Sequence assessment of comigrating AFLP™ bands in *Echinacea* — implications for comparative biological studies

Subbaiah M. Mechanda, Bernard R. Baum, Douglas A. Johnson, and John T. Arnason

Abstract: The extent of sequence identity among clones derived from monomorphic and polymorphic AFLP™ polymorphism bands was quantified. A total of 79 fragments from a monomorphic band of 273 bp and 48 fragments from a polymorphic band of 159 bp, isolated from individuals belonging to different populations, varieties, and species of *Echinacea*, were cloned and sequenced. The monomorphic fragments exhibited above 90% sequence identity among clones within samples. Sequence identity within variety ranged from 82.78% to 94.87% and within species from 75.82% to 98.9% and was 57.97% in the genus. The polymorphic fragments exhibited much less sequence identity. In some instances, even two clones from the same fragment were different in their size and sequence. Within sample, clone sequence identity ranged from 100% to 51.57%, within variety from 33.33% to 100% in one variety, and from 23.66% to 45% within species and was as low as 1.25% within the genus. In addition, sequences of the same size were aligned to verify the nature of their sequence dissimilarity/similarity. Within each size group, identical sequences were found across species and varieties. In general, comigrating bands cannot be considered homologous. Thus, the use of AFLP™ band data for comparative studies is appropriate only if the results emanating from such analyses are considered as approximations and are interpreted as phenotypic but not genotypic.

Key words: AFLP markers, false homologies.

Résumé : Le degré d'identité de la séquence nucléotidique parmi des clones dérivés d'amplicons AFLP™ monomorphes ou polymorphes a été mesuré. Au total, 79 produits correspondant à un fragment monomorphe de 273 pb et 48 produits d'un amplicon polymorphe de 159 pb ont été séquencés. Ces produits provenaient d'individus appartenant à diverses populations, variétés ou espèces d'échinacée. Les produits de l'amplicon monomorphe étaient identiques à plus de 90 % parmi les clones d'un échantillon. Au sein d'une variété, l'identité variait entre 82,78 % et 94,78 %. Elle se situait entre 75,82 % et 98,9 % au sein des espèces et elle était de 57,97 % globalement pour le genre. Les produits de l'amplicon polymorphe montraient beaucoup moins d'identité. Dans certains cas, même deux clones provenant du même amplicon étaient de taille et de séquence distinctes. Au sein d'un échantillon, l'identité variait entre 51,57 % et 100 % ; au sein d'une variété, cela allait de 33,33 % à 100 % ; au sein de l'espèce, l'identité variait entre 23,66 % et 45 %, tandis qu'au sein du genre, elle pouvait n'être que de 1,25 %. De plus, les produits de la même taille ont été alignés pour vérifier la nature de leur similitude/dissimilitude. Au sein de chaque groupe (selon la taille), des séquences identiques ont été trouvées parmi les espèces et les variétés. De façon générale, des amplicons de même taille ne peuvent pas être considérés comme étant homologues. Ainsi, l'emploi de données AFLP pour des études de comparaison est approprié uniquement lorsque celles-ci sont acceptées comme étant des approximations et sont interprétées comme étant des données phénotypiques et non génotypiques.

Mots clés : marqueurs AFLP, fausses homologies.

[Traduit par la Rédaction]

Received 24 March 2003. Accepted 3 September 2003. Published on the NRC Research Press Web site at http://genome.nrc.ca on 13 January 2004.

Corresponding Editor: G.J. Scoles.

S.M. Mechanda. Agriculture and Agri-Food Canada, 960 Carling Avenue, Ottawa, ON K1A 0C6, Canada, and Ottawa–Carleton Institute of Biology, University of Ottawa, 30 Marie Curie, Ottawa, ON K1A 0C7, Canada.
B.R. Baum.[1] Agriculture and Agri-Food Canada, 960 Carling Avenue, Ottawa, ON K1A 0C6, Canada.
D.A. Johnson and J.T. Arnason. Ottawa–Carleton Institute of Biology, University of Ottawa, 30 Marie Curie, Ottawa, ON K1A 0C7, Canada.

[1]Corresponding author (e-mail: baumbr@agr.gc.ca).

 doi: 10.1139/G03-094

Introduction

Amplified restriction fragment polymorphism (AFLP™) has become a popular technique for rapidly generating molecular markers (Vos et al. 1995). The AFLP method involves complete digestion of the genomic DNA using restriction enzymes, followed by ligation of the resulting fragments with adapters. Subsequently, a subset of the restriction fragments are amplified by PCR using selective AFLP primers and the resulting amplimers are then resolved by denaturing polyacrylamide gel electrophoresis (PAGE) (Vos et al. 1995; reviewed in Mueller and Wolfenbarger 1999; Savelkoul et al. 1999). Depending on the purpose, the AFLP fragments may be detected either by radiolabelling the fragments (Vos et al. 1995) or by silver staining the gel (Chalhoub et al. 1997). When compared with other technologies such as restriction fragment length polymorphism (RFLP), AFLP has proven to be a reproducible and reliable methodology with a high multiplex ratio (Powell et al. 1996). The AFLP marker system has been used in a broad range of applications such as genotyping including pathotyping and fingerprinting (Mueller and Wolfenbarger 1999), medical diagnostics, chromosomal mapping, and generating data for systematics, population genetic analysis, breeding, cultivar identification, germplasm management, and many other applications. It has been used in bacteria, fungi, plants, insects, animals, and humans (Mueller and Wolfenbarger 1999; Savelkoul et al. 1999). For 2002 alone, the National Library of Medicine (PubMed, NCBI) lists about 100 titles for population studies and 20 titles for phylogenetic studies.

Several problems associated with the use of AFLP data have been discussed by Robinson and Harris (1999). While a valuable tool for many applications, the use of AFLP in comparative biology studies has often made the assumption that comigrating AFLP fragments are homologous, i.e., bands migrating at the same position in the gel were derived from the same chromosomal region. But this may not always be the case, as comigrating bands may consist of sequences from different regions of the genome, and a single nucleotide mutation at a potential restriction site would result in the lack of a band at the expected size. In all cases, bias in the data used for population or evolutionary studies is introduced.

Previous research employing RFLP and random amplified polymorphic DNA (RAPD) markers had demonstrated that the same bias may be at work. Zande and Bijlsma (1995) described these limitations when RAPD data were used and reported that fragments of similar length are not identical among species belonging to different subgroups of *Drosphila*. There are indications from recent studies that these limitations are also a concern when AFLP data are used for estimating relatedness. Rouppe van der Voort et al. (1997) assessed allele specificity of AFLP markers in five remotely related potato genotypes by cloning and sequencing 20 markers from two genotypes, of which 19 were nearly identical. Wong et al. (2001) electroblotted AFLP products and probed with 11 individual markers, of which two were monomorphic and nine were polymorphic. Badr et al. (2002) made the claim that comigrating bands are identical in terms of sequence in closely related species of *Hordeum*. Ipek and Simon (2003) sequenced 79 AFLP fragments that were used in diversity analysis in garlic and found 95% of the comigrating fragments to be identical.

While carrying out an analysis of the biodiversity of *Echinacea* that used AFLP data for population genetics, numerical taxonomic, and phylogenetic analyses (Mechanda et al.[2]), we became concerned about the validity of this assumption. We were especially concerned, as *Echinacea* is an open-pollinated species with a few reported triploid ($2n$ = 33) introgressants between diploid ($2n$ = 22) and tetraploid ($2n$ = 44) plants. In this study, we set out to test the assumption that size similarity of comigrating AFLP bands indicates sequence identity. By expanding the analysis to a relatively large collection of sequences derived from monomorphic and polymorphic bands, at four taxonomic levels (genus, species, variety, and population), these results should give a clearer indication of the sequence similarity among identical bands.

Materials and methods

Plant material

Leaves from single plants, roots, and seeds of 58 natural populations of *Echinacea* from Canada and mainly from the midwestern and southeastern United States were collected. In the related investigation, 435 single plant samples of the 58 populations were used (Mechanda et al.[2], their Table 1). For AFLP analysis, single leaves from each population were collected on ice and frozen in liquid nitrogen and stored at –80 °C.

DNA isolation

The cetyltrimethylammonium bromide method (Doyle and Doyle 1987) and the Qiagen Kit for DNA isolation (Qiagen, Mississauga, Ont.) were used to extract DNA from single leaves. DNA concentration was measured using a Hoefer DynaQuant™ minifluorometer (Hoefer Pharmacia Biotech, Calif.) and diluted to 50 ng/μL using TE (10 mM Tris, 1 mM EDTA, pH 8.0) or sterile water.

AFLP procedure and data scoring

AFLP was performed according to Vos et al. (1995) using the AFLP® Analysis System-I (Invitrogen Life Technologies, Calif.) with some minor modifications as described in Baum et al. (2001). Briefly, 500 ng of DNA was digested with *Eco*RI and *Mse*I and then adapters specific to the restriction sites were ligated overnight at room temperature. The fragments with adapters were amplified with AFLP primers having one selective nucleotide in the preamplification reaction with 20 cycles of 94 °C for 30 s, 56 °C for 60 s, and 72 °C for 60 s using the DNA Engine Thermocycler PTC 200 (MJ Research, Waltham, Mass.). The preamplification reactions were diluted 1:50 using TE. Following preliminary tests with 10 primer sets, all DNA accessions were evaluated with three primers sets, each with

[2] S.M. Mechanda, B.R. Baum, D.A. Johnson, and J.T. Arnason. Analysis of diverstiy of natural populaterion and commerical lines of *Echinacea*, using AFLP. Submitted.

three selective nucleotides (shown in bold), based on amplification of an optimum number of bands per gel (80–100) and the maximum number of polymorphisms detected per primer set: set 1, D10Eco-GACTGCGTACCAATTC**ACC**/Mse-GATGAGTCCTGAGTAA**CCG**; set 2, G10Eco-GACTGCGTACCAATTC**AGC**/Mse-GATGAGTCCTGAGTAA**CCG**; set 3, D12Eco-GACTGCGTACCAATTC**ACC**/Mse-GATGAGTCCTGAGTAA**CGC**).

The *Eco*RI primer was end labelled with ^{33}P using T4 polynucleotide kinase (Invitrogen Life Technologies) and selective amplification was carried out with three selective nucleotides on both the *Eco*RI and *Mse*I primers with one cycle at 94 °C for 30 s, 65 °C for 30 s, and 72 °C for 60 s. For the next 12 cycles, the annealing temperature was lowered 0.7 °C from 65 °C to 56 °C followed by 23 cycles performed at 94 °C for 30 s, 56 °C for 30 s, and 72 °C for 60 s. The PCR products were electrophoresed on 6% polyacrylamide gel with the AFLP band size marker 30- to 330-bp AFLP™ DNA ladder (Invitrogen Life Technologies) or a 100-bp DNA ladder (Promega, Wis.) in one lane and a representative sample in another lane (see Fig. 1). The gels were dried and autoradiographed using Kodak diagnostic X-OMAT film (Eastman Kodak, Rochester, N.Y.) at room temperature for 2–3 days. The representative samples were used to align different gels for band orientation, comparison, and scoring.

Cloning and sequencing

The AFLP reactions were repeated for selected samples (two to six for each species and variety) without the ^{33}P-labelled primer. Instead, the gels were electrophoresed and stained using the Silver Sequence DNA sequencing system (Promega). A well-resolved monomorphic band and a polymorphic band at the same position across the lanes were selected at random, excised from the gels, eluted by the crush and soak method (Sambrook et al. 1989), and then reamplified for 23 cycles using the same primer pair and conditions used to generate the band. The purified 127 bands were cloned using the TOPO TA cloning® kit for Sequencing (Invitrogen). Plasmid DNA was isolated using the QIAprep Miniprep (Qiagen) plasmid kit and the purified DNA was resuspended in water. The inserts from two to five plasmids per sample were sequenced in both directions using the ABI PRISM dye terminator cycle kit and run on an ABI PRISM 377 DNA sequencer (Perkin Elmer-Applied Biosystems, Calif.).

Sequence alignment

To determine if the comigrating bands are identical across species and varieties, the sequences from both the monomorphic band and those of the polymorphic band across the lanes were aligned separately with CLUSTAL W (Thompson et al. 1994), displayed with GENEDOC (Nicholas and Nicholas 1997), and their degree of similarities assessed. The sequences were summarized separately for the clones within samples, for the clones within varieties, for the clones within species, and for all the clones within the genus, each time at the threshold similarity of 100%, and the percent identical sequences were calculated.

Results

A well-resolved monomorphic band (273 bp) and a well-resolved polymorphic band (159 bp) were excised and cloned from the samples of each species/variety (Fig. 1). A total of 79 clones obtained from the monomorphic bands were sequenced and aligned. For the polymorphic band, 48 clones were sequenced and aligned.

Monomorphic bands

Initially, at least three to five clones from three to five samples from a taxon were sequenced. Preliminary sequence alignments for several taxa (*Echinacea atrorubens* var. *atrorubens*, *Echinacea atrorubens* var *neglecta*, *Echinacea pallida* var. *angustifolia*, and *Echinacea purpurea*) revealed that very high identity within each taxon existed, and therefore, fewer clones were sequenced for the other taxa (Table 1). In all cases, the expected band size was conserved among all samples and consistently, the sequence identity was above 90%, but never 100%, among clones within samples. For example, the analysis of sequences from the monomorphic band in three populations of *Echinacea atrorubens* var. *paradoxa* (Fig. 1) demonstrates the high similarity but not complete sequence identity (Fig. 2). Note that in this particular example, one population, EPP-1B, accounts for more variation than the other two. Sequence identity dropped slightly within varieties and dropped further within species in *E. pallida* (75.82% identity) (Table 1). All of the sequences for the genus *Echinacea* when aligned together were only 58.97% identical. Note that the percent dissimilarity increases through the four taxonomic levels from population to genus.

Polymorphic bands

Two clones from each of the five *E. pallida* var. *angustifolia* samples were sequenced, and while duplicates had similarities of 99–100%, identity among the five samples was only 54.03%. In contrast, *Echinacea pallida* var. *simulata* and *Echinacea pallida* var. *tennesseensis* exhibited much lower similarities (Table 1). The alignment presented in Fig. 3 demonstrates the lack of sequence identity among polymorphic bands migrating to the same position (159 bp) and taken from the same gel of *E. atrorubens* var. *paradoxa* shown in Fig. 1. In contrast with the results above for the monomorphic bands where conservation of size was observed, length differences were observed among different clones selected from the same polymorphic AFLP band, e.g., EPP-2B, EPP-2A, and EPP-2C differ widely in size, even though they have the same adapter/primer sequences at both ends. In general, the recovery of sequences with the expected length of 159 bp was very inconsistent, and sequences ranging from 74 to 161 bp were recovered (Table 1). For *E. pallida* var. *angustifolia*, *E. pallida* var. *pallida*, *E. pallida* var. *simulata*, and *E. purpurea*, the expected fragment length was found among the clones with only minor variation, 158 and 161 bp, and then in only two clones (Table 1). In the other taxa, clones of the expected sequence length were not found at all in our sample, resulting in sequences ranging from 74 to 161 bp (Table 1). For example, in *E. pallida* var. *tennesseensis*, none of the 10 sequences had the expected size of 159 bp. Four of the 10

were identical in length (125 bp) and in sequence. Five clones of equal length (108 bp) could be assigned to three groups containing two, two, and one clone, respectively (Fig. 4). The last of the 10 clones, ET52-2 from the same sample as ET52-1, was different in length (91 bp) and very different in nucleotide sequence (Fig. 4).

To attempt to determine whether the observed heterogeneity was related to sequence, we performed alignments across species for each of the fragment size classes. The alignments of 15 sequences that were 159 bp revealed four different groups including (*i*) two of *Echinacea pallida* var. *sanguinea*, (*ii*) three of *E. purpurea*, (*iii*) nine sequences, of which one was from *Echinacea laevigata*, four from *E. pallida* var. *angustifolia*, two *E. atrorubens* var. *atrorubens*, and one each from *E. purpurea* and *E. pallida* var. *sanguinea*, and (*iv*) a single sequence of *E. pallida* var. *sanguinea* (Fig. 5). In the same alignment at position 97–100, an *Mse*I restriction site (TTAA) was detected in two sequences from *E. pallida* var. *sanguinea* (ES5-3, ES5-5) (see Discussion). The alignment of six sequences that were 108 bp revealed two different kinds: (*i*) three sequences of *E. pallida* var. *tennesseensis* and one sequence of *E. laevigata* and (*ii*) two sequences in *E. pallida* var. *tennesseensis* (not shown). All of the 91-bp fragments were identical (not shown) but stemmed from the following taxa: *E. atrorubens* var. *atrorubens* (four clones), *E. atrorubens* var. *negelcta* (one clone), *E. atrorubens* var. *paradoxa* (three clones), *E. laevigata* (one clone), and *E. pallida* var. *tennesseensis* (one clone). These results demonstrate that sequences of comigrating polymorphic bands from different species and their varieties vary widely in sequence length and similarity.

Within variety, sequence identity ranged from 23% identity to 64% (discounting *E. atrorubens* var. *atrorubens* where too few comparable clones were investigated). Within species, sequence identity ranged from 23% to 46%, and within the genus, the identity dropped to a very low 1.25% (Table 1).

Discussion

False homologies

The essence of comparative biology, especially at the genetic and taxonomic levels, depends on on homology; however, homology is a relative concept and testing for state of homology is required (Patterson 1988). Because a state of homology can be inferred from so many criteria, it becomes important to specify the precise context or criteria used when comparisons are assessed. Thus, AFLP comigrating bands may appear to be homologous based on the criterion of mobility on a PAGE gel when, with respect to sequence identity, they may or may not be homologous. Two AFLP bands of identical mobility and sequence may have originated from different parts of the genome, as might be the case among sequences in *E. atrorubens* var. *paradoxa* (Table 1, sample 2; Fig. 3) and *E. pallida* var. *tennesseensis* (Table 1, samples 3 and 5; Fig. 4). The noncoding nature of AFLPs may help to provide an explanation for the variability and multiple length variants that are present at many loci, which are rarely strictly dominant (Wong et al. 2001). Other factors such as the possibility of replication errors during PCR of the silver-stained fragments as well as presence of multiple alleles may contribute to the variation within samples.

In both RAPDs and AFLPs, the resulting fragments are random due to the distribution of the primer annealing sites in the former and restriction sites in the latter. In case of RAPDs, Zande and Bijlsma (1995) reported that fragments of similar length are not necessarily identical between species belonging to different subgroups of *Drosphila*. Several AFLP studies have indicated fragment heterogeneity. In barley, inserts from colonies of a single transformation event were typically not identical (Shan et al. 1999). In soybean, direct conversion of the AFLP fragment to a sequence tagged site marker was not feasible due to the small band size and heterogeneity of the bands (Meksem et al. 2001). Rouppe van der Voort et al. (1997) sequenced 20 out of 117 putatively homologous AFLP bands in a mapping study of potato and reported that 95% were identical. In the present study, 79 out of 435 putatively homologous monomorphic bands and 48 out of 159 putatively homologous polymorphic bands among the 435 positions on the gels were sequenced. Identities ranging from 75% to 98% among the four species (Table 1) were found for the monomorphic bands at a given position on the gel, but the sequence identity of the polymorphic bands at a given position was considerably lower. Therefore, the claim that comigrating bands are identical in terms of sequence in closely related species of *Hordeum* (Badr et al. 2002) is not valid for *Echinacea*, possibly because of the extensive outbreeding nature of the genus (also see Robinson and Harris (1999) for a general discussion). This hypothesis can be tested in barley, as some species are known to be outbreeding. Therefore, because of incomplete homologies with respect to sequence identity, AFLP data might be treated as phenotypic as opposed to genotypic data for the assessment of biodiversity, population genetics, and related studies and analyzed accordingly as we did in our population genetic study of *Echinacea* (Mechanda et al.[2]).

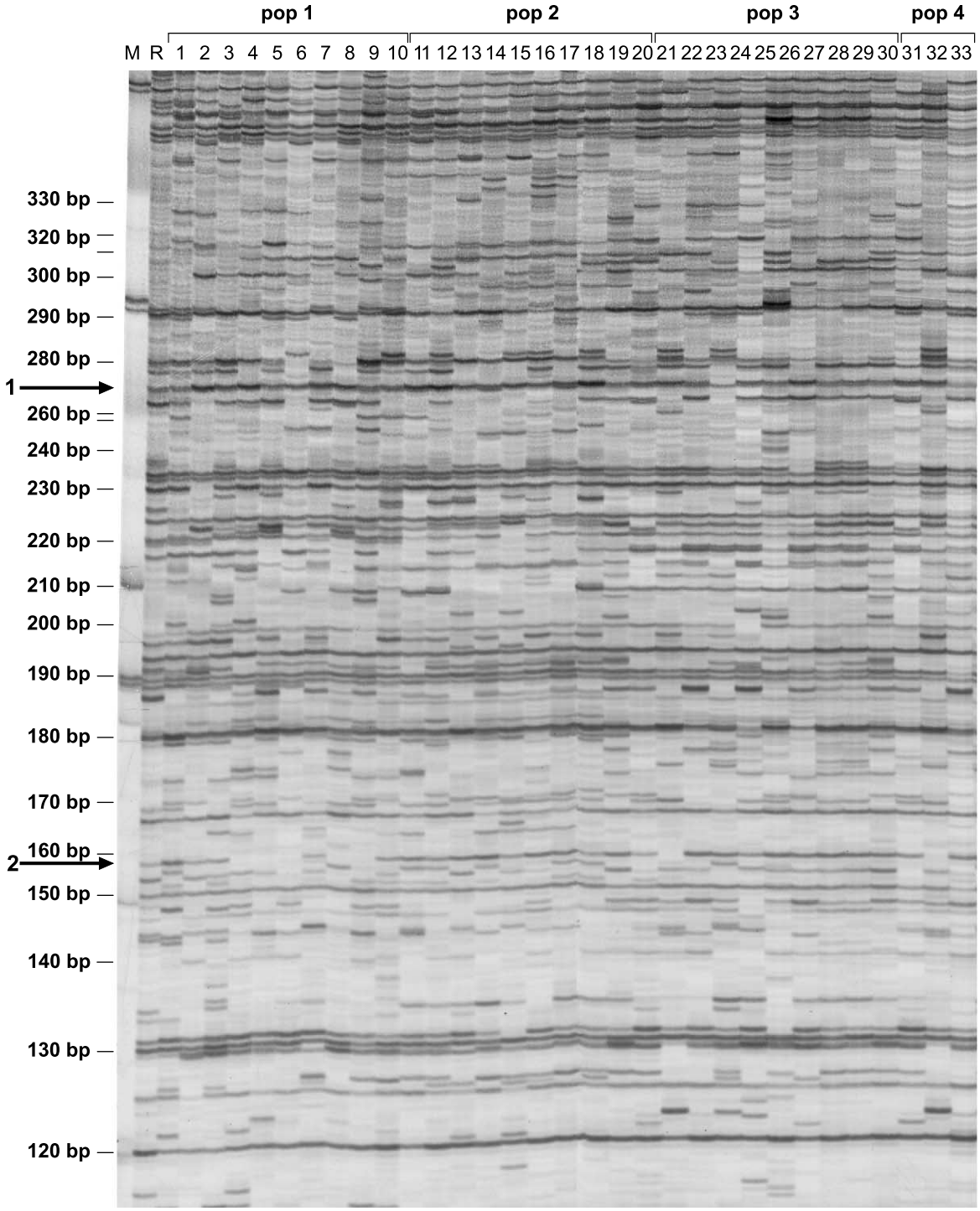
Monomorphic bands

A band of 273 bp, which was the size of the excised band, can normally be easily resolved from bands of similar sizes on a polyacrylamide gel. Furthermore, on the silver-stained gel used to isolate the desired fragment, it is much easier to manually resolve bands with similar mobility with better sensitivity than radioactively labelled fragments (Chalhoub et al. 1997). The highest level of identical sequences was obviously within samples, lower within varieties, much lower within species, and considerably lower within the genus (Table 1). The high degree of identity between cloned sequences suggests that the observed sequence differences did not arise from the accidental copurification of bands with different origins.

Polymorphic bands

We chose the 159-bp bands across the gel for comparison because at that size, they can easily be distinguished by one nucleotide difference on a PAGE gel. The level of putative homology, i.e., identity of sequences, was much lower than among the monomorphic bands (Table 1). Furthermore, we found a lack of identity in both length and sequence among

Fig. 1. AFLP silver-stained gel containing DNA samples from four populations of *E. atrorubens* var. *paradoxa* amplified with No. 2 primer set combination (*Eco*RI + AGC and *Mse*I + CCG). Arrow 1 points to the monomorphic band and arrow 2 points to the polymorphic band that was excised, cloned, and sequenced. Lane M, molecular weight marker; lane R, reference individual used for between-gel comparisons.



different clones taken from the same fragment, such as in *E. atrorubens* var. *paradoxa* EPP-2B and EPP-2C (Fig. 3) and *E. pallida* var. *tennesseensis* clones ET2-1 and ET2-2 (Fig. 4). However, when fragments of similar size were separately aligned, there was complete identity among the fragments of 91 bp. In the other fragments, 108 and 159 bp, the sequences could be grouped into different kinds, each with sequences of complete identity across species and varieties.

What is the origin of this variation? Since the fragments likely originate from different parts of the genome, conceivably from repetitive DNA, an extreme G + C content may influence gel mobility. Examination of these examples

Table 1. Similarities between a monomorphic band and a polymorphic AFLP comigrating band in *Echinacea*.

Variety/species/genus	Sample No.	No. of sequenced clones	Actual band size (bp)	% similarity of sequences within sample	% similarity of sequences within variety	% similarity of sequences within species	% similarity of sequences within genus
Monomorphic band, size expected 273 bp							
E. atrorubens var. *atrorubens*	1	4	273	98.54			
	2	3	273	95.60			
	3	7	273	94.13			
Total	3	14			91.24		
Mean ± SD				96.09±2.25			
E. atrorubens var. *neglecta*	1	5	273	93.40			
	2	5	273	93.77			
	3	3	273	95.23			
Total	3	13			89.74		
Mean ± SD				94.13±0.97			
E. atrorubens var. *paradoxa*	1	4	273	94.13			
	2	4	273	98.53			
	3	4	273	96.33			
Total	3	12			91.57		
Mean ± SD				96.33±2.2			
E. atrorubens	9	39	273			80.65	
Mean ± SD					90.85±0.98		
E. laevigata	1	1	273				
	2	1	273				
Total	2	2				98.90	
E. pallida var. *angustifolia*	1	3	273	96.33			
	2	3	273	91.20			
	3	3	273	90.47			
	4	3	273	92.30			
	5	3	273	91.94			
	6	2	273	97.06			
Total	6	17			82.78		
Mean ± SD				93.22±2.78			
E. pallida var. *pallida*	1	1	273				
	2	1	273				
Total	2	2			92.30		
E. pallida var. *simulata*	1	1	273				
	2	1	273				
Total	2	2			88.64		
E. pallida var. *tennesseensis*	1	1	273				
	2	1	273				
Total	2	2			94.87		
E. pallida	12	23				75.82	
Mean ± SD					89.65±0.98		
E. purpurea	1	3	273	90.11			
	2	3	273	91.57			
	3	3	273	91.94			
	4	3	273	94.13			
	5	3	273	97.06			
Total	5	15				79.48	
Mean ± SD					92.96±2.71		
Echinacea	28	79	273				58.97
Mean ± SD						83.71±10.33	
Polymorphic band, size expected 159 bp							
E. atrorubens var. *atrorubens*	1	2	91				
	2	2	91				
Total	2	4			100.00		
E. atrorubens var. *neglecta*	1	1	91				
	2	1	74				

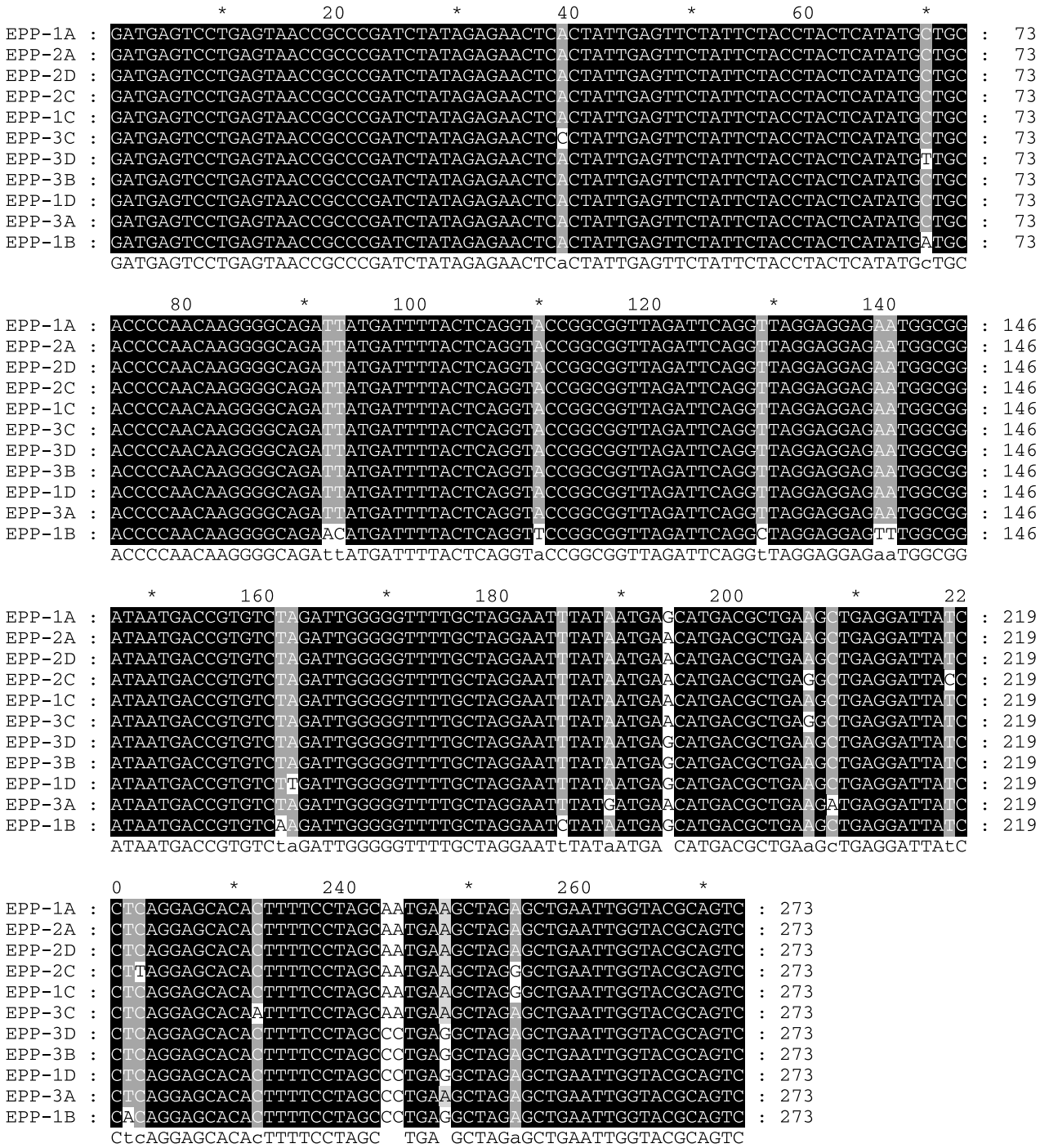
Table 1 (*concluded*).

Variety/species/genus	Sample No.	No. of sequenced clones	Actual band size (bp)	% similarity of sequences within sample	% similarity of sequences within variety	% similarity of sequences within species	% similarity of sequences within genus
Total	2	2			64.83		
E. atrorubens var. *paradoxa*	1	2	125	100.00			
	2	3	91, 158	39.87			
	3	2	91	100.00			
Total	3	7			46.40		
Mean ± SD				42.66±34.71			
E. atrorubens	7	13				37.60	
Mean ± SD					70.41±27.23		
E. laevigata	1	1	159				
	2	1	108				
	3	1	91				
Total	3	3				31.44	
E. pallida var. *angustifolia*	1	2	159	100.00			
	2	2	159	99.37			
	3	2	159	100.00			
	4	2	161	99.37			
Total	4	8			54.03		
Mean ± SD				99.62±0.36			
E. pallida var. *pallida*	1	2	159	100.00			
	2	2	91,159	100.00			
Total	2	4			33.33		
E. pallida var. *simulata*	1	2	159	62.26			
	2	2	159	51.57			
	3	1	158				
Total	*3	*5			42.13		
Mean ± SD				56.92±7.56			
E. pallida var. *tennesseensis*	1	2	125	100.00			
	2	2	125	100.00			
	3	2	108	68.51			
	4	2	108	100.00			
	5	2	91,108	57.40			
Total	5	10			36.00		
Mean ± SD				85.18±20.67			
E. pallida	14	27				23.66	
Mean ± SD					39.76±8.75		
E. purpurea	1	1	159				
	2	1	159				
	3	1	159				
	4	1	159				
	5	1	160				
Total	5	5				45.00	
Echinacea	27	48	74–159				1.25
Mean ± SD						34.43±9.07	

Note: Similarities computed from alignment of nucleotide sequences at the 100% similarity threshold.

shows no major differences in the proportion of G + C content, thus eliminating this explanation. Another possibility is that during the addition of adapters, ligation of digested fragments would lead to the amplification of a sequence that is actually two sequences, thus resulting in false homology. The detection of *Mse*I sites in two sequences from *E. pallida* var. *sanguinea* (ES5-3, ES5-5) suggests that this has occurred (Fig. 5). Physical factors affecting gel resolution may explain discrepancies among fragment length and sequence identity, e.g., reptation of DNA molecules (entangled DNA fragments diffusing through the pores of a gel under the influence of a driving electric field) may result in comigration of fragments of different sizes (Rousseau et al. 2000), or differences in mobility, diffusion, and dispersion of comigrating fragments on sequencing gels may be due to nonoptimal conditions of separation on the gel (Brahmasandra et al. 2001). Thus, a band could possibly contain a mixture of entangled DNA fragments of varying sizes as revealed by cloning and sequencing. We speculate that this phenomenon may occur at the lowest molecular weights and (or) at the lower part of

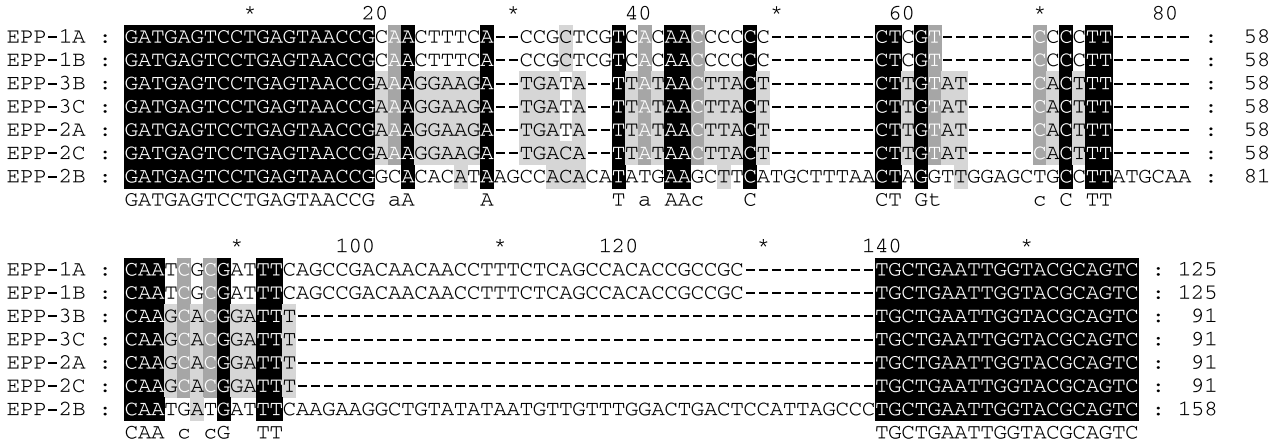
Fig. 2. Example of alignment of sequences from AFLP comigrating monomorphic fragments of 273-bp band size in *E. atrorubens* var. *paradoxa* (refer to Fig. 1, arrow 1 and text). Four clones from one population (EPP-1: EPP98-0606-2 Camden Co., Mo.), three clones from one population (EPP-2: EPP98-0605-2, Benton Co., Mo.), and four clones from EPP-3 (EPP 0823, Benton Co.). Black, 100% identity; dark grey, 80% identity; light grey, 60% identity; white, less than 50% identity.



the acrylamide gels where the resolution of the bands is normally clearer than at the upper part. This idea could be tested by repeating the analysis with a polymorphic band of higher molecular weight. Gel systems that use capillary electrophoresis may avoid some of these problems. Irrespective, the reliability of gel-based AFLPs as genotypic markers for comparative purposes remains questionable. Instead, they may be used in comparative investigations, such as taxonomy, genotyping, and fingerprinting (Mueller and LaReesa Wolfenbarger 1999), by treating them as phenotypic markers.

Although Wong et al. (2001) inferred varying codominance status among AFLP polymorphic sites, they based their findings on experiments that used cloned fragments as

Fig. 3. Example of alignment of sequences from AFLP comigrating polymorphic fragments of expected 159-bp band size in *E. atrorubens* var. *paradoxa* (refer to Fig. 1, arrow 2 and text) demonstrating sequence difference among clones. Two clones from one population (EPP-1: EPP98-0606-2, Camden Co.), three clones from one population (EPP-2: EPP98-0605-2, Benton Co.,; and two clones from EPP-3 (EPP 0823, Benton Co.). Black, 100% identity; dark grey, 80% identity; light grey, 60% identity; white, less than 50% identity.



probes of Southern blots. It can be assumed that presence of hybridization signals may also be affected by interlane mobility shifts. Southern hybridization may yield a false indication of sequence identity. We did not clone any site at which the fragments could not be detected, i.e., band absent, among the polymorphic bands. This is because it can always be assumed that a fragment may be present even though visually undetectable. Excised bands were equally unambiguous on both radioactive and silver-stained gels.

Theoretical consideration

Several criteria have been put forward for estimating nucleotide divergence. These are based on RAPD analysis (Clark and Lanigan 1993) but also apply to AFLPs: (*i*) homology of bands of the same size in different species should be demonstrated, (*ii*) true nucleotide sequence divergence should not exceed ~10%, and (*iii*) single nucleotide substitutions at the priming sites are assumed to result in a loss of amplification (criterion 8 in Clark and Lanigan 1993). In this study, we have demonstrated for the monomorphic fragments that sequence divergence did not exceed ~10% within samples and within varieties. But within species, the second criterion was not satisfied (Table 1). Monomorphic bands are not considered to be useful for comparative studies, as they occur in all samples, including all of the species in *Echinacea*. The polymorphic band investigated in this study did not meet any of Clark and Lanigan's (1993) criteria but might carry useful information when used as phenotype. It has been assumed that mapped AFLP markers from all of the regions of the genome would be more reliable for diversity analysis than unmapped ones. However, such unmapped markers, which are widely distributed over the genome, can be used for genetic diversity analysis (Virk et al. 2000).

Implications

In the special case when specific fragments that are mapped on chromosomes are used as AFLP markers for comparative studies, then their homology might be expected probably at the species level and below. A recent study of garlic (*Allium sativum*) found that 95% of comigrating AFLP fragments sharing the same position on PAGE contained homologous sequences (Ipek and Simon 2003). However, it is not practical to sequence all bands, and Ipek and Simon (2003) have sequenced only the mapped marker bands. Proof of homology cannot be achieved with confidence by Southern blotting in lieu of cloning and sequencing, as the presence of primer sequences in the AFLP fragments may result in nonspecific cross-hybridization.

Nonhomologous products might bias genetic distance estimates between taxa (Lamboy 1994) and diversity measures. Using nonhomologous and nonindependent AFLP fragment data as characters will increase homoplasy (Bremer 1991). However, for the purpose of identification, the use of AFLP banding patterns as phenotypic characters is deemed appropriate. Diagnostic characters are in some sense phenetic, and in general, morphological characters have proven to be among the most useful ones for diagnostic purposes and for determining either phenetic or phylogenetic relationships. Therefore, despite its shortcomings, treating AFLPs as phenotype data is a viable alternative and makes use of a technique that is reliable, easy to use, with highly reproducible results, and a system with a high multiplex ratio.

The results of a scientific investigation are limited by the sampling method used to collect the data set. In AFLP, variables such as primer choice, selective amplification, and now the size range of the bands scored may influence one's results. The upper part of the AFLP gel contains bands that may be more difficult to resolve, while the lower part of the gel is usually well resolved. In this investigation, the bands (273 bp) from the upper middle portion of the gel (Fig. 1) exhibited a much higher sequence identity than the bands (159 bp expected) from the lower part of the gel. It appears that the upper middle part of the AFLP gel might be more reliable in terms of sequence identity. Caution is indicated when exploiting AFLP data. In studies where the data are treated as another phenotype, the gel patterns may be useful without proving sequence identity. However, when extending

Fig. 4. Example of alignment of sequences from AFLP comigrating polymorphic fragments of expected 159-bp band size in *E. pallida* var. *tennesseensis* (refer to Fig. 1, arrow 2 and text) demonstrating sequence heterogeneity among clones. Four of the 10 shown are identical in length (125 bp) and in sequence. Five clones are of equal length (108 bp). Twenty clones from sample ET51-3 and one clone from ET52-1 have identical sequences. The sixth clone (ET52-2) from the same sample as ET52-1 is different in length with 91 bp and very different in nucleotide sequence. Black, 100% identity; dark grey, 80% identity; light grey, 60% identity; white, less than 50% identity.

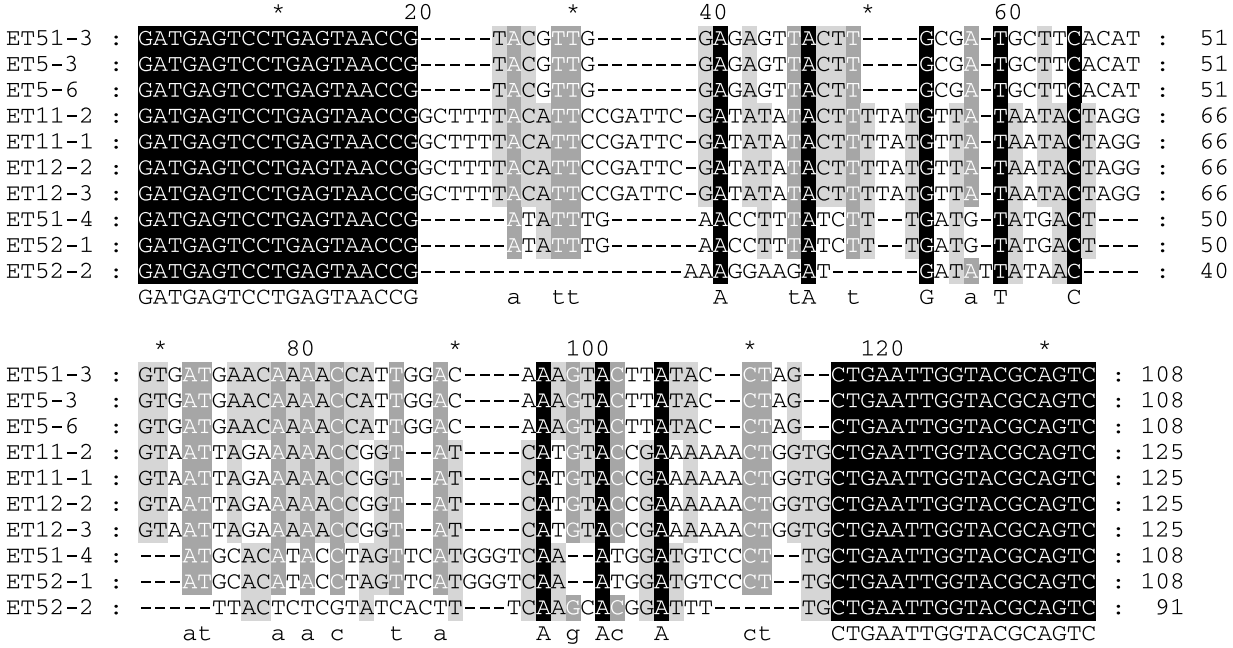
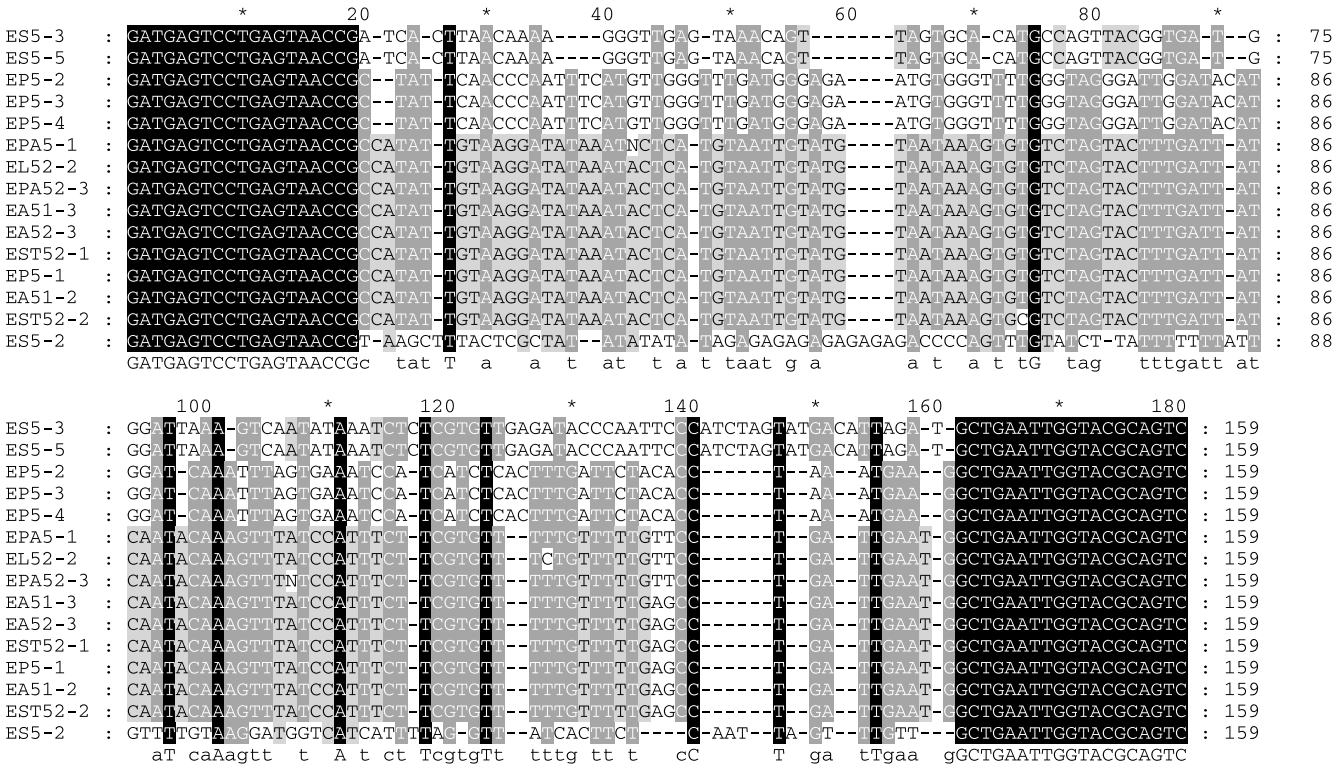


Fig. 5. Example of alignment of sequences from AFLP comigrating polymorphic fragments of identical size, 159 bp separated from the other sizes, across species and varieties. ES, *E. pallida* var. *sanguinea*; EP, *E. purpurea*; EPA, *E. atrorubens* var. *atrorubens*; EL, *E. laevigata*; EA and EST, *E. pallida* var. *angustifolia*. Note the four different types of sequences (see text). (First letters in label are according to field labels.)



the use of AFLPs to comparative studies such as genetic diversity estimates or phylogenetic analyses, the extent of sequence identity for all of the comigrating bands may first need to be demonstrated.

Acknowledgements

The financial support of the Trout Lake Farms LLC and Amway/Nutralite Corporation is gratefully acknowledged. We thank Drs. Stephen Molnar and Nicholas Tinker, both at ECORC, for their critical comments on an earlier draft of this manuscript.

References

Badr, A., El Shazly, H., El Rabey, H., and Watson, L.E. 2002. Systematic relationships in *Lathyrus* sect. *Lathyrus* (Fabaceae) based on amplified fragment length polymorphism (AFLP) data. Can. J. Bot. **80**: 962–969.

Baum, B.R., Mechanda, S.M., Livesey, J.F., Binns, S.E., and Arnason, J.T. 2001. Predicting quantitative phytochemical markers in single *Echinacea* plants or clones from their DNA fingerprints. Phytochemistry, **56**: 543–549.

Brahmasandra, S.N., Burke, D.T., Mastrangelo, C.H., and Burns, M.A. 2001. Mobility, diffusion and dispersion of single-stranded DNA in sequencing gels. Electrophoresis, **22**: 1046–1062.

Bremer, B. 1991. Restriction data from cpDNA for phylogenetic reconstruction: is there only one accurate way of scoring? Plant Syst. Evol. **175**: 39–54.

Chalhoub, B.A., Thibault, S., Laucou, V., Rameau, C., Höfte, H., and Cousin, R. 1997. Siver staining and recovery of AFLP™ amplification products on large denaturing polyacrylamide gels. Biotechniques, **22**: 216–220.

Clark, A.G., and Lanigan, C.M.S. 1993. Prospects for estimating nucleotide divergence with RAPDs. Mol. Biol. Evol. **10**: 1096–1111.

Doyle, J.J., and Doyle, J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. **19**: 11–15.

Ipek, M., and Simon, P. 2003. Sequence characterization of polymorphic AFLP fragments in garlic (*Allium sativum* L.). Plant and Animal Genome, Final Abstracts Guide, San Diego, Calif. Poster 685. p. 245.

Lamboy, W. 1994. Computing genetic distance similarity coefficient from RAPD data, the effects of PCR artifacts. PCR Methods Appl. **4**: 31–37.

Meksem, K., Ruben, E., Hyten, D., Triwitayakorn, K., and Lightfoot, D.A. 2001. Conversion of AFLP bands into high-throughput DNA markers. Mol. Gen. Genet. **265**: 207–214.

Mueller, U.G., and LaReesa Wolfenbarger, L. 1999. AFLP genotyping and fingerprinting. Trends Ecol. Evol. **14**: 389–394.

Nicholas, K.B., and Nicholas, H.B., Jr. 1997. Genedoc: a tool for editing and annotating multiple sequence alignments. Multiple sequence alignment and shading utility, version 2.6.002. Available from http://www.psc.edu/biomed/genedoc [accessed 2000].

Patterson, C. 1988. Homology in classical molecular biology. Mol. Biol. Evol. **5**: 603–625.

Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, M.J., Tingey, S.V., and Rafalski, A. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellites) markers for germplasm analysis. Mol. Breed. **2**: 225–235.

Robinson, J.P., and Harris, S.A. 1999. Amplified fragment length polymorphisms and microsatellites: a phylogenetic perspective. *In* Which DNA marker for which purpose? Chap. 12. Final Compendium of the Research Project Development, optimization and validation of molecular tools for assessment of biodiversity in forest trees in the European Union. *Edited by* E.M. Gillet. DGXII Biotechnology FW IV Research Programme Molecular Tools for Biodiversity. Available from http://webdoc.sub.gwdg.de/ebook/y/1999/whichmarker/index.htm [accessed 1999].

Rouppe van der Voort, J.N.A.M., van Zandvoort, P., van Eck, H.J., Folkerstma, R.T., Hutten, R.C.B., Draaistra, J., Gommers, F.J., Jacobsen, E., Helder, J., and Bakker, J. 1997. Use of allele specificity of comigrating AFLP markers to align genetic maps from different potato genotypes. Mol. Gen. Genet. **255**: 438–447.

Rousseau, J., Drouin, G., and Slater, J.W. 2000. Gel electrophoretic mobility of single stranded DNA: the two reptation field-dependent factors. Electrophoresis, **21**: 1464–1470.

Sambrook, J., Fritsch, E.F., and Maniatis, T.1989. Molecular cloning. A laboratory manual. 3 vols. Cold Spring Harbor Laboratory Press, Plainview, N.Y.

Savelkoul, P.H.M., Aarts, H.J.M., de Haas, J., Dijkshoorn, L., Duim, B., Otsen, M., Rademaker, J.L.W., Schouls, L., and Lenstra, J.A. 1999. Amplified fragment length polymorphism analysis: the state of an art. J. Clin. Microbiol. **37**: 3083–3091.

Shan, X., Blake, T.K., and Talbert, L.E. 1999. Conversion of AFLP markers to sequence- specific PCR markers in barley. Theor. Appl. Genet. **98**: 1072–1078.

Thompson, J.D., Higgins, D.G., and Gibbons, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighing, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**: 4673–4680.

Virk, P.S., Newbury, H.J., Jackson, M.T., and Ford-Lloyd, B.V. 2000. Are mapped markers more useful for assessing genetic diversity? Theor. Appl. Genet. **100**: 607–613.

Vos, P., Hogers, R., Bleeker, M., Van de Lee, T., Hornes, M., Freijters, A.,.Pot, J., Peleman, J., Kuiper, M., and Zabeau, M. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. **21**: 4407–4414.

Wong, A., Forbes, M.R., and Smith, M.L. 2001. Characterization of AFLP markers in damselflies: prevalence of codominant markers and implications for population genetic applications. Genome, **44**: 677–684.

Zande, L. van de, and Bijlsma, R. 1995. Limitations of the RAPD technique in phylogeny reconstruction in Drosophila. J. Evol. Biol. **8**: 645–656.