

Nuclear DNA-based markers for plant evolutionary biology

A. E. STRAND, J LEEBENS-MACK* and B. G. MILLIGAN

*Department of Biology, New Mexico State University, Las Cruces, NM 88003, USA, *Department of Biology, Vanderbilt University, Nashville, TN 37235, USA*

Abstract

While DNA-based markers can provide a wealth of information for the study of plant evolutionary biology, progress is limited by the lack of primers available for PCR. To overcome this limitation, we outline a protocol for developing oligonucleotide primers targeting regions of low copy-number nuclear genes. This protocol is intended to lead to universally useful primer sets. To test our approach, we designed eight primer sets and tested their abilities to amplify targets from representatives of each dicot and one monocot subclass. Five of the eight primer sets amplified targets from at least five of the seven taxa and thus exhibited broad taxonomic usefulness; the remaining primers were rather specific, however, and amplified targets from at most three taxa. In only one primer-taxon combination was a complex multiple-banded amplification produced. Overall, the protocol outlined proved quite useful at identifying broadly applicable primers targeted to low copy-number nuclear genes. Wider application of this approach should be effective at greatly increasing the amount of genetic information available for a diversity of plant nuclear genomes.

Keywords: PCR, primers, single copy DNA, genealogies

Received 2 April 1996; revision accepted 15 July 1996

Introduction

One of the most important recent advances in evolutionary biology is the development of genealogical approaches to quantify a diversity of evolutionary processes. The development of suitable molecular markers has not kept up with theory, however. This is mainly because of the requirement for information on both allele frequencies and historical relationships among alleles. Randomly amplified polymorphic DNA (RAPD) markers and microsatellites, while clearly useful in many contexts, require substantial additional time and effort to provide genealogical information. Markers currently in use that easily yield genealogical information include the organelle genomes and the noncoding regions of the nuclear ribosomal repeats. These markers, however, may not evolve quickly enough for use in population-level analyses (Wolfe 1987; but see Strand *et al.* 1996). Recent progress in marker development has been based on the detection of nucleotide sequence variation within introns of a few

specific protein-coding genes (Lessa 1992; Slade *et al.* 1993; Palumbi & Baker 1994; Taylor *et al.* 1995). Use of any limited set of markers, however, can lead to sampling artefacts in which the inferred gene genealogies differ from the branching history of the organisms from which the genes were sampled. This problem is best overcome by increasing the number of independently evolving genes studied (Pamilo & Nei 1988; Doyle 1992), something that can only be accomplished by developing markers for a large array of nuclear genes. To date, the difficulty of doing so has severely limited empirical progress based on introducing genealogical analyses into evolutionary biology, despite the increasing recognition of their utility (Avice 1989; Slatkin & Maddison 1989; Hudson 1990; Milligan *et al.* 1994; Milligan 1996).

Our aim is to implement a general protocol that can be used to develop new primer sets targeted at amplifying low copy-number nuclear markers from a taxonomically diverse group of organisms. The strategy presented is based on information contained in the publicly accessible DNA sequence databases; as the databases accumulate information, the method increases in effectiveness and generality. We focus on primer design, both because that

Correspondence: Allan E. Strand. Tel.: +1-505-646-1150; Fax: +1-505-646-5665. E-mail: astrand@evolution.nmsu.edu

represents the rate-limiting step in obtaining PCR-amplified DNA to characterize genetically and because methods for screening and genotyping the resulting products are well-covered elsewhere (Lessa & Applebaum 1993; Myers *et al.* 1987).

Materials and methods

Primer design protocol

While there are many approaches to primer design, we have sought a strategy that uses readily available information, is likely to yield successful amplification in a taxonomically diverse array of plants, generates relatively simple PCR amplification products, and may identify regions of nucleotide site variation even within single populations. The following steps outline a strategy that meets all these criteria in our experience.

Identification of potential target regions. Genes of particular interest to molecular biologists because of the roles they play in primary and secondary compound biosynthesis are excellent targets for three reasons. First, there typically exists a great deal of DNA sequence information for these genes that can be used to design primers. Second, there exists the potential to link evolutionary studies of these genes to molecular studies of the function, regulation, and expression of the same genes, thereby providing much greater insight than possible in the absence of such information. Finally, copy number for biosynthetic genes is generally low. For example, isozyme surveys of a wide range of plant taxa report low copy numbers for most isozyme-encoding genes (Gottlieb 1982; Wendel & Weeden 1989).

Potential target regions that include one or more introns within a gene are likely to be most useful. Intron sequences typically evolve at a much higher rate than exon sequences (Li & Graur 1991) and are therefore more likely to contain variable nucleotide sites suitable for use as markers. Second, in those cases in which amplification targets multiple loci, intron lengths often differ among them (Reif *et al.* 1985). As a result, products corresponding to individual loci can be separated and analysed alone.

Availability of information on gene structure from genomic DNA sequences is therefore important in the primer design process. Unfortunately, a great many genes are represented in the public databases only as sequences derived from cDNAs and therefore contain no information on gene structure. Intron position, however, generally evolves much more slowly than primary DNA sequences (Kersanach *et al.* 1994), so for the purposes of primer design, intron positions within cDNA sequences can be inferred based upon their positions in the few available genomic sequences. Finally, it is advantageous to identify

target regions that span two introns rather than one. DNA sequence data from the intervening exon can provide a gene-specific tag useful both in verifying the identity of the amplified product and in providing sites for internal sequencing primers.

Designing primer pairs. Identification of potential targets focuses attention on specific genomic regions. However, selection of oligonucleotide primer sequences that amplify the desired region in a diverse array of genetically uncharacterized plant species remains a challenging problem. Several criteria seem important to meet simultaneously. Most importantly, evolutionarily conserved segments within each target region must be identified as these are the most likely sites for successful primer annealing across a diversity of taxa. The kinetic characteristics of potential primers in the polymerase chain reaction are also important considerations. Such undesirable features as the possibility of self-annealing, primer-dimer formation, and incompatible melting temperatures can be checked with computer programs such as PRIMER version 0.5 (Lincoln *et al.* 1991). In some cases, primer pairs with suboptimal melting temperatures may be necessary due to the paucity of conserved regions in the sequences. Finally, the length of potential primer sites, their G–C content (especially at the 3' end), and the position of the 3' end relative to the reading frame are useful to consider. By placing the critical 3' end at a second codon position nucleotide, the possibility of a mismatch preventing annealing may be reduced. In some cases, the only available conserved sites may be significantly shorter than our ideal goal of 20 nucleotides; in those cases either an inosine may be included at variable internal sites or primers degenerate at those sites may be constructed.

Although all criteria were considered, in most cases the available conserved regions dictated the primer sites given the degree of sequence variation evident among the sequences examined. Thus, the procedure for identifying conserved regions plays a central role in the practical development of primers. For each target region, all available sequences can be extracted from GENBANK, EMBL, or other sequence databases. The XYLEM package of utilities (Fristensky 1993) which automates the parsing of GENBANK feature tables is useful for extracting the coding sequences from each accession, which may then be aligned by visual inspection, by using CLUSTALW (Thompson *et al.* 1994), or by using other alignment software. Regions of completely or nearly conserved sequences, readily identified from these alignments, are potential primer sites to be evaluated according to our other criteria. Increasingly, sequences derived from multiple copies within taxa are available. Because the evolutionary history of these gene copies is generally unknown, it seems reasonable to ignore the issue of

orthology vs. paralogy and to consider each sequence of equal importance when identifying primer sites from the alignments. Future refinements of this procedure will need to consider this issue further.

Example primer sets

We implemented the protocol above and designed the eight primer pairs described in Table 1. All of these primers target genes of known biochemical importance and putative low copy number. In addition, calmodulin was reported to contain highly conserved coding sequences (Marshak *et al.* 1984). Although chalcone synthase is known in some cases to exist as a multigene family, at least in *Petunia*, each locus has a distinct intron length and can therefore be distinguished from other loci (Reif *et al.* 1985).

General utility of oligonucleotides

Oligonucleotides were screened in polymerase chain reactions against a wide taxonomic range representing seven subclasses of angiosperms (*sensu* Cronquist 1981). The taxa chosen (subclasses in parentheses) were *Aquilegia longissima* (Magnoliidae), *Baptisia leucophaea* (Rosidae), *Paulownia tomentosa* (Asteridae), *Quercus gambelii* (Hamamelidae), *Raphanus sativa* (Dilleniidae), *Silene alba* (Caryophyllidae) and *Zea mays* (Commelinidae). This taxonomic breadth was intended to assess the general utility of these oligonucleotides as primers for the study of a diversity of genetically uncharacterized plant taxa.

Each of the primer pairs listed in Table 1 were included in PCR reactions with each study taxon. In addition, primers that anneal to the internal transcribed spacer region of nuclear ribosomal DNA (White *et al.* 1990) were used as positive controls. Reactions were composed of Tris

(pH 9.0) at a concentration of 10 mM, KCl at 50 mM, MgCl₂ at 1.5 mM, Triton X-100 at 0.1% (v/v), NP-40 at 0.05% (v/v), Promega-B *Taq* polymerase at 0.033 U/μL, each dNTP at 0.1 mM, and each primer at 100 nM. Final reaction volumes were 15 μL including 5–100 ng of genomic template DNA. All reactions were replicated using annealing temperatures of 42, 45 and 50 °C. Cycling conditions using an MJ Research thermal cycler were: one 2-min period at 95 °C followed by 35 cycles composed of a 1-min segment at 95 °C, a 90-s segment at one of the annealing temperatures, and a 2-min segment at 72 °C, followed finally by a 9-min period at 72 °C.

Results and Discussion

The number of taxa in which each primer pair successfully amplified varied greatly (Table 2), ranging from amplification in all taxa (*G3pdh*) to amplification in only a single taxon (*Aat*). However, five of the eight primer sets amplified in at least five of the seven taxa, an indication that our primer design protocol successfully identified broadly useful oligonucleotides. A priori one might expect a positive relationship between the number of sequences used in designing a primer and the number of taxa in which the primer successfully amplified. For these primers and taxa, however, there was a nonsignificant linear regression regardless of whether the independent variable was the raw number of sequences, the number of taxa, or the number of subclasses (correlations 0.41, 0.39, and 0.45, respectively) used to design a pair of primers. Although no relationship existed between the number of sequences used to design each primer pair and the number of taxa in which they amplified, in all but two cases (*Pgi* in *Baptisia* and *Chi* in *Raphanus*), primer pairs designed from members of a particular subclass amplified in members of that same subclass. These results demonstrate that sequences from

Table 1 Putative loci, primer designations and their sequences for oligonucleotides designed in this study. The number of sequences used in designing each primer is denoted by *n*

Putative locus	Primer	<i>n</i>	Sequence (5' → 3')
Alcohol dehydrogenase (<i>Adh</i>)	ADHX2F	4	TACTTITGGGAACGIAAGGTA
	ADHX4R	4	TCICCIACACTCTCIACAAT
Aspartate aminotransaminase (<i>Aat</i>)	AATX5F	3	GCTATTCAAGAGAACAG
	AATX7R	3	TCAACACCAGTAGGGTTA
Calmodulin (<i>Cam</i>)	CAMX1F	14	AGCCTNTTCGACAAGGATGG
	CAMX2R	14	AGTGANCGCATCACAGTT
Chalcone isomerase (<i>Chi</i>)	CHIX1F	4	TNNTTCCTCGGCGGGCGC
	CHIX4R	4	TCCCCGATNATGGNCTCCA
Chalcone synthase (<i>Chs</i>)	CHSX1F	12	AGGAAAAATTCAAGCGCATG
	CHSX2RN	12	TTCAGTCAAGTGCATGTAACG
Glyceraldehyde 3-phosphate dehydrogenase (<i>G3pdh</i>)	GPDX7F	3	GATAGATTTGGAATTGTTGAGG
	GPDX9R	3	AAGCAATTCAGCCTTGG
Phosphoglucose isomerase (<i>Pgi</i>)	PGIX12F	2	TCTCTICAGTAIGGCTT
	PGIX14R	2	AATGATACATTCCATCACCT
Triose phosphate isomerase (<i>Tpi</i>)	TPIX4FN	3	AAGGTCATTGCATGTGTGG
	TPIX6RN	3	CTTTACCAGTTCCAATAGCCC

Table 2 PCR product sizes amplified by each primer pair for each taxon. All numbers represent the length of a particular band in base pairs. Asterisk indicates that the locus identity has been confirmed by direct sequencing of the band and a similarity search against GENBANK. Superscripts refer to the annealing temperatures at which bands appear (a = 42 °C, b = 45 °C, c = 50 °C). The calmodulin primers amplify multiple bands in *Baptisia* at 45 and 50 °C (Fig. 1)

	Putative locus							
	<i>Adh</i>	<i>Aat</i>	<i>Cam</i>	<i>Chi</i>	<i>Chs</i>	<i>G3Pdh</i>	<i>Pgi</i>	<i>Tpi</i>
Forward Primer	ADHX2F	AATX5F	CAMX1F	CHIX1F	CHSX1F	GPDx7F	PGIX12F	TPIX4FN
Reverse Primer	ADHX4R	AATX7R	CAMX2R	CHIX4R	CHSX2RN	GPDx9R	PGIX14R	TPIX6R
<i>Aquilegia</i>	75 ^{ab} *380 ^{abc} 950 ^a		200 ^a	220 ^{ab} 550 ^{abc}	*200 ^{abc}	250 ^{ab} *1000 ^{abc}		300 ^b 500 ^b 1800 ^b
<i>Baptisia</i>	*550 ^{abc}	*540 ^{abc}	mult ^{bc}		320 ^{abc} 830 ^a	300 ^{ab} 1050 ^c		200 ^{abc} 700 ^{bc} 1400 ^{bc}
<i>Paulownia</i>	500 ^{abc}		650 ^{abc} 1300 ^{bc}	300 ^{abc} 400 ^{abc}	700 ^c 900 ^{abc}	1600 ^{bc} 200 ^{ab} 1000 ^a 1200 ^{ac}		450 ^{ab}
<i>Quercus</i>	400 ^{ab} 900 ^{abc}		800 ^{bc}			450 ^{ab} 1200 ^{abc} 1600 ^{bc}		
<i>Raphanus</i>	250 ^{abc}		550 ^{abc}		150 ^{abc}	150 ^{ab} 850 ^b	550 ^{bc}	400 ^{abc}
<i>Silene</i>			400 ^c	620 ^{abc}		150 ^c 200 ^{abc}		750 ^c
<i>Zea</i>	600 ^{ab} 1150 ^{ab} 2500 ^{ab}		450 ^{abc} 1150 ^{abc}	350 ^c 1100 ^{abc}	500 ^a 900 ^{ab}	1200 ^a 450 ^{abc} 700 ^{abc} 1300 ^{abc}	1000 ^{bc}	900 ^b 1150 ^b

other taxa within a taxon's subclass can provide enough information to design primers that amplify low copy-number nuclear DNA from that taxon. This is also encouraging for the prospect of applying our protocol to the study of hitherto unstudied taxa.

Reactions run on most primer pair-taxon combinations produced multiple bands on agarose gels. Figure 1 illustrates the banding patterns obtained when each pair of primers was used in reactions with *Baptisia* DNA. Although simply looking at banding patterns does not address the issue of the degree and type of homology among bands, detailed population-level analyses of two bands (*Adh* in *Baptisia* and the 380-bp *Adh* band in *Aquilegia*) suggest that each represents a separate locus. Because each amplification product should include an intron, the differences in size among the bands could, in general, be attributable to variation in intron length among duplicated loci.

There is both indirect and direct evidence that the bands amplified represent members of the target low-copy number gene family. In the case of isozyme loci, the number of bands amplified is less than or equal to the number of isozyme loci resolved by horizontal starch-gel electrophoresis (Weeden & Wendel 1989; Gottlieb 1982) for all but one gene (*Tpi*). In the case of four genes (*Adh* in

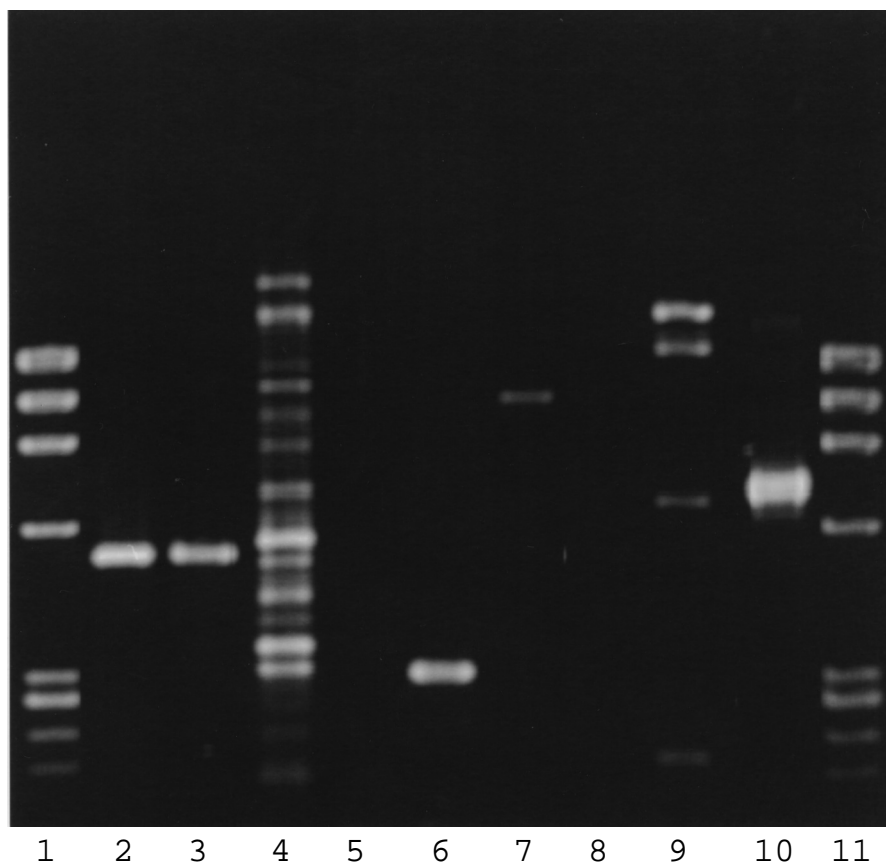
Aquilegia and *Baptisia*, *Chs* and *G3pdh* in *Aquilegia*, and *Aat* in *Baptisia*) the identities of the amplified fragments were verified by performing a similarity search of their sequences against the entire GENBANK.

Once a particular band has been sequenced, it is possible to design new primers that are specific to that band in the taxon of interest. We have successfully designed primers specific to the 380-bp *Adh*, 1000-bp *G3pdh*, and the 200-bp *Chs* bands in *Aquilegia*. Single amplification products can then easily be screened for variation by a number of methods: direct sequencing, RFLP analysis with four-base restriction endonucleases (McCauley 1994), heteroduplex analysis, single-stranded conformation polymorphism (SSCP), or denaturing gradient gel electrophoresis (DGGE) (Lessa & Applebaum 1993; Slade *et al.* 1993; Strand *et al.* 1996).

Conclusions

Nuclear DNA-based markers that contain a record of their own genealogical history are currently underutilized in evolutionary biology, primarily because of the paucity of oligonucleotides suitable for amplifying variable regions of the nuclear genome (Milligan *et al.* 1994). These types of markers have considerable potential as sources of

Fig. 1 Agarose gel showing the products resulting from amplifications of *Baptisia* DNA using each primer pair in this study. All amplification products on this gel were produced in reactions with an annealing temperature of 50°C. The contents of each lane are as follows: lane (2) *Aat*, (3) *Adh*, (4) *Cam*, (5) *Chi*, (6) *Chs*, (7) *G3pdh*, (8) *Pgi*, (9) *Tpi*, and (10) Nuclear Ribosomal DNA internal transcribed spacer region. Lanes 1 and 11 contain a size standard (Stratagene PhiX174/*Hae*III).



information on a wide array of questions in evolutionary biology, including estimation of mating systems (Milligan 1996), effective population sizes (Kuhner *et al.* 1995), migration rates (Slatkin & Maddison 1989), introgression rates (Leebens-Mack 1995), selection (Hudson *et al.* 1987), and phylogenies (both intra and interspecific). Studies of molecular systematics and evolution within and among species have become quite common, yet most investigations have focused on a small set of genes. This paper provides a suite of potentially useful DNA-based markers which can be used to address any of the questions outlined above. However, the utility of these markers will still have to be assessed on a per taxon basis. At least two issues will have to be resolved for each marker–taxon combination: the amount of variability present for each marker in the taxon of interest and the type of homology relating any variants observed. The former issue may rapidly be resolved using one of the mutation detection methods outlined above; the latter may best be resolved by observing the segregation of variants in experimental crosses. More importantly than providing a suite of new nuclear markers, this paper provides a general protocol to follow when designing primers to amplify low copy-number nuclear DNA-based markers. It is our hope that the set of available oligonucleotide primers will be greatly expanded in the near future so that the potential of genealogical analyses of

naturally occurring genetic variation will be realized. To this end we are constructing a database of these and other primer sequences, and we welcome submissions from other geneticists and evolutionary biologists.

Acknowledgements

We would like to acknowledge C. W. DePamphilis, D. Howard, D. McCauley, and M. Varagona for kindly supplying us with template DNA from the variety of taxa used in this study. In addition, T. Barkmann, M. Cain, N. Young, and two anonymous referees provided valuable feedback on this work. This research was supported in part by an NSF dissertation improvement grant, an EPA graduate fellowship, and an NMSU Biology Department fellowship awarded to A.E.S., and NSF #8906129 awarded to B.G.M.

References

- Avice JC (1989) Gene trees and organismal histories: A phylogenetic approach to population biology. *Evolution*, **43**, 1192–1208.
- Cronquist A (1981) An integrated system of classification of flowering plants. Columbia University Press, New York.
- Doyle J (1992) Gene trees and species trees: molecular systematics as one-character taxonomy. *Systematic Botany*, **17**, 144–163.
- Fristensky B (1993) Feature expressions: creating and manipulating sequence datasets. *Nucleic Acids Research*, **21**, 5997–6003.
- Gottlieb LD (1982) Conservation and duplication of isozymes in plants. *Science*, **216**, 373–380.

- Hudson R, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics*, **116**, 153–159.
- Hudson RR (1990) Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology*, **7**, 1–44.
- Kersanach R, Brinkman H, Liaud M-F, Zhang D-X, Martin W, Cerff R (1994) Five identical intron positions in ancient duplicated genes of eubacterial origin. *Nature*, **367**, 387–389.
- Kuhner MK, Yamato J, Felsenstein J (1995) Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. *Genetics*, **140**, 1421–1430.
- Leebens-Mack J (1995) *An investigation of the consequence of natural hybridization between two East Texas Baptisia (Fabaceae) species*. PhD thesis, University of Texas, Austin, Texas.
- Lessa EP (1992) Rapid surveying of DNA sequence variation in natural populations. *Molecular Biology and Evolution*, **9**, 323–330.
- Lessa EP, Applebaum G (1993) Screening techniques for detecting allelic variation in DNA sequences. *Molecular Ecology*, **2**, 119–129.
- Li W-H, Graur D (1991) *Fundamentals of Molecular Evolution*. Sinauer, Sunderland, Massachusetts.
- Lincoln S, Daly M, Lander E (1991) *PRIMER: a computer program for automatically selecting PCR primers*. Technical report, MIT Center of Genome Research and Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, MA 02142.
- Marshak DR, Clarke M, Roberts DM, Watterson DM (1984) Structural and functional properties of calmodulin from the eukaryotic organism, *Dictyostelium discoideum*. *Biochemistry*, **23**, 2891–2899.
- McCauley DE (1994) Contrasting the distribution of chloroplast DNA and allozyme poly morphism among local populations of *Silene alba*: Implications for studies of gene flow in plants. *Proceedings of the National Academy of Sciences, USA*, **91**, 8127–8131.
- Milligan BG (1996) Estimating long-term mating systems using DNA sequences. *Genetics*, **142**, 619–627.
- Milligan BG, Leebens-Mack J, Strand AE (1994) Conservation genetics: beyond the maintenance of marker diversity. *Molecular Ecology*, **3**, 423–435.
- Myers RM, Maniatis T, Lerman LS (1987) Detection and localization of single base changes by denaturing gradient gel electrophoresis. *Methods in Enzymology*, **155**, 501–527.
- Palumbi S, Baker CS (1994) Contrasting population structure from intron sequences and mtDNA. *Molecular Biology and Evolution*, **11**, 426–435.
- Pamilo P, Nei M (1988) Relationships between gene trees and species trees. *Molecular Biology and Evolution*, **5**, 568–583.
- Reif H, Niesbach U, Deumling B, Saedler H (1985) Cloning and analysis of two genes for chalcone synthase from *Petunia hybrida*. *Molecular and General Genetics*, **199**, 208–215.
- Slade RW, Moritz C, Heideman A, Hale P (1993) Rapid assessment of single copy nuclear DNA variation in diverse species. *Molecular Ecology*, **2**, 359–373.
- Slatkin M, Maddison WP (1989) A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics*, **123**, 603–613.
- Strand AE, Milligan BG, Pruitt CM (1996) Are populations islands? Analysis of chloroplast DNA variation in *Aquilegia*. *Evolution*, **50**, 1822–1829.
- Taylor MFJ, Shen Y, Kreitman ME (1995) A population genetic test of selection at the molecular level. *Science*, **270**, 1497–1499.
- Thompson J, Higgins D, Gibson T (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.
- Weeden NF, Wendel JF (1989) Genetics of plant isozymes. In: *Isozymes in Plant Biology* (eds Soltis DE, Soltis PS), Vol. 4 of *Advances in Plant Sciences*. Dioscorides Press, Portland, Oregon.
- Wendel JF, Weeden NF (1989) Visualization and interpretation of plant isozymes. *Isozymes in Plant Biology* (eds Soltis DE, Soltis PS), pp. 5–45. Vol. 4 of *Advances in Plant Sciences* Dioscorides Press, Portland, Oregon.
- White T, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal RNA genes for phylogenetics. In: *PCR Protocols. A Guide to Methods and Applications* (eds Innis, MA, Gelfand DH, Sninsky JJ, White TJ), pp. 315–322. Academic Press, San Diego, CA.
- Wolfe KH, Li W-H, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences, USA*, **84**, 9054–9058.

The authors' research centres on extracting demographic and evolutionary information from patterns of molecular variation found in samples of natural populations. In addition to developing molecular markers, we are developing genealogical analysis techniques and applying them to study mating systems, effective population size, gene flow and introgression.
