (Table 2). A total of 35 samples, representing seven species, was identified by their ITS sequences. The taxonomic range of species identified was broad, including a conifer and four genera of woody dicots. By far, the most common species was *Q. fusiformis* ($N = 18$).

At the Powell Cave site, we were initially unable to identify the *Bumelia lanuginosa* root samples because our reference database did not include *B. lanuginosa*. *B. lanuginosa*, a small tree or shrub, was left out of our database because we were unaware of the presence of the species at the site and it did not occur above any of the other caves we sampled. Access to Powells Cave is restricted to three times per year, and since our root samples were collected during the winter, the naked branches of *B. lanuginosa* were not recognized. Our inability to identify the *B. lanuginosa* roots with our initial set of reference sequences led us to consider other species that might be at the site or possible fungal contamination. A BLAST search (Altschul *et al*. 1997; Zhang & Madden 1997) on GenBank showed no similarity to any fungal ITS sequences and very poor similarity to any plant sequences. In spring 1999, we obtained *B. lanuginosa* leaves (kindly provided by T. Wendt) and identified the roots by their perfect alignment with the reference *B. lanuginosa* sequence. We subsequently returned to Powells Cave during the 1999 growing season and identified several *B. lanuginosa* individuals near the cave.

Discussion

As long as DNA can be extracted and purified from below-ground plant parts, our method of identification using the ITS region can reliably identify all samples to the level of genus and, in most cases to the level of species. Our experience and that of Steenkamp *et al*. (1994) suggests that extraction and purification of DNA from roots can be challenging. In our study, it was the most difficult aspect of using DNA for species identification. We and others (Steenkamp *et al*. 1994) have found that more highly buffered extraction solutions are essential to maintain pH in a suitable range for dissolving the extracted DNA. In our work, solutions buffered at low levels, such as the Doyle & Doyle (1987) 2× CTAB buffer, became too acidic (data not shown). However, pH is clearly not the sole determinant of successful DNA extraction since 1 M Tris and 1 M boric acid 1× CTAB solutions had very different abilities to extract DNA from roots later in the season. Further development of extraction buffers for woody roots is desirable to find ones that eliminate the need for gel purification following extraction. This is especially important when only small quantities of DNA are likely to be recovered due to limited quantities of root tissue.

Although we did not determine what caused the seasonal variation in the ability of the 1 M Tris 1× CTAB extraction buffer to successfully harvest DNA from woody roots, two possibilities seem particularly likely. First, the proportion of DNA in the roots may have decreased as the season progressed because the roots were less actively growing and becoming more highly lignified or supported with cellulose. Second, the concentration of secondary compounds may have increased as the season progressed. In some plants, the concentration of secondary compounds in young leaves is lower than in mature leaves (e.g. Langenheim 1994; Forbes *et al*. 1995; Lerdau *et al*. 1995; Llusià & Peñuelas 2000). It may be that a similar process takes place in the roots of the species studied here. Roots that were ultimately determined to be *Quercus*, a genus known to have large amounts of tannin, were the most difficult from which to extract and purify DNA. Prior to gel purification, *Quercus* DNA solutions were obviously contaminated with pigmented compounds which could only be separated from the DNA by agarose electrophoresis.

Universality and accuracy of the method

Use of the ITS region for identification of roots should have broad applicability to all green plants, opening avenues of investigation into below-ground processes in plants that have previously been difficult or impossible to study (Jackson *et al*. 1999). Because the ITS region can be amplified throughout green plants using the same set of primers (Baldwin *et al*. 1995), the ITS of any below-ground part from which DNA can be extracted successfully and purified can be used to identify species. The most telling aspect of the method's accuracy was our discovery that we had overlooked *Bumelia lanuginosa* above-ground at

Powells Cave only after we identified *B. lanuginosa* root ITS sequences from the site.

Even so, the ITS region sequence by itself will not always provide an unambiguous means of identifying all species as is evident from our ITS sequences for *Ilex* and *Quercus* and examples from the systematics literature (Soltis & Kuzoff 1993; Bain & Jansen 1995; Bakker *et al*. 1995; Baldwin *et al*. 1995; Francisco-Ortega *et al*. 1997; Schilling *et al*. 1998). In our study, surveys of the above-ground vegetation allowed us to make unambiguous assignments of species, especially *Quercus*, but there will undoubtedly be communities where this will not be the case. For these situations, more rapidly evolving regions or markers will be needed if DNA is to be used to identify roots. Some candidates for this level of identification are randomly amplified polymorphic DNA (RAPD; Rieseberg *et al*. 1994; Chowdari *et al*. 1998; Pejic *et al*. 1998), intersimple sequence repeats (ISSRs; Zietkiewicz *et al*. 1994; Godwin *et al*. 1997; Blair *et al*. 1999; Prevost & Wilkinson 1999) and amplified fragment length polymorphisms (AFLPs; Vos *et al*. 1995; Hongtrakul *et al*. 1997; Pejic *et al*. 1998; Barcaccia *et al*. 1999; Krauss 1999; Palacios *et al*. 1999; Zhu *et al*. 1999). All three methods have been shown to generate large amounts of variation, often allowing individuals to be fingerprinted. The AFLP method is especially promising because it works with very small quantities of DNA and is highly repeatable (Vos *et al*. 1995; Hongtrakul *et al*. 1997; Waugh *et al*. 1997; Zhu *et al*. 1999).

Another area of concern for closely related species is intraspecific polymorphisms and hybridization. Even our small number of intraspecific samples of *Quercus* revealed that ITS region polymorphisms can make species-level identification by sequence similarity impossible. Other genera that experience high levels of interspecific hybridization are likely to present similar problems. The only way to identify such problems is to sample a sufficiently large number of individuals from the area of study so that the intraspecific polymorphisms can be characterized. It should also be noted that polymorphisms due to hybridization will likely be a problem for all nuclear DNA regions and organellar markers where gene flow is bidirectional.

Finally, because the same primers that are used to amplify the ITS region in plants will also amplify fungal ITS regions (White *et al*. 1990), we were concerned initially that fungi on roots or mycorrhizal associations with the roots would interfere with our ability to obtain unambiguous plant ITS region sequences from roots. This was not a problem even though all of the species we identified by root DNA form mycorrhizal associations (Brundrett 1991). Few of our root samples were from larger roots that would not be expected to have mycorrhizal infections, and there are no data addressing whether roots at the depths we sampled are normally infected by mycorrhiza. However, during the development of our techniques, we extracted DNA and amplified the ITS region from shallow roots of several woody species in our reference database. These roots should have been infected, but, here again, we did not detect fungal contaminants in our ITS sequences (data not shown). It is therefore likely that the concentration of fungal DNA in roots will be too low relative to root DNA to interfere with successful amplification of plant ITS sequences.

Comparison of methods

Bobowski *et al*. (1999) recently published a method for identifying below-ground plant parts using RFLPs detected in *rbcL*, a chloroplast encoded subunit of RUBISCO, following PCR-amplification. Both the method of Bobowski *et al*. (1999) and our method amplify regions of DNA using primers that are well conserved throughout green plants. Also, both methods require only a small amount of purified DNA to successfully amplify their respective regions. However, use of the DNA sequence from the ITS region has several advantages over restriction digests of *rbcL* for identifying plants to species. The ITS region usually evolves more rapidly than *rbcL* (Palmer *et al*. 1988; Baldwin *et al*. 1995) and is routinely used by plant systematists for lower level phylogenetic reconstructions rather than *rbcL* because of its rapid evolution (Baldwin *et al*. 1995). Therefore, the ITS region should resolve the identity of below-ground plant structures with greater accuracy in a larger number of cases than *rbcL*. In addition, DNA sequence data are more likely to reveal diagnostic differences than RFLPs because RFLPs reveal only changes at restriction sites or length variation large enough to be detected on agarose gels. DNA sequences reveal all possible diagnostic changes.

Although the DNA sequence of the ITS region usually identifies plants to species, the cost of sequencing reactions may be greater than that of restriction digests. We therefore suggest the following strategy for the identification of below-ground plant parts. When the species of interest are in different genera, either *rbcL* or the ITS region could be amplified and surveyed for diagnostic RFLPs. When species are in the same genus or are known to have diverged recently, the reference ITS sequences could be searched for diagnostic restriction sites. If consistent diagnostic restriction sites can be identified, then restriction digests could be employed on PCR-amplified ITS sequences for the below-ground samples. If not, then the project could proceed with DNA sequencing.

Molecular methods for identifying below-ground parts of plants have opened new avenues of research concerning below-ground ecological processes in plants (Bobowski *et al*. 1999; Jackson *et al*. 1999), and new techniques should continue to expand the questions that can be addressed. We are currently developing fingerprinting methods for roots, which should make it possible to take the final step

of unambiguously connecting individual roots to above-ground stems in nonclonal species, allowing study of above- and below-ground plant physiological processes on a per plant basis in natural systems.

Acknowledgements

We thank Amy George and Florencia Perret for assistance in the laboratory and two anonymous reviewers for their comments on the manuscript. CRL gratefully acknowledges the support of the National Science Foundation. RBJ gratefully acknowledges support from the National Science Foundation, the Andrew W. Mellon Foundation, the United States Department of Agriculture and the InterAmerican Institute for Global Change Research.

References

Altschul SF, Madden TL, Schaeffer AA *et al*. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389–3402.

Auken OW, Ford AL, Stein A, Stein AG (1980) Woody vegetation of upland plant communities in the southern Edwards Plateau. *Texas Journal of Science*, **32**, 23–35.

Bain JF, Jansen RK (1995) A phylogenetic analysis of the aureoid *Senecio* (Asteraceae) complex based on ITS sequence data. *Plant Systematics and Evolution*, **195**, 209–219.

Bakker FT, Olsen JL, Stam WT (1995) Evolution of nuclear rDNA ITS sequences in the *Cladophora albida/sericea* clade (Chlorophyta). *Journal of Molecular Evolution*, **40**, 640–651.

Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ (1995) The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden*, **82**, 247–277.

Barcaccia G, Albertini E, Falcinelli M (1999) AFLP fingerprinting in *Pelargonium peltatum*: its development and potential in cultivar identification. *The Journal of Horticultural Science and Biotechnology*, **74**, 243–250.

Blair MW, Panaud O, McCouch SR (1999) Inter-simple sequence repeat (ISSR) amplification for analysis of microsatellite motif frequency and fingerprinting in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics*, **98**, 780–792.

Bobowski BR, Hole D, Wolf PG, Bryant L (1999) Identification of roots of woody species using polymerase chain reaction (PCR) and restriction fragment length polymorphism analysis. *Molecular Ecology*, **8**, 485–491.

Brundrett MC (1991) Mycorrhizas in natural ecosystems. *Advances in Ecological Research*, **21**, 171–313.

Caldwell MM, Richards JH, Manwaring JH, Eissenstat DM (1987) Rapid shifts in phosphate acquisition show direct competition between neighbouring plants. *Nature*, **327**, 615–616.

Casper BB, Jackson RB (1997) Plant competition underground. *Annual Review of Ecology and Systematics*, **28**, 545–570.

Chowdari KV, Venkatachalam SR, Davierwala AP, Gupta VS, Ranjekar PK, Govila OP (1998) Hybrid performance and genetic distance as revealed by the $(GATA)_4$ microsatellite and RAPD markers in pearl millet. *Theoretical and Applied Genetics*, **97**, 163–169.

DNASTAR I (1999a) *MEGALIGN 4.00 edit*. DNASTAR, Inc., Madison, WI.

DNASTAR I (1999b) *SEQMAN II 4.00 edit*. DNASTAR, Inc., Madison, WI.

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

Dumolin-Lapègue S, Demesure B, Fineschi S, Le CV, Petit RJ (1997) Phylogeographic structure of white oaks throughout the European continent. *Genetics*, **146**, 1475–1487.

Dumolin-Lapègue S, Kremer A, Pettit RJ (1999) Are chloroplast and mitochondrial DNA variation species independent in oaks. *Evolution*, **53**, 1406–1413.

Eckenrode VK, Arnold J, Meagher RB (1984) Comparison of the nucleotide sequence of soybean 18S rRNA with the sequences of other small-subunit rRNAs. *Journal of Molecular Evolution*, **21**, 259–269.

Forbes TDA, Pemberton IJ, Smith GR, Hensarling CM (1995) Seasonal variation of two phenolic amines in *Acacia berlandieri*. *Journal of Arid Environments*, **30**, 403–415.

Francisco-Ortega J, Santos Guerra A, Hines A, Jansen RK (1997) Molecular evidence for a Mediterranean origin of the Macaronesian endemic genus *Argyranthemum* (Asteraceae). *American Journal of Botany*, **84**, 1595–1613.

Godwin ID, Aitken EAB, Smith LW (1997) Application of inter simple sequence repeat (ISSR) markers to plant genetics. *Electrophoresis*, **18**, 1524–1528.

Hongtrakul V, Huestis GM, Knapp S (1997) Amplified fragment length polymorphisms as a tool for DNA fingerprinting sunflower germplasm: genetic diversity among oilseed inbred lines. *Theoretical and Applied Genetics*, **95**, 400–407.

Jackson RB (1999) The importance of root distributions for hydrology, biogeochemistry, and ecosystem functioning. In: *Integrating Hydrology, Ecosystem Dynamics, and Biogeochemistry in Complex Landscapes* (eds Tenhunen J, Kabat P), pp. 219–240. John Wiley and Sons, Chichester, UK.

Jackson RB, Moore LA, Hoffmann WA, Pockman WT, Linder CR (1999) Ecosystem rooting depth determined with caves and DNA. *Proceedings of the National Academy of Sciences of the USA*, **96**, 11387–11392.

Krauss SL (1999) Complete exclusion of nonsires in an analysis of paternity in a natural plant population using amplified fragment length polymorphism (AFLP). *Molecular Ecology*, **8**, 217–226.

Langenheim JH (1994) Higher plant terpenoids: a phytocentric overview of their ecological roles. *Journal of Chemical Ecology*, **20**, 1223–1280.

Lerdau M, Matson P, Fall R, Monson R (1995) Ecological controls over monoterpene emissions from Douglas-Fir (*Pseudotsuga menziesii*). *Ecology*, **76**, 2640–2647.

Llusià J, Peñuelas J (2000) Seasonal patterns of terpene content and emission from seven Mediterranean woody species in field conditions. *American Journal of Botany*, **87**, 133–140.

Moritz C, Hillis DM (1996) Molecular systematics. Context and controversies. In: *Molecular Systematics* (eds Hillis DM, Moritz C, Mable BK), 2nd edn., pp. 1–16. Sinauer Associates, Sunderland, MA.

Nepstad DC, de Carvalho CR, Davidson EA *et al.* (1994) The role of deep roots in the hydrological and carbon cycles of Amazonian forests and pastures. *Nature*, **372**, 666–669.

Palacios C, Kresovich S, Gonzalez-Candelas F (1999) A population genetic study of the endangered plant species *Limonium dufourii* (Plumbaginaceae) based on amplified fragment length polymorphism (AFLP). *Molecular Ecology*, **8**, 645–657.

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

Pejic I, Ajmone-Marsan P, Morgante M *et al.* (1998) Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs and AFLPs. *Theoretical and Applied Genetics*, **97**, 1248–1255.

Prevost A, Wilkinson MJ (1999) A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. *Theoretical and Applied Genetics*, **98**, 107–112.

Rieseberg LH, Allan GL, Bui Y *et al.* (1994) Genetic fingerprinting of various native California cultivars. *Madroño*, **41**, 30–38.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning*, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Schenk HJ, Callaway RM, Mahall BE (1999) Spatial root segregation: are plants territorial? *Advances in Ecological Research*, **28**, 145–180.

Schilling EE, Linder CR, Noyes R, Rieseberg LH (1998) Phylogenetic relationships in *Helianthus* (Asteraceae) based on nuclear ribosomal DNA internal transcribed spacer region sequence data. *Systematic Botany*, **23**, 177–187.

Soltis PS, Kuzoff RK (1993) ITS sequence variation within and among populations of *Lomatium grayi* and *L. laevigatum* (Umbelliferae). *Molecular Phylogenetics and Evolution*, **2**, 166–170.

Steenkamp J, Wiid I, Lourens A, Van Helden P (1994) Improved method for DNA extraction from *Vitis vinifera*. *American Journal of Enology and Vitaculture*, **45**, 102–106.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–4680.

Van Valen L (1976) Ecological species, multispecies, and oaks. *Taxon*, **25**, 233–239.

Vos P, Hogers R, Bleeker M *et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.

Waugh R, Bonar N, Baird E *et al.* (1997) Homology of AFLP products in three mapping populations of barley. *Molecular and General Genetics*, **255**, 311–321.

White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: *PCR Protocols: A Guide to Methods and Applications* (eds Innis MA, Gelfand DH, Sninsky JJ, White TJ), pp. 315–324. Academic Press, San Diego, CA.

Whittemore AT, Schaal B (1991) Interspecific gene flow in sympatric oaks. *Proceedings of the National Academy of Sciences of the USA*, **88**, 2540–2544.

Zhang J, Madden TL (1997) POWERBLAST: a new network BLAST application for interactive or automated sequence analysis and annotation. *Genome Research*, **7**, 649–656.

Zhu JH, Stephenson P, Laurie DA, Li W, Tang D, Gale MD (1999) Towards rice genome scanning by map-based AFLP fingerprinting. *Molecular and General Genetics*, **261**, 184–195.

Zietkiewicz E, Rafalski A, Labuda D (1994) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics*, **20**, 176–183.

Zimmer EA, Martin SL, Beverly SM, Kan YW, Wilson AC (1980) Rapid duplications and loss of genes coding for α chains of hemoglobin. *Proceedings of the National Academy of Sciences of the USA*, **77**, 2158–2162.

The work reported here was a collaboration between ecologists and evolutionary biologists to develop molecular approaches to plant identification that can be used broadly by researchers studying below-ground plant patterns and processes. R. Linder studies adaption from the level of molecular genetics to populations and is interested in the relationship between micro- and macroevolution. R. Jackson studies physiological controls on ecosystem functioning and feedbacks between global change and the biosphere, including a project to identify the rooting depth and resource uptake of different plant species and ecosystem types. L. Moore studies the contributions of whole root systems to plant performance and ecosystem function.
