NOTE TO USERS

The CD is not included in this original manuscript.

This reproduction is the best copy available.

uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Subbaiah M. Mechanda
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Biology)
GRADE / DEGREE

Department of Biology
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Molecular Analysis of Genetic Diversity in the Native Populations of *Echinacea*

TITRE DE LA THÈSE / TITLE OF THESIS

D. Johnson
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

B. Baun
CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

J. Arnason

G. Carmody

T. Dickinson

G. Drouin

M. Smith

Gary W. Slater
LE DOYEN DE LA FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTDOCTORALES /
DEAN OF THE FACULTY OF GRADUATE AND POSTDOCORAL STUDIES

MOLECULAR ANALYSIS OF GENETIC DIVERSITY IN THE NATIVE POPULATIONS OF *ECHINACEA*.

SUBBAIAH M. MECHANDA

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies
University of Ottawa
In partial fulfillment of the requirements for the degree for the Ph.D. Degree
From the Ottawa-Carleton Institute of Biology

Faculty of Science
Department of Biology
university of Ottawa
Ottawa, Ontario
Canada

Your file Votre référence
ISBN: 0-494-10995-5
Our file Notre référence
ISBN: 0-494-10995-5

ABSTRACT

Echinacea is a North American medicinal plant that was used for medicinal purposes by Native Americans. Modern uses of *Echinacea* preparations include the treatment of colds and influenza, wound healing, and treatment of candidiasis. While demand is increasing, the natural populations of *Echinacea* are dwindling due to wild crafting and the loss of habitat caused by urban development. Understanding of the genetic diversity of natural populations is crucial in the management and conservation of the natural populations of *Echinacea*. Genetic diversity of natural populations of *Echinacea* was estimated using AFLP markers. Population genetic analysis of the AFLP data put the variation at 40% among and within the populations. Canonical analysis supported the four species concept of the morphometric analyses, however phylogenetic analysis supported some of the four species, but there was very little support by all the other analyses of the eight varieties of *Echinacea*. No correlations were found between the genetic distances and the geographic distances at the population level. The correlations between data sets and primer pairs was high at the species level, but was lower at the variety level. AFLP markers that were species or variety specific could not be identified. The comigrating bands that are monomorphic within and among species and varieties are 90% identical in their nucleotide sequences while the polymorphic bands within and among species and varieties are only 25% identical. The assumption that comigrating AFLP fragments have identical sequences within species and varieties is not valid in *Echinacea*. Based on these findings, AFLP markers were treated as phenotypes rather than genotypes and as such used for phylogenetic analysis of *Echinacea* populations. Phylogenetic analysis supported only two species, *E. purpurea*

and *E. laevigata* that separated into clear groups with all the individuals together, whereas *E. atrorubens* and *E. pallida* were separated into groups, but the individuals were scattered among the groups. The varieties could not be grouped, as the individuals were scattered among different groups. However discriminant analysis of the AFLP data revealed bands that could be used to identify the four species and only three of the eight varieties *Echinacea*.

Tissue culture procedures for mass propagation of *Echinacea purpurea* (L.) Moench for producing uniform plants from leaves of mature plants were established. However, the method was not applicable for all the other species and varieties of *Echinacea*.

RÉSUMÉ

Les échinacées sont des plantes médicinales qu'utilisaient jadis les Autochtones d'Amérique du Nord. Elles entrent encore aujourd'hui dans la composition de médicaments, qui sont utilisés pour soigner le rhume et la grippe, hâter la cicatrisation des plaies et traiter la candidose et font l'objet d'une demande croissante. Les populations naturelles d'échinacées sont en régression à cause de la cueillette en milieu sauvage et de la destruction des habitats due à l'urbanisation. Or, leur gestion et leur conservation dépendent étroitement de nos connaissances sur leur diversité génétique. Une analyse réalisée à l'aide de marqueurs AFLP a permis d'établir à 40 % la variation génétique qui existe au sein des populations naturelles d'échinacées et entre ces populations. L'analyse typologique et PCO ont corroboré les résultats de l'analyse morphométrique selon laquelle nous aurions affaire à quatre espèces, mais elles n'ont pas donné de résultats concluants quant à l'existence de huit variétés. En effet, ces séries de données présentent des corrélations fortes en ce qui concerne les espèces, mais plus faibles en ce qui concerne les variétés. Par ailleurs, il semble n'exister aucune corrélation entre les distances génétiques et les distances géographiques au niveau des populations. Nous n'avons pas réussi à trouver de marqueurs AFLP spécifiques aux espèces ni aux variétés. Toutefois, l'analyse discriminante des données AFLP a révélé des bandes qui pourraient servir à distinguer les espèces ainsi que trois des variétés. L'hypothèse selon laquelle les fragments AFLP de taille identique afficheraient les mêmes séquences au sein de chaque espèce ou variété n'est pas valide. La proportion des bandes comigratrices qui sont identiques au sein d'une espèce ou variété ainsi qu'entre espèces ou variétés atteint 90 % dans le cas de la bande monomorphe étudiée, mais seulement 25 % dans le cas de la

bande polymorphe. Compte tenu de ces résultats, nous avons assimilé les marqueurs AFLP à des phénotypes plutôt qu'à des génotypes, et nous les avons utilisés pour une analyse phylogénique des populations. Cette analyse nous a permis de confirmer l'existence de deux des espèces (*Echinacea purpurea* et *E. laevigata*), qui correspondent chacune à un groupe distinct. Dans le cas des deux autres (*E. atrorubens* et *E. pallida*), les sujets étudiés se répartissent entre plusieurs groupes. De même, il n'a pas été possible de regrouper les variétés, puisque leurs sujets sont également dispersés entre divers groupes. L'analyse typologique et PCO ont permis de corroborer les quatre espèces récentes, mais non les varieties.

La méthode de culture de tissus pour la propagation massive d'*Echinacea purpurea* (L.) Moench en vue de la production de plants uniformes à partir de feuilles de plants adultes a été établie. Toutefois, cette méthode n'était pas applicable à toutes les autres espèces et variétés d'*Echinacea*.

ACKNOWLEDGEMENTS

I would like to thank my Co-supervisors Drs. Bernard R. Baum and Douglas A. Johnson, for helping me in all the aspects of my doctoral thesis. I am very grateful to them for the guidance, inspiration, and insights into various aspects of data collection, analysis and writing during the compilation of my thesis.

I am highly appreciative of my research committee members Drs. John T. Arnason, George R. Carmody and Myron M. Smith for the excellent input during the work and helpful comments during meetings and in the writing of the thesis.

Financial support from the Trout Lake Farms/ Amway / Nutralite Corporation and MediPlant Inc, paved the way for this work and thesis. I am very indebted to them all. The provision of laboratory space, growth chambers, equipments and supplies during my work from Agriculture and Agri-Food Canada, Research Branch, ECORC Ottawa, is highly appreciated. The technical help, patience and cooperation from my co-workers, Grant Bailey, Vahab Soleimani, Tara Potter, Kathy Upton, and Dr. Shannon Binns were commendable, I am very grateful to them all.

I am very grateful to my wife Kaveri, sons Nikhil and Vikram for their support during my research, thesis work and understanding for all the missed quality time with them. I would like to thank my parents, Mr. M. A. Muthappa and Mrs. Tara Muthappa and my in-laws Mrs. Nanjamma Chinnappa and Mr. B. M. Chinnappa for their encouragement and support throughout my education. I would like to thank my brother Lt. Col. M. M. Poonacha, sister Mrs. Neelamma Aiyappa and the families for all their understanding and support. Finally, I would like to thank Gouri Chinnappa and Vidas Vitkus for their understanding and support during my doctoral studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2, 4-D	2, 4- Dichlorophenoxyacetic acid
AFLP	Amplified restriction fragment polymorphism
AMOVA	Analysis of Molecular Variance
BA	Benzylaminopurine
BLAST	Basic Local Alignment Search Tool
bp	base pairs
CDA	Canonical discriminant analysis
CTAB	Cetyltrimethylammonium bromide
DNA	Deoxyribo Nucleic Acid
EDTA	Ethylene diamine tetra-acetic acid
GDE	Genetic diversity estimate
IBA	Indole-3-butyric acid
MS	Murashige and Skoog medium
NAA	∝-naphthaleneacetic acid
NCBI	National Centre for Biotechnology Information
NJ	Neighbor Joining
NTSYS-pc	Numerical Taxonomic Systems for personal computers
PAGE	Polyacrylamide Gel Electrophoresis
PCA	Principal component analysis
PCO	Principal coordinate analysis
PCR	Polymerase Chain Reaction
PPM	Plant Preservation Mixture.

RAPD	Random Amplified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism
SCAR	Sequence Characterized Amplified Regions
SSR	Simple Sequence Repeats
STS	Sequence Tagged Sites
WPM	Woody plant medium
UPGMA	Unweighted Pair Group Method with Arithmetic Averaging

CHAPTER ONE

General introduction

The overall objectives of this study were to estimate the level of genetic diversity in the natural populations and some commercial lines of *Echinacea*; to develop Amplified Restriction Fragment Polymorphism (AFLP) markers that could identify the different groups of native populations of *Echinacea*; and, to develop tissue culture methods for the propagation of uniform *Echinacea* plants, both for commercial reasons and as part of an effort to replenish the dwindling native populations of *Echinacea*. This thesis is presented in five chapters to meet all the objectives. In the first chapter, a literature review is presented that describes *Echinacea* in general, the use of AFLP for estimating genetic diversity in different areas of biology and its various applications, and finally, a brief introduction to tissue culture of *Echinacea*. In chapter two, one of the major assumptions of the AFLP method, that comigrating bands are identical in sequence has been assessed using *Echinacea* as a multilevel taxon model. In chapter three, a detailed analysis of the diversity of natural populations and commercial lines of *Echinacea* using AFLP is presented. Sequencing of the comigrating bands and the genetic diversity of *Echinacea* were done at the same time, but based on the results of the sequences, AFLP data was used as phenotypic data for further analysis. In chapter four, a tissue culture method for the direct regeneration of *E. purpurea* (L.) Moench is described. In the fifth and final chapter, all the results that have been generated during this study have been put into perspective and the editorial comments for each chapter have been addressed with general conclusions from this work.

Echinacea chemistry, taxonomy and distribution

Echinacea is a native North American plant, commonly known as the purple coneflower. Traditionally *Echinacea* has been used by the Native American people for treating respiratory problems, sores, wounds and a variety of other ailments such as small pox and mumps (Brochers et al. 2000; Hobbs 1998). According to monographs from the German Commission E Monograph of *Echinacea*, modern uses of the herb include treatment of colds and influenza, lung infections, wound healing and candidiasis (Blumenthal et al. 2000). It has become one of the most popular medicinal botanicals in Europe and North America, consistently among the top ten medicinal plants in terms of sales (Briskin, 2000). Each year about $ 300 million (US) is spent just on *Echinacea* alone despite paucity of clinical evidence that it is an effective treatment (Barret, 2003).

Three species of *Echinacea*, *E. purpurea*, *E. angustifolia* DC. var. *angustifolia* and *E. pallida* (Nutt.) Nutt. (sensu McGregor 1968) are commonly sold as phytomedicines (Brochers 2001; Percival 2000). However, there is confusion among these species in commercial trade as well as sales of other non-commercial *Echinacea* species in the market (Hobbs 1990). Different formulations use different parts of the plant including leaf, roots or flowers (Bauer 1999). The concentrations of active ingredients such as cichoric acid and alkylamides varies depending on the species, the type of formulation, the time of harvest, storage and processing and the extraction procedure (Bauer 1999). These complications have been responsible for problems in maintaining quality control, and assessing efficacy in the clinical trials (Bauer 1999; Kigler 2002). Thus both morphological and molecular identification of the species and varieties in natural populations of *Echinacea* is an important priority.

Five classes of chemical constituents are under investigation for active ingredients, such as immunomodulators, and for standardizing the formulations. These include caffeic acid derivatives, alkylamides, polyacetylenes, glycoproteins and polysaccharides (Bauer 1998; Awang 1999). Echinacoside, a caffeoylic acid derivative is variable in the roots and has been used to distinguish between *E. angstifolia* and *E. pallida* (Awang 1999). Cichoric acid has been found to be variable between *E. purpurea* and *E. angustifolia* and has been used to distinguish between them (Bauer 1998). Alkylamide contents also vary among the *Echinacea* species and varieties; can be used to distinguish between *E. purpurea* and *E. angustifolia* (Bauer and Wagner 1991). The caffeic acid derivatives such as cichoric acid, echinacoside, chlorogenic acid, alkylamides, mainly isobutyl amides have been assigned some of the therapeutic effects of *Echinacea* (Letchamo et al. 1999).

Binns et al. (2002b) determined the phytochemical variation in the native populations of *Echinacea* using high-pressure liquid chromatography (HPLC). The amount of cichoric acid was highest in the variety *E. pallida* (Nutt.) Nutt. var. *sanguinea* (Nutt.) but the highest amounts of alkamides 1-3 and 7 were found in the species *E. purpurea*. Further, they used canonical discriminant analysis (CDA) to analyse the phytochemical variations among the four species and nine varieties. Analyses of this data lead to a revision of the taxonomy (described below, see Table 1.1). Binns et al. (2002b) found the revised species to be phytochemically distinct, with cynarin and echinacoside showing the greatest amount of variation in the CDA. The varieties did not show the same distinctness with cichoric acid and echinacoside being the most important

phytochemicals. Cichoric acid and ketoalkene can be used as markers for germplasm improvement and quality control of *Echinacea* phytomedicines.

Echinacea belongs to tribe Heliantheae in the family Asteraceae. *Echinacea* is a diploid (2n=22) plant; it is an open pollinated species with a high level of introgressants, that are triploid (2n=33) and some even tetraploid (2n=44), (McGregor 1968). McGregor (1968) was the first to publish a complete taxonomic treatment of the genus *Echinacea* using morphological characters such as leaf and stem trichomes, leaf lamina, root types and chromosome numbers on wide scale sampling. The distribution of natural populations of *Echinacea* extends from the Atlantic drainage area of US and into Canada, but not Mexico (McGregor 1968; McKeown 1999). They are adventives in the New England, but are rare in the Great Lakes area, scanty above 1371 meters towards the Rocky Mountains and were not found in New Mexico, southern Texas and Florida.

The distribution of natural populations of *Echinacea* species and varieties has been described by McGregor (1968) and Binns et al. (2002a). *E. angustifolia* was found in the dry prairies and barrens from Minnesota to Saskatoon and further south to Oklahoma, Kansas and Texas. *E. angustifolia* DC. var. *strigosa* McGregor was found in the rocky prairies with gypsiferous soils in Kansas, central Oklahoma and extreme north central Texas. *E. atrorubens* Nutt. was found in heavy moist prairie soils in the prairies, in a narrow band from Houston Texas, to Ardmore, Oklahoma, and Topeka, Kansas. *E. laevigata* was found in the open woods and fields in Pennsylvania, Virginia and south to mountains of Georgia. *E. pallida* was distributed in the rocky prairies, open wooded hillsides and glades of Northern Indiana, Illinois, Iowa, western Missouri, Arkansas,

Table 1.1. Revised classification of *Echinacea*. The species and varieties of *Echinacea*, using morphological characters and chromosome counts by McGregor (1968) and morphometric and chemotaxonomic analysis by Binns et al. (2002a).

McGregor (1968)	Binns et al. (2002a)
E. atrorubens	*E. atrorubens* var. a*trorubens*
E. paradoxa var. *neglecta*	*E. atrorubens* var. *neglecta*
E. paradoxa var. *paradoxa*	*E. atrorubens* var. *paradoxa*
E. laevigata	*E. laevigata*
E. angustifolia var. *angustifolia*	*E. pallida* var. *angustifolia*
E. angustifolia var. *strigosa*	*E. pallida* var. *angustifolia*
E. pallida	*E. pallida* var. *pallida*
E. sanguinea	*E. pallida* var. *sanguinea*
E. simulata	*E. pallida* var. *simulata*
E. tennesseensis	*E. pallida* var. *tennesseensis*
E. purpurea	*E. purpurea*

northeastern Texas, and North to Nebraska. *E. sanguinea* Nutt. was distributed ranging from north to southeast Oklahoma and southwest Arkansas. The distribution of *E. simulata* McGregor, extended from eastern Missouri, southwest to north central Arkansas. *E. paradoxa* (Norton) Britton var. *paradoxa* was found in the rocky prairies, glades and bald knobs of the central and western Ozark region of Missouri and north-central Arkansas. *E. paradoxa* (Norton) Britton var. *neglecta* (McGregor) was found in the rocky prairies and open wooded hillsides of the Arbuckle Mountain area of Oklahoma. *E. purpurea* was found in the rocky open woods, thickets and prairies of Georgia to Louisiana and eastern Oklahoma, north to Ohio, Michigan, Illinois and Iowa. *E. simulata* (McGregor) was distributed in the rocky wooded hillsides and prairies of north central Arkansas, eastern Missouri, western Illinois and west-central Kentucky. *E. sanguinea* Nutt. was distributed in the sandy open Pine Barrens and prairie of Southeastern Oklahoma, southwestern Arkansas, western Louisiana and eastern Texas. *E. tennesseensis* (Beadle) Small was distributed in the dry gravelly hills and Cedar barrens in LaVergne Tennessee. These descriptions guided the collection of wild plant materials that were used in this thesis (Fig.1.1)

According to McGregor's (1968) classification, nine species and two varieties of *Echinacea* were described (Table 1.1). He evaluated the differences among the taxa based on the morphological characters and chromosome counts only and did not employ any numerical taxonomic methods. Recently Baum et al. (1999) and Binns et al. (2001) discovered inconsistencies among these descriptions of the taxa and encountered practical difficulties with the taxonomic identification keys.

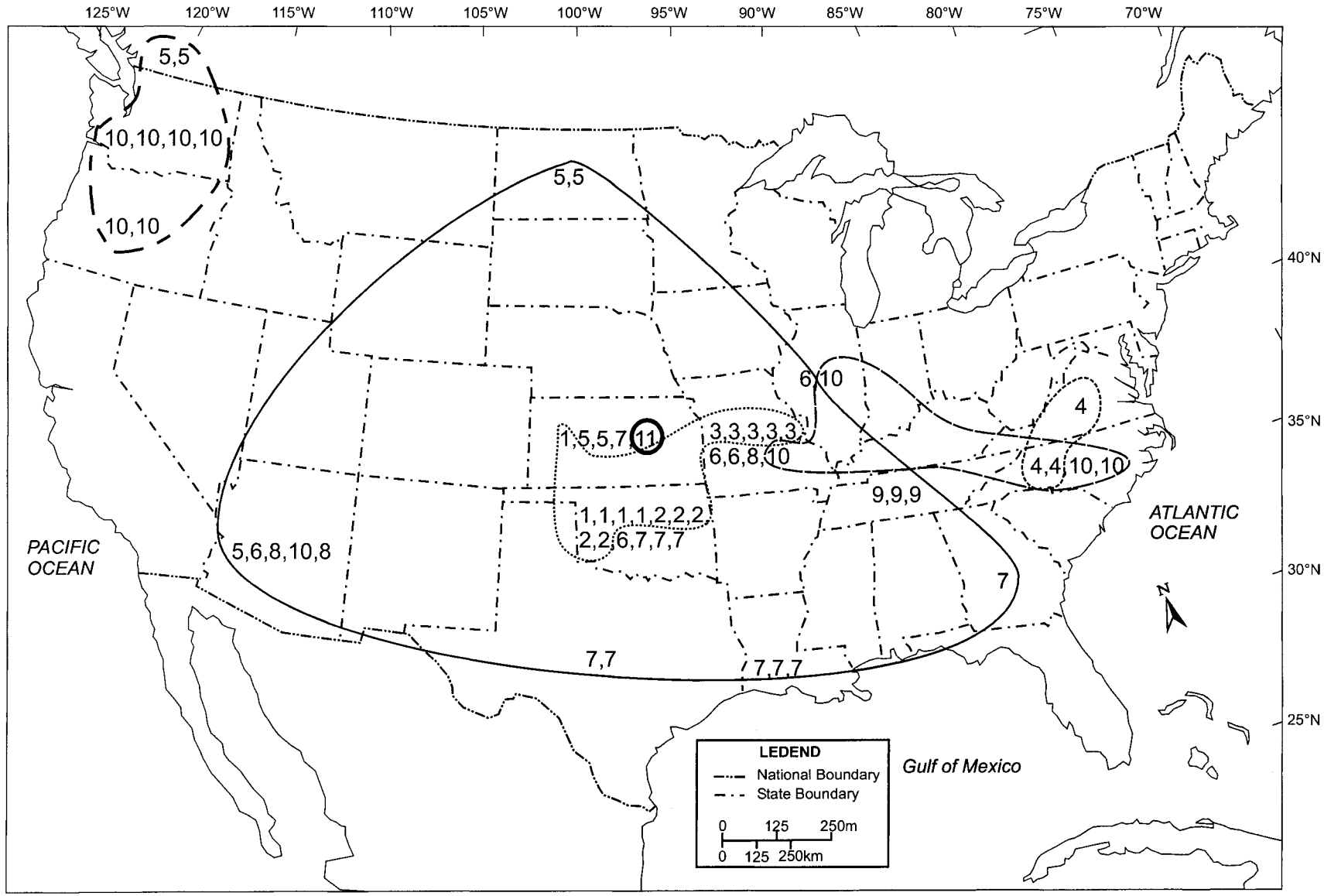
Large number of natural populations and herbarium specimens were collected by Binns et al. (2001) from a wide range of sites in North America. They revisited the distribution of *Echinacea* using the distribution map of McGregor (1968) as a guide. The center of diversity according to both the studies was in the region that contains Arkansas, Oklahoma, Kansas and Missouri. The range extended from east of the Rocky Mountains, south to Gulf coast, north and west across the Great Plains to eastern Montana and southern Saskatoon, Canada and east to the Appalachians (Binns et al. 2001; McKeown 1999).

The following (Fig.1.1) was the distribution of natural population of *Echinacea* species and varieties as reported by Binns et al. (2001). *E. purpurea* was found in the wooded hillsides and banks near waterways in the western limits of Iowa to Arizona, Louisiana and east to Pennsylvania in the north and Georgia in the south. *E.* subgenus *pallida* was found in the habitats ranging from Canadian prairies south along the high plains to Texas, bordered by Montana, Wyoming and Colorado in the west through low plains and pine oak forests of Minnesota, Wisconsin and Illinois, east to Ohio and south to Louisiana, the cedar glades of Tennessee and adventives eastward. *E. pallida* var. *pallida* ranged from the dry prairies, openings in wooded hillsides and low plains of Illinois, Minnesota, Iowa to eastern Nebraska and Kansas, western Missouri and south to northeastern Texas. *E. pallida* var. *angustifolia* has been found in rocky, dry prairies and limestone outcrops from Manitoba and Saskatoon, Canada south along the high plains to central Texas and west to Montana, Wyoming and northwest Colorado. *E. pallida* var. *sanginea* was collected from heavy clay-loam soils in pine-oak forest clearings from Texarkana south along eastern Texas border and throughout western Louisiana.

Fig. 1.1. Distribution of the species and varieties of *Echinacea* which were sampled from the wild populations representing geographical ranges of each putative species and variety for AFLP analysis. McGregor (1968) was used as the starting point for the collection sites and contour lines were based on the historical data, germplasm collections, herbarium collections and surveys of Binns et al (2002a).

The species and varieties are labelled as: 1. *E. atrorubens* var. *atrorubens*. 2. *E. atrorubens* var. *neglecta* 3. *E. atrorubens* var. *neglecta*. 4. *E. laevigata* 5. *E. pallida* var. *angustifolia*. 6. *E. pallida* var. *pallida*. 7. *E. pallida* var. *sanguinea*. 8. *E. pallida* var. *simulata*. 9. *E. pallida* var. *tennesseensis* 10. *E. purpurea*

11. Hybrid. Geographic boundaries of species: _____, *E. pallida*; _ _ _ _ _ _, *E. purpurea*; ....., *E. atrorubens*; ----------, *E. laevigata*. O, Hybrid, _ _ _ _ _ _ _ _ _ cultivated populations.


125°W
120°W
115°W
110°W
105°W
100°W
95°W
90°W
85°W
80°W
75°W
70°W
40°N
35°N
30°N
25°N
5,5
10,10,10,10
10,10
5,5
6,10
4
1,5,5,7,11
3,3,3,3,3
6,6,8,10
4,4,10,10
9,9,9
1,1,1,1,2,2,2
2,2,6,7,7,7
5,6,8,10,8
7
7,7
7,7,7
PACIFIC OCEAN
ATLANTIC OCEAN
Gulf of Mexico
N
LEDEND
National Boundary
State Boundary
0 125 250m
0 125 250km

E. pallida var. *simulata* was collected in clearings in wooded areas and edges of rocky prairies in north central Arizona, eastern Missouri, western Illinois, Kentucky, and Tennessee and rarely to the southeast. *E. pallida* var. *tennesseensis* was distributed in Cedar glade endemic to central arid regions in Tennessee. *E. atrorubens* was collected from prairies and lowland stony-clay soils of Texas and Oklahoma north to Topeka, Kansas in narrow band. *E. atrorubens* var. *paradoxa* was found in Ozarkian uplands of western Missouri, dipping into Arizona; endemic in limestone glades, bald knobs and outcrops, rare. *E. atrorubens* var. *neglecta* was endemic to rocky upland prairies and openings of wooded hillsides in Arbuckle Mountains, Oklahoma. *E. laevigata* (Boynton & Beadle) Blake was found in open wooded hillsides and well drained rocky, upland soils of Virginia, North Carolina, South Carolina and Georgia. Taxonomic classification of *Echinacea* using multivariate morphometric analyses and chemotaxonomic methods was revised by Binns et al. (2002a) and established two subgenera, four species and nine varieties for *Echinacea*. They also provided descriptions of the taxa with easy to identify taxonomic keys. Their study was the most comprehensive treatment of the genus *Echinacea* to date.

However, use of morphological characters for taxonomic purposes suffers from several drawbacks (Weins 2001; Carr et al. 2003). Morphological characters are not always inheritable, are influenced by environmental factors and are sometimes ambiguous. This makes homology assessment and quantification difficult. Moreover, morphological and phytochemical identification are not possible on the products that are sold in the marketplace by the pharmaceutical industry. In order to overcome these problems and to corroborate the recent morphometric classification of *Echinacea*,

molecular methods are necessary. One of the objectives of this study was to identify AFLP markers which are species or variety specific and corroborate the DNA based methods to the morphological methods.

The first molecular study by Urbatsch et al. (1995) used restriction site variation in the chloroplast DNA to infer the phylogeny of *Echinacea*, sub tribe Rudbeckinae. They observed low levels of chloroplast DNA variation in the multiple samples within species. Phylogeny based on the internal rDNA repeat transcribed spacer sequence 2 (ITS-2) of the rDNA repeat was established by Urbatsch et al. (2000), but was limited to eight specimens representing only seven species of *Echinacea* out of the nine species and three varieties described previously (McGregor, 1968). The genetic diversity of *Echinacea* was estimated using Random Amplified Polymorphic DNA (RAPD) analysis (Kaptaeyn et al. 2002) on few accessions of the three commercially important *Echinacea* species. However, RAPD has been determined to be not very reproducible and robust (Jones et al. 1997). After the revision of classification based on morphology of *Echinacea* (Binns et al. 2002a) some of their species have become varieties and hence the molecular data should be re evaluated. The AFLP genetic diversity studies of Kim and Still (2003); Kim et al. (2004) were carried out using only 39 plants representing nine species and three varieties (sensu McGregor) of *Echinacea* and also based on Binns et al. (2002a).

These studies were not very comprehensive and did not include all the species and varieties of *Echinacea*. A more comprehensive study that analyzes large populations collected across a broad range of distribution of *Echinacea* will help to clarify the species and varieties of this genus. To enhance the *Echinacea* germplasm and understand the diversity in the natural populations, a project was initiated. As indicated above a

morphometric study based on the morphological characters and chemical content of the natural populations and commercial lines of *Echinacea* was recently completed (Binns et al. 2002a). The present study was initiated to determine if molecular support could be established for the morphometric revision of *Echinacea*. Moreover, molecular markers are also useful in estimating the genetic diversity in this species, in identifying individuals and duplicates in germplasm collections and in identifying genetically diverse parents for developing breeding and mapping populations.

Variation in morphological characters, geographical distribution and variation, cyto-genetic relationships and markers, breeding systems, cross-compatibility and biochemical markers have been used to determine relationships within and among species. These characters are central to taxonomy, but have very limited resolving power due to the small number of variables present in these characters (Karp et al.1997). Molecular markers in general are variable, cover the genome, are unambiguous, lack phenotypic plasticity, are genetically interpretable as heritable entities and are easily quantified in population genetics and systematics (Powell et al. 1996; Karp et al. 1997). The wild relatives of a species are useful sources of genetic variability for many agronomic traits and have great potential for *Echinacea* improvement. To make efficient use of the desirable pool of genetic diversity, the genomic relationships and intra and inter species genetic diversity has to be assessed (Karp et al. 1997).

A good understanding of information on the degree and distribution of genetic variation is important for critical evaluation and conservation of biodiversity. An array of molecular techniques is available for estimating genetic diversity at the molecular level (Karp et al. 1996). They are Restriction Fragment Length Polymorphisms (RFLP),

RAPD, microsatellites, AFLP, Single Nucleotide Polymorphism (SNP) and other similar techniques. RFLP markers are generated by the difference in the size of DNA fragments that are generated by restriction enzymes and hybridised with a probe. This is a very laborious technique and needs prior sequence information (Powell et al. 1996). RAPD markers are generated by amplification of random DNA segments with short arbitrary primers and are not very repeatable and reliable (Williams et al. 1990; Mueller and Wolfenbarger 1999; Powell et al. 1996)). Microsatellites are simple sequence repeats consisting of two to six base pairs that are repeated in tandem and are amplified using PCR primers specific for the regions flanking the repeat. As with RFLPs, prior sequence knowledge is required in this case of the flanking regions for primer design (Mueller and Wolfenbarger 1999; Karp et al 1996; Powell et al. 1996). SNPs are the most abundant source of genetic variation among individuals within a species and require the prior knowledge of the sequences of the expressed sequence tags or the non-coding regions for the species, to design primers for the amplification of the SNP by PCR (Rafalski 2002). The information generated from all these techniques has to be analysed by appropriate methods to get an estimate of the genetic diversity (Savelkoul et al. 1999). AFLP has become a very popular technology due to the possible generation of large number of markers from microgram amounts of DNA with no *a priori* knowledge of the sequence of the genome (Mueller and Wolfenbarger 1999; Savelkavoul et al. 1999). They are highly reproducible, are able to resolve complex genetic structures with high levels of resolution with high multiplex ratio (Powell et al. 1996) and are reliable with informative multilocus probes (Winfield et al. 1998). Thus the AFLP technique was the method of choice for this study.

Application of AFLP

The analysis of AFLP generated data requires that fragments of the same size be identical within and among samples from a population and in fact the AFLP bands that have been mapped were assumed to be homologous with respect to co-migrating bands in diverse populations within and between species (Badr et al. 2002). Homologous genes are genes that are related through a common evolutionary ancestor. Orthologous genes are those homologues that are present in different organisms and have evolved from a common ancestral gene by speciation. Paralogous genes are present in the same organism or different organisms and have evolved from a common ancestral gene by a gene duplication event. If this gene duplication event took place before a speciation event, there are paralogous genes in different genomes.

Although AFLP fragments of a specific length are not necessarily representative of a specific locus across genetic backgrounds, most of them do represent a locus (Rouppe van der voort et al. 1999). The AFLP method generates data sets that reliably reflect differentiation between entire individual genomes and it is possible that a fraction of the variation seen in AFLP is phylogenetically accurate (Hedren et al. 2001). In population genetics studies it is also assumed that AFLP markers segregate as dominant markers exhibiting presence or absent status for a given locus. However this assumption is unfounded and the application of AFLP data to population genetics may not be appropriate (Wong et al. 2001).

The following applications of AFLP, show the wide range of organisms with a broad array of areas where it has been very popular (Mueller and Wolfenbarger 1999; Savelkavoul et al. 1999).

Mapping

AFLPs were primarily used for genomic mapping to bridge the gap between genetic and physical maps (Vos et al. 1995). Linkage maps of the parental genomes in dihaploid rose were deduced using AFLP markers to identify Quantitative Trait Loci (QTL) markers linked to thorn density of the shoots (Crespel et al. 2002). High-density genetic maps suitable for applications such as map-based gene cloning were developed for Petunia using AFLP data (Strommer et al. 2002). Indirect selection of agronomically important traits and construction of genetic linkage map and QTL analysis of the tetraploid water yam was achieved using co-dominantly scored AFLPs (Mignouna et al.2002). Hoarau et al. (2002) used 1000 AFLP markers to map the QTLs that were used to identify complex yield components in sugarcane.

Variety identification

Variety/cultivar identification in plant breeding, such as Canadian durum wheat cultivar identification, was carried out using PCR-based, cultivar specific, Sequence Tagged Sites (STS) markers derived from AFLP (Soleimani et al. 2002). Lefebvre et al. (2001) used AFLP derived genetic distances to produce Unweighted Pair Group Method with Arithmetic Averaging (UPGMA) based dendrograms for variety identification, to discriminate between five varieties of pepper hybrids used in commercial production and to protect the breeder's rights. Rapeseed cultivars were identified for registration using genetic distances derived from two AFLP primer combinations by cluster and principal coordinate analyses (PCO) (Lombard et al 2000). De Rick et al. (2001) assessed the distinctness, uniformity and stability of sugar beet varieties by cluster and AMOVA analyses of AFLP data and used these for the protection of breeder's right.

Genetic diversity

Genetic diversity among Arabica cultivars of coffee and their identification using cluster analysis of AFLP markers was assessed for coffee improvement (Steiger et al. 2002). Genetic relationships among Papaya cultivars, breeding lines, unimproved germplasm and related species were estimated by cluster analysis of the AFLP data (Kim et al. 2002). McGregor et al. (2002) used AFLP markers for germplasm management, to validate taxonomic classification, to investigate the extent of redundancy and to study the distribution of genetic diversity across the collection area using cluster analysis and AMOVA of data generated from two AFLP primer pairs on two plants per accession on the entire *Solanum* germplasm collection. AFLP markers were used to identify 67 grape clones, by cluster analysis of the AFLP polymorphic bands, thus permitting the elimination of redundant germplasm and identification of the specific clones based on the AFLP pattern (Cerevera et al. 1998). AFLP patterns from very diverse geographic origins of the common vetch collections were evaluated using AFLP fingerprints by cluster and PCO analyses to identify the collection to the original sites of collection for utilization in *ex situ* germplasm collections (Potokina et al. 2002).

Taxonomy

In systematics and phylogeny of plants AFLP has recently been used at the species, subspecies and population levels as detailed in the following examples. Closely related species of *Lactuca* were identified by phenetic analyses of AFLP data by cluster and PCO analyses and were corroborated with the ITS-1 sequence study (Koopman et al. 2001). Reduction of species in the wild potato *Solanum* section *Petota* series *Longipedicellata* was achieved by cluster and parsimony analyses using AFLP data (Berg

et al. 2002). Genetic and phylogenetic relationships of Egyptian *Hordeum* L. wild plants and their cultivated relatives were inferred by parsimony analysis of the AFLP data (El-Rabey 1999). Kim and Still (2003) and Kim et al. (2004) reported the phenetic relationships and genetic diversity among the nine species and three varieties of *Echinacea* by neighbor joining and bootstrap and PCO analyses of the AFLP data. They found AFLP data supported only two clades within *Echinacea* and did not find any congruence with the taxonomic classification of *Echinacea* by Binns et al. (2002a). Ecological studies used AFLP data to estimate genetic diversity and identification of species by cluster and PCO analyses and to establish consistency between morphology and molecular based approaches (Coart et al. 2002).

Animal genetics

In animal genetics, Buntjer et al. (2002) used AFLP fingerprinting to infer phylogeny of the wild and domesticated cattle (*Bos taurus*) species by neighbours joining and bootstrap analyses to resolve anomalies in the phylogeny inferred from studies of mitochondrial DNA. A linkage map of Chicken (*Gallus domesticus*) was established using STSs derived from flanking regions of AFLP markers (Knorr et al. 2001). Giannasi et al. (2001) used neighbour joining and principal component analyses (PCA) of the AFLP data for determination of the species tree for 24 specimens of a medically important snake, white-lipped tree viper, *Trimeresurus albolabris* for species-specific antivenom production and conservation management. AFLP data was analysed by PCA and PCO analyses to discriminate species of Caribbean Anolis lizards for biodiversity assessment (Ogden and Thorpe 2002).

Microbiology

In microbiology, Guan et al. (2002) estimated genetic diversity of *E. coli* bacteria and discriminated among isolates from animal and human sources by discriminant analysis of AFLP data and used the information for epidemiological studies of *E. coli* for predicting the source of faecal pollution of water. Taxonomic identification of *Bacillus anthrasis* by cluster, cladistic and phenetic analyses using the AFLP data was used to identify a single strain as the cause of an epidemic (Keim et al.1997). Strong influence of geography on the population structure of yeast *Pichia kluyveri*, the monophyly of one variety and utility of maintaining another variety was resolved using parsimony analysis of the AFLP data combined with morphology by Ganter et al. (2000).

These examples demonstrate that AFLPs can be applied to any species from any Kingdom and used for mapping genes, estimating genetic diversity, species and population separation, individual identification, and phylogenetic studies. The binary data set derived from AFLP fingerprints can be analysed by a wide variety approaches including cluster analysis, PCO, AMOVA, discriminant analysis and phylogenetic analysis. These approaches result in identifying individuals that are useful in breeding for better crops and livestock, in the conservation and management of germplasm, and identification of species and varieties in taxonomy.

Tissue culture

Commercially grown *Echinacea* are propagated by seed or clonally using the taproot or underground rhizomes (Henthorne 2000; Adam 2002). The seeds are stratified prior to sowing and grown in flats under standard greenhouse conditions. Germination of the seed takes about 10-20 days. Eight–week-old seedlings are transplanted to the field in

the fall for cultivation and these plants take 2-4 years to reach maturity. For clonal propagation taproot or underground rhizomes are used. Aerial shoots arise as basal rosettes of leaves from the rhizomes and can be detached from the mother plant.

Echinacea is commercially cultivated for phytomedicinal purposes. In addition there is extensive wild crafting and collection from natural populations. To replenish the wild populations, there is an increasing interest in the application of *in vitro* tissue culture methods for mass production of uniform plants of these medicinal herbs. In addition the quality of the phytomedicines from *Echinacea* could be affected by contamination due to the presence of bacteria, fungi and environmental pollutants (Choffe et al. 2000), adulteration with other plant species (Hobbs, 1994) and natural variation in the active component of the preparation due to variability in plants and the environment plus handling during processing (Murch et al. 2000; Bauer, 1999). The loss of habitat and wild crafting of *Echinacea* from natural populations have resulted in species such as *E. purpurea*, *E. laevigata* and *E. tennesseensis* being declared close to extinction (U.S. Department of the Interior, 2004). Thus, there is an increasing interest in the application of *in vitro* tissue culture methods for the mass production of uniform plants of these medicinal herbs to ensure quality of the herbal products.

The reintroduction of the mass produced tissue cultured plants into their natural habitat has been considered as an option to replenish their populations (Coker and Camper 2000). Rare and endangered plant species such as *Cosmos atrosanguineus*, the chocolate cosmos plant have been regenerated by tissue culture and stored securely for long term by cryo-preservation (Wilkinson et al. 2002). Medicinally important plants such as *Thasia garganica*, drias plant (Makunga et al. 2003) the clustered sawwort,

Saussurea obvallata (DC.) a rare, threatened medicinal herb (Joshi and Dhar 2003) and Devil's claw, *Harpagophytum procumbens* and *H. zeyheri* (Levielle and Wilson 2002) have been propagated by tissue culture methods and reintroduced into their natural settings for their conservation. Gockel et al (1992) were the first to establish a tissue culture method for the regeneration of *Echinacea. angustifolia*; however they were not successful with *E. purpurea*, the species that is most widely cultivated commercially. Successful regeneration of plants from *E. purpurea* by direct somatic embryogenesis and indirect shoot organogenesis using immature petioles has been established (Choffe et al. 2000). Camper and Coker (2000) were able to produce large number of *E. purpurea* plants by tissue culture using explants from two-month-old sterile seedlings. Choffe et al. (2001) reported root organogenesis from hypocotyls and cotyledon explants of *E. purpurea*. Regeneration of *E. purpurea* from leaf explants via callus culture was developed by (Koroch et al. 2002). Different methods of *in vitro* regeneration were established for *E. angustifolia*, *E. pallida*, *E. paradoxa* and *E. purpurea*, all of which are commercially cultivated for phytomedicines (Lakshmanan et al. 2002).

The regeneration methods described above were developed only for the three commercially important species and used material derived from germinated sterile seedlings. This approach is time consuming. Other species and varieties, which may be chemically superior and agronomically desirable can be selected directly from the natural population and propagated by tissue culture of mature leaves for phytomedicine production and conservation of the natural populations. Thus a method which uses mature leaves from the field collected samples and is not dependent on the sterile seedling, is critical for mass propagation of non commercial and nearly extinct species.

The method that was developed in this study using *E. purpurea* as a model may be applicable to all the other species with few modifications.

Rationale for the Study

Genetic diversity assessment of plant species was traditionally based on morphological characters, agronomical characteristics and recently on molecular methods. However, to date very few comprehensive genetic diversity estimates of natural populations and cultivated *Echinacea* populations have been established. Thus, genetic diversity estimates derived from such data may poorly portray the actual genetic diversity of the natural populations. In this study variation at the DNA level was used to obtain genetic diversity estimates among natural populations of *Echinacea*. Furthermore this data will enable the comparison with and corroboration of, the classification of the natural populations derived from morphological and chemical markers. AFLP markers can be used for the identification and authentication of *Echinacea* at the farm and industries levels and also provide an opportunity to study an emerging crop at an early stage of selection. Tissue culture methods may be the logical route for production of uniform elite lines of *Echinacea* for pharmaceutical industry and for restoring the natural populations of the nearly extinct groups.

Hypotheses and objectives of the study.

Hypothesis 1:

Molecular methods estimate genetic diversity accurately in the natural populations of *Echinacea*.

Objective 1:

Estimate the amount of genetic diversity within and among the native populations of

Echinacea using AFLP.

Hypothesis 2:

Native populations of *Echinacea* can be grouped based on genetic distances into species and varieties.

Objective 2:

Separate the natural populations of *Echinacea* based on genetic distance and test against the current morphometric classification. Identify diagnostic fragments using canonical discriminant analysis and phylogenetic analysis.

Hypothesis 3:

Comigrating AFLP fragments are species or variety specific and homologous in *Echinacea*.

Objective 3:

Clone and sequence the comigrating AFLP fragments to determine their sequence identity.

Hypothesis 4:

Every species or varieties of *Echinacea* can be multiplied using a combination of auxin and cytokinin.

Objective 4:

Develop a method using NAA and BA to produce multiple plantlets from the mature leaves of all the species and varieties of *Echinacea*.

CHAPTER TWO

Sequence assessment of comigrating AFLP® bands in *Echinacea* - implications for comparative biological studies.

This chapter is a reproduction of the paper published in Genome: 47:15-25 (2004), by Subbaiah M. Mechanda, Bernard R. Baum, Douglas A. Johnson and John T. Arnason. The validity of the assumption in AFLP that comigrating bands in closely related species are related by descent and are identical in sequence was assessed and meets part of the objective 3 of the thesis. Comigrating monomorphic and polymorphic AFLP bands were isolated, cloned and sequenced from the four species and eight varieties of *Echinacea*. The sequences were aligned and sequence identities were determined separately for the monomorphic and polymorphic bands at 100% similarity threshold. The monomorphic bands were 75 to 98% similar at the species level whereas the polymorphic bands were 33 to 45% similar. This may introduce bias into the data set and thus the genetic diversity measures may be biased. Using AFLP data as phenotype rather than genotype may be an alternative for overcoming this bias. [This chapter is a reproduction of the original paper. Editorial comments have not been included and instead have been addressed in the Discussion.]

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 on 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 (Chaloub 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 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 was 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 is 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) electro-blotted AFLP products and probed with eleven 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*. Ipeck and Simon (2003) sequenced 79 AFLP fragments which were used in diversity analysis in garlic and found 95% of the co-migrating 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. 2004), we became concerned about the validity of this assumption. We were

especially concerned about the similarity of comigrating AFLP bands, 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 study, 435 single plant samples of the 58 populations were used (Mechanda et al. 2004b: 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 (10mM Tris, 1mM 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 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 pre-amplification 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 pre-amplification 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 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:

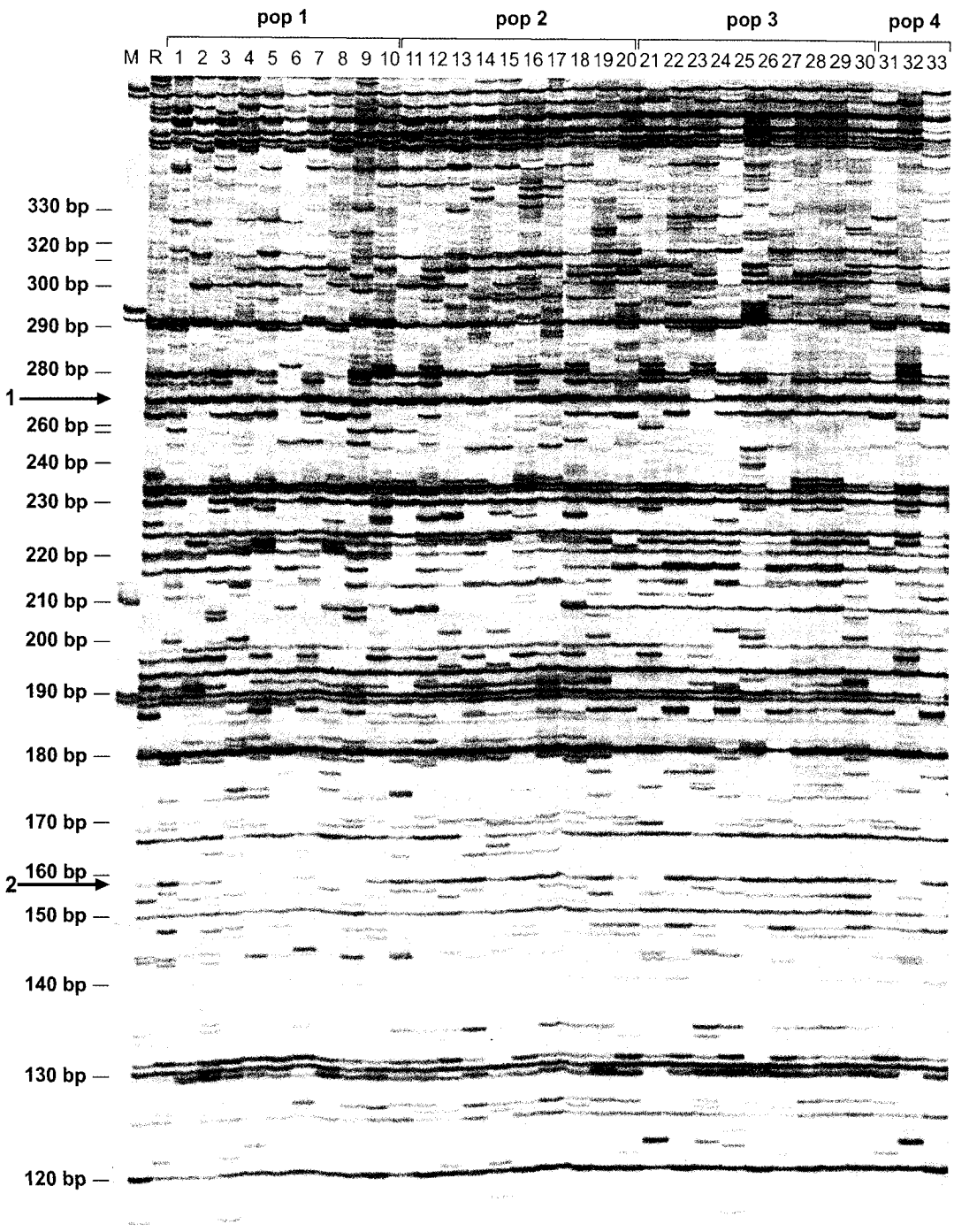
set 1, D10 Eco-GACTGCGTACCAATTC**ACC**/Mse-GATGAGTCCTGAGTAA**CCG**,

set 2, G10 Eco-GACTGCGTACCAATTC**AGC**/Mse-GATGAGTCCTGAGTAA**CCG**,

set 3, D12 Eco-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, 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, 30-330 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. 2.1). The gels were dried and autoradiographed using Kodak diagnostic X-OMAT film (Eastman Kodak, Rochester, N.Y.) at room temperature

Fig.2.1. AFLP silver-stained gel containing DNA samples from four populations of *E. atrorubens* var. *paradoxa* amplified with No. 2 (G-10) primer set combination (*Eco*R1+AGC and *Mse*1+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.


pop 1
pop 2
pop 3
pop 4
M R 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33
330 bp
320 bp
300 bp
290 bp
280 bp
1
260 bp
240 bp
230 bp
220 bp
210 bp
200 bp
190 bp
180 bp
170 bp
160 bp
2
150 bp
140 bp
130 bp
120 bp

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), then reamplified for 23 cycles using the same primer pair and conditions used to generate the band. The purified 127 bands were cloned using TOPO TA cloning® kit for Sequencing (Invitrogen). Plasmid DNA was isolated using the QIAprep Miniprep (Qiagen) plasmid kit and the purified DNA was re-suspended 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 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. 2.1). 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 2.1). 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. 2.1) demonstrates the high similarity but not complete sequence identity (Fig. 2.2). 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 *Echinacea pallida* (75.82% identity) (Table 2.1). All of the sequences for the genus *Echinacea* when aligned together were only 58.97% identical. Note the percent dissimilarity increases through the four taxonomic levels, from population to genus.

Table 2.1. Similarities between a monomorphic 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		

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
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	

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
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		

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
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			

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
	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 83.71±10.33
Mean ± SD							
Polymorphic band, size expected 159bp							
E. atrorubens var. atrorubens	1	2	91				
	2	2	91				
Total	2	4			100.00		

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
E. atrorubens var. *neglecta*	1	1	91				
	2	1	74				
Total	2	2			64.83		
E. atrorubens var. *paradoxa*	1	2	125	100			
	2	3	91, 158	39.87			
	3	2	91	100			
Total	3	7			46.40 42.66±34.71		
Mean ± SD							
E. atrorubens	*7	*13				37.60 70.41±27.23	
Mean ± SD							
E. laevigata	1	1	159				

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
	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			

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	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			

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	5	10			36.00 85.18±20.67		
Mean ± SD							
E. pallida	14	27				23.66 39.76±8.75	
Mean ± SD							
E. purpurea	1	1	159				
	2	1	159				
	3	1	159				
	4	1	159				
	5	1	160				
	5	5				45.00	
Echinacea	27	48	74-159				1.25

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
Mean ± SD							34.43±9.07

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

Fig. 2.2. Example of alignment of sequences from AFLP comigrating monomorphic fragments of 273-bp band size in *E. atrorubens* var. *paradoxa* (refer to Fig. 2.1, arrow 1 and text). Four clones from one population (EPP-1: EPP98-0606-2 Cambden 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.

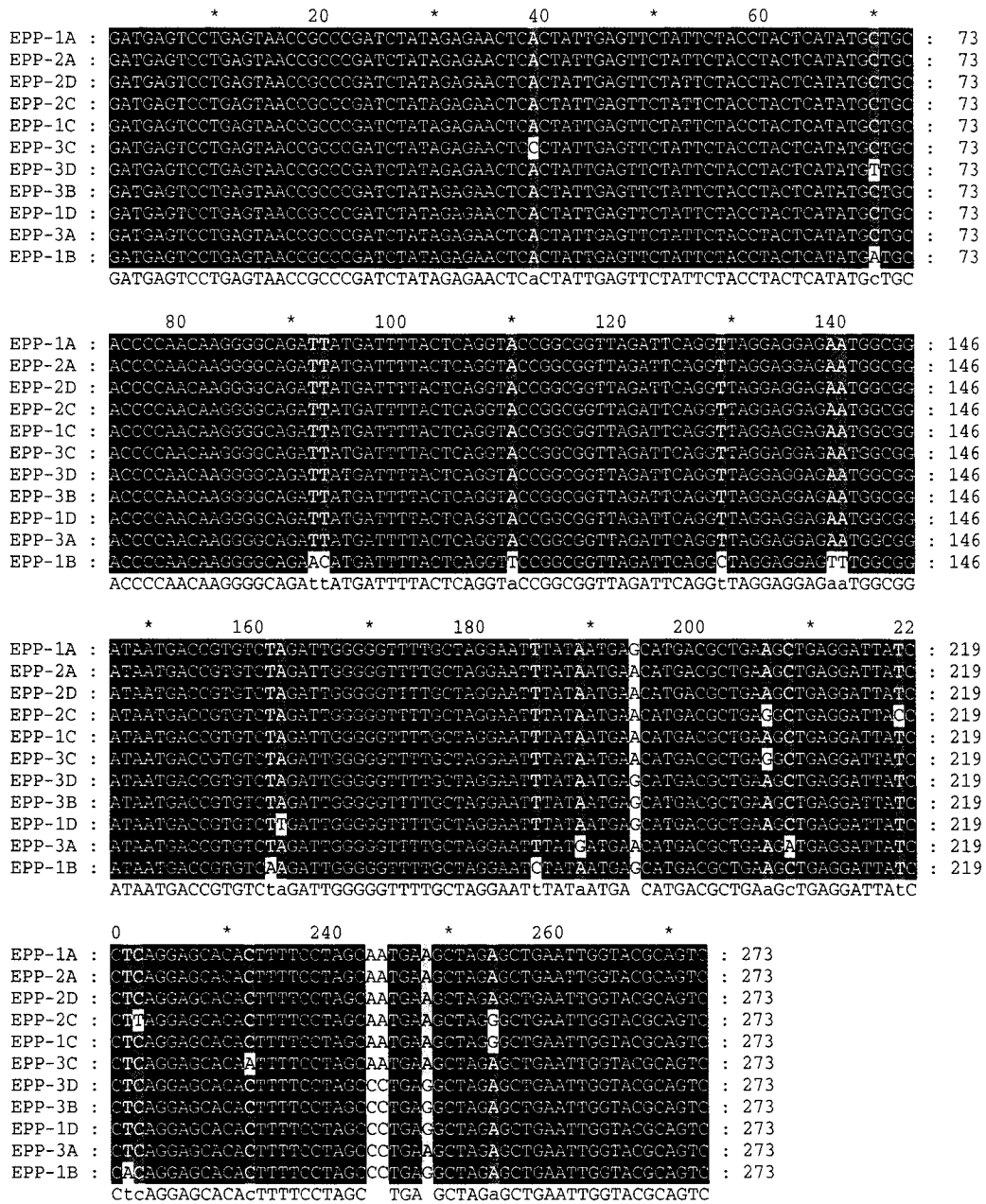
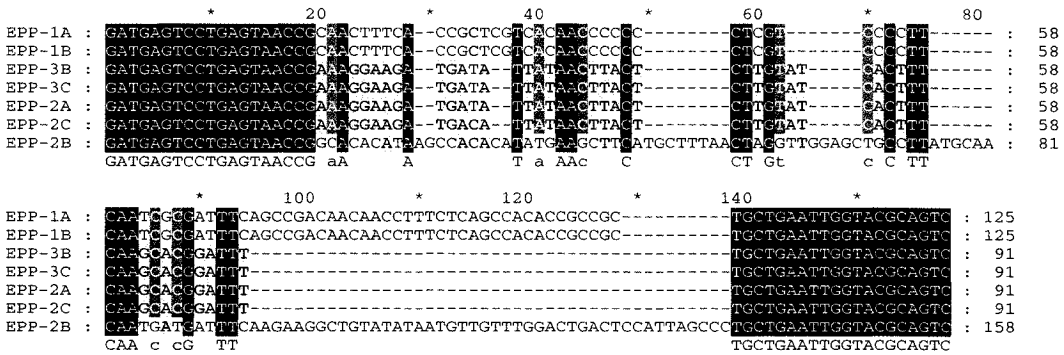

EPP-1A : GATGAGTCCTGAGTAACCGCCCGATCTATAGAGAACTCACTATTGAGTTCTATTCTACCTACTCATATGCTGC : 73
EPP-2A : GATGAGTCCTGAGTAACCGCCCGATCTATAGAGAACTCACTATTGAGTTCTATTCTACCTACTCATATGCTGC : 73
EPP-2D : GATGAGTCCTGAGTAACCGCCCGATCTATAGAGAACTCACTATTGAGTTCTATTCTACCTACTCATATGCTGC : 73
EPP-2C : GATGAGTCCTGAGTAACCGCCCGATCTATAGAGAACTCACTATTGAGTTCTATTCTACCTACTCATATGCTGC : 73
EPP-1C : GATGAGTCCTGAGTAACCGCCCGATCTATAGAGAACTCACTATTGAGTTCTATTCTACCTACTCATATGCTGC : 73
EPP-3C : GATGAGTCCTGAGTAACCGCCCGATCTATAGAGAACTCCCTATTGAGTTCTATTCTACCTACTCATATGCTGC : 73
EPP-3D : GATGAGTCCTGAGTAACCGCCCGATCTATAGAGAACTCACTATTGAGTTCTATTCTACCTACTCATATGTTGC : 73
EPP-3B : GATGAGTCCTGAGTAACCGCCCGATCTATAGAGAACTCACTATTGAGTTCTATTCTACCTACTCATATGCTGC : 73
EPP-1D : GATGAGTCCTGAGTAACCGCCCGATCTATAGAGAACTCACTATTGAGTTCTATTCTACCTACTCATATGCTGC : 73
EPP-3A : GATGAGTCCTGAGTAACCGCCCGATCTATAGAGAACTCACTATTGAGTTCTATTCTACCTACTCATATGCTGC : 73
EPP-1B : GATGAGTCCTGAGTAACCGCCCGATCTATAGAGAACTCACTATTGAGTTCTATTCTACCTACTCATATGATGC : 73
GATGAGTCCTGAGTAACCGCCCGATCTATAGAGAACTCaCTATTGAGTTCTATTCTACCTACTCATATGcTGC
EPP-1A : ACCCCAACAAGGGGCAGATTATGATTTTACTCAGGTACCGGCGGTTAGATTCAGGTTAGGAGGAGAATGGCGG : 146
EPP-2A : ACCCCAACAAGGGGCAGATTATGATTTTACTCAGGTACCGGCGGTTAGATTCAGGTTAGGAGGAGAATGGCGG : 146
EPP-2D : ACCCCAACAAGGGGCAGATTATGATTTTACTCAGGTACCGGCGGTTAGATTCAGGTTAGGAGGAGAATGGCGG : 146
EPP-2C : ACCCCAACAAGGGGCAGATTATGATTTTACTCAGGTACCGGCGGTTAGATTCAGGTTAGGAGGAGAATGGCGG : 146
EPP-1C : ACCCCAACAAGGGGCAGATTATGATTTTACTCAGGTACCGGCGGTTAGATTCAGGTTAGGAGGAGAATGGCGG : 146
EPP-3C : ACCCCAACAAGGGGCAGATTATGATTTTACTCAGGTACCGGCGGTTAGATTCAGGTTAGGAGGAGAATGGCGG : 146
EPP-3D : ACCCCAACAAGGGGCAGATTATGATTTTACTCAGGTACCGGCGGTTAGATTCAGGTTAGGAGGAGAATGGCGG : 146
EPP-3B : ACCCCAACAAGGGGCAGATTATGATTTTACTCAGGTACCGGCGGTTAGATTCAGGTTAGGAGGAGAATGGCGG : 146
EPP-1D : ACCCCAACAAGGGGCAGATTATGATTTTACTCAGGTACCGGCGGTTAGATTCAGGTTAGGAGGAGAATGGCGG : 146
EPP-3A : ACCCCAACAAGGGGCAGATTATGATTTTACTCAGGTACCGGCGGTTAGATTCAGGTTAGGAGGAGAATGGCGG : 146
EPP-1B : ACCCCAACAAGGGGCAGAACATGATTTTACTCAGGTTCCGGCGGTTAGATTCAGGCTAGGAGGAGTTTGGCGG : 146
ACCCCAACAAGGGGCAGAttATGATTTTACTCAGGTaCCGGCGGTTAGATTCAGGtTAGGAGGAGaaTGGCGG
EPP-1A : ATAATGACCGTGTCTAGATTGGGGGTTTTGCTAGGAATTTATAATGAGCATGACGCTGAAGCTGAGGATTATC : 219
EPP-2A : ATAATGACCGTGTCTAGATTGGGGGTTTTGCTAGGAATTTATAATGAACATGACGCTGAAGCTGAGGATTATC : 219
EPP-2D : ATAATGACCGTGTCTAGATTGGGGGTTTTGCTAGGAATTTATAATGAACATGACGCTGAAGCTGAGGATTATC : 219
EPP-2C : ATAATGACCGTGTCTAGATTGGGGGTTTTGCTAGGAATTTATAATGAACATGACGCTGAGGCTGAGGATTATC : 219
EPP-1C : ATAATGACCGTGTCTAGATTGGGGGTTTTGCTAGGAATTTATAATGAACATGACGCTGAAGCTGAGGATTATC : 219
EPP-3C : ATAATGACCGTGTCTAGATTGGGGGTTTTGCTAGGAATTTATAATGAACATGACGCTGAGGCTGAGGATTATC : 219
EPP-3D : ATAATGACCGTGTCTAGATTGGGGGTTTTGCTAGGAATTTATAATGAGCATGACGCTGAAGCTGAGGATTATC : 219
EPP-3B : ATAATGACCGTGTCTAGATTGGGGGTTTTGCTAGGAATTTATAATGAGCATGACGCTGAAGCTGAGGATTATC : 219
EPP-1D : ATAATGACCGTGTCTTGATTGGGGGTTTTGCTAGGAATTTATAATGAGCATGACGCTGAAGCTGAGGATTATC : 219
EPP-3A : ATAATGACCGTGTCTAGATTGGGGGTTTTGCTAGGAATTTATGATGAACATGACGCTGAAGATGAGGATTATC : 219
EPP-1B : ATAATGACCGTGTCAAGATTGGGGGTTTTGCTAGGAATCTATAATGAGCATGACGCTGAAGCTGAGGATTATC : 219
ATAATGACCGTGTCtaGATTGGGGGTTTTGCTAGGAATtTATaATGA CATGACGCTGAaGcTGAGGATTAtC
EPP-1A : CTCAGGAGCACACTTTTCCTAGCAATGAAGCTAGAGCTGAATTGGTACGCAGTC : 273
EPP-2A : CTCAGGAGCACACTTTTCCTAGCAATGAAGCTAGAGCTGAATTGGTACGCAGTC : 273
EPP-2D : CTCAGGAGCACACTTTTCCTAGCAATGAAGCTAGAGCTGAATTGGTACGCAGTC : 273
EPP-2C : CTTAGGAGCACACTTTTCCTAGCAATGAAGCTAGGGCTGAATTGGTACGCAGTC : 273
EPP-1C : CTCAGGAGCACACTTTTCCTAGCAATGAAGCTAGGGCTGAATTGGTACGCAGTC : 273
EPP-3C : CTCAGGAGCACAATTTTCCTAGCAATGAAGCTAGAGCTGAATTGGTACGCAGTC : 273
EPP-3D : CTCAGGAGCACACTTTTCCTAGCCCTGAGGCTAGAGCTGAATTGGTACGCAGTC : 273
EPP-3B : CTCAGGAGCACACTTTTCCTAGCCCTGAGGCTAGAGCTGAATTGGTACGCAGTC : 273
EPP-1D : CTCAGGAGCACACTTTTCCTAGCCCTGAGGCTAGAGCTGAATTGGTACGCAGTC : 273
EPP-3A : CTCAGGAGCACACTTTTCCTAGCCCTGAAGCTAGAGCTGAATTGGTACGCAGTC : 273
EPP-1B : CACAGGAGCACACTTTTCCTAGCCCTGAGGCTAGAGCTGAATTGGTACGCAGTC : 273
CtcAGGAGCACAcTTTTCCTAGC TGA GCTAGaGCTGAATTGGTACGCAGTC

Fig. 2.3. Example of alignment of sequences from AFLP comigrating polymorphic fragments of expected 159-bp band size in *E. atrorubens* var. *paradoxa* (refer to Fig. 2.2, arrow 2 and text) demonstrating sequence difference among clones. Two clones from one population (EPP-1: EPP98-0606-2 Cambden 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.


EPP-1A
EPP-1B
EPP-3B
EPP-3C
EPP-2A
EPP-2C
EPP-2B

Polymorphic bands

Two clones from each of the five *E. pallida* var. *angustifolia* samples were sequenced, and while duplicates had percent 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 percentage similarities (Table 2.1). The alignment presented in Fig. 2.3 demonstrates the lack of sequence identity among polymorphic bands migrating at the same position (159bp) and taken from the same gel of *E. atrorubens* var. *paradoxa* shown in Fig. 2.1. In contrast to 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 2.1). 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 bp and 161 bp, and then in only two clones (Table 2.1). In the other taxa, clones of the expected sequence length were not found at all in our sample, resulting in sequences ranging from 74 bp to 161 (Table 2.1). 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. 2.4). The last of the 10 clones,

Fig. 2.4. Example of alignment of sequences from AFLP comigrating polymorphic fragments of expected 159-bp band size in *E. pallida* var. *tennesseensis* (refer to Fig. 2.1, 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). Two 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.

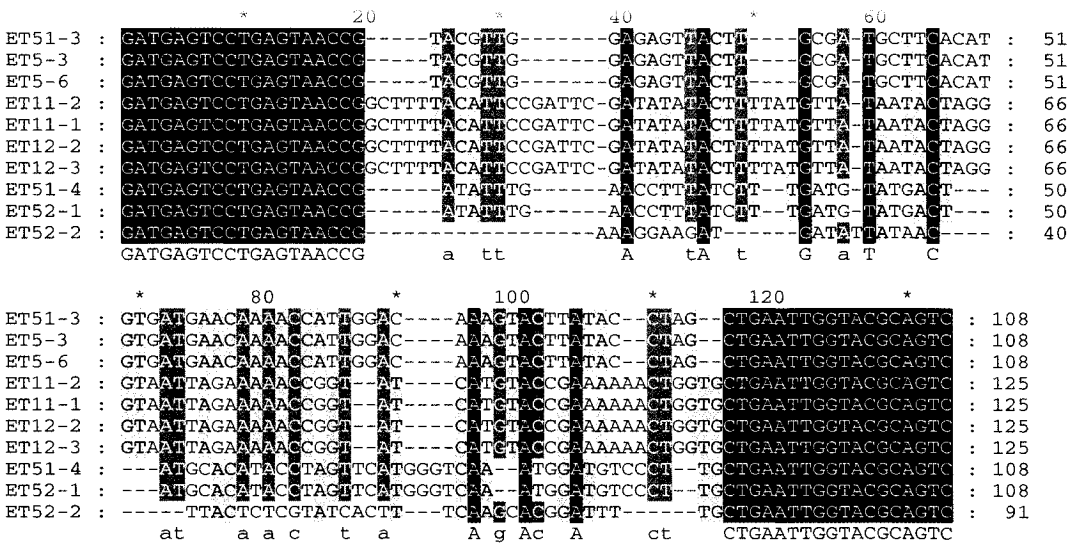
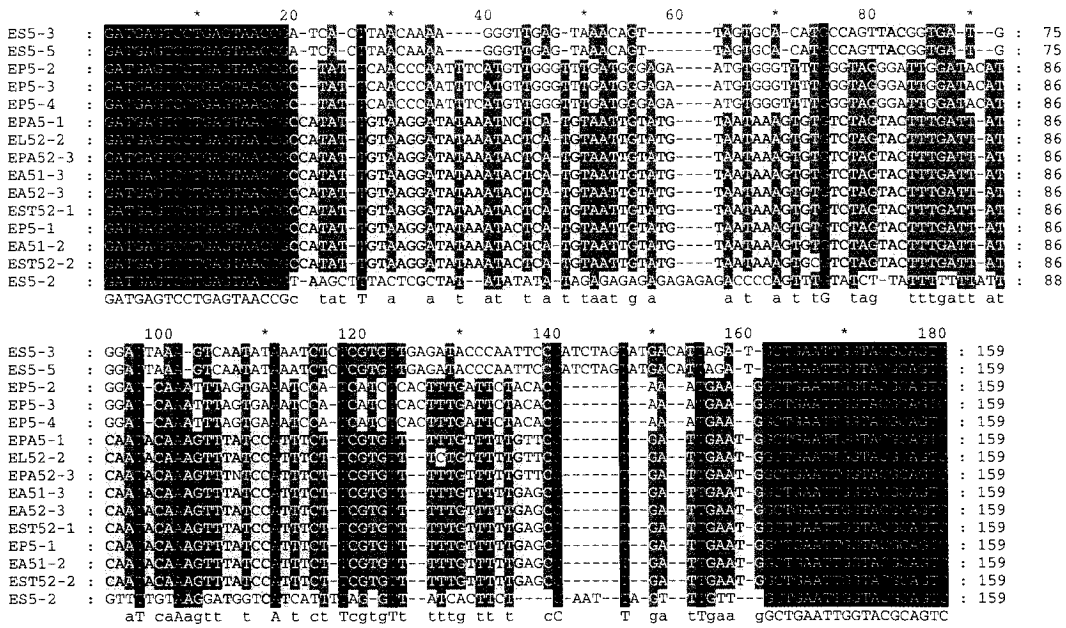


Fig. 2.5. Example of alignment of sequences from AFLP comigrating polymorphic fragments of identical size, 159 bp separated from the other sizes, across species and varieties. ES, *E. pallida* var. *sanguinea*; EP, *E. purpurea*; EPA, *E. atrorubens* var. *atrorubens*; EL, *E. laevigata*; EA and EST: *E. pallida* var. *angustifolia*. Note the four different types of sequences (see text). (First letters in label are according to field labels).



ET52-2 from the same sample as ET52-1, was different in length (91 bp) and very different in nucleotide sequence (Fig.2.4). To attempt to determine whether the observed heterogeneity as 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. 2.5). In the same alignment at position 97-100, an *Mse*1 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*; (*ii*) two sequences in *E. pallida* var. *tennesseensis* (not shown). All 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 2.1).

Discussion

False assumption of homology based on fragment length

The essence of comparative biology, especially at the genetic and taxonomic levels, depends upon 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 2.1, sample 2; Fig. 2.3) and *E. pallida* var. *tennesseensis* (Table 2.1, samples 3 and 5; Fig. 2.4). The noncoding nature of AFLPs may help to provide an explanation for the variability and multiple length variants which 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 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 (Mechanda et al. 2004a) were sequenced. Identities ranging from 75% to 98% among the four species (Table 2.1) 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 due to the extensive out breeding 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 out breeding. Therefore, because fragments of the same length are not homologous, 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. 2004b).

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 labeled 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 2.1). 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 2.1). Furthermore, we found lack of identity in both length and sequence among different clones taken from the same fragment, such as in *E. atrorubens* var. *paradoxa* EPP-2B and EPP-2C (Fig. 2.3) and *E. pallida* var. *tennesseensis* clones ET2-1 and ET2-2 (Fig. 2.4). However, when fragments of similar size were separately aligned, there was complete identity among the fragments of 91 bp. In the other fragments, 108 bp 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 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. 2.5). Physical factors affecting gel resolution may explain discrepancies among fragment length and sequence identity, e.g., reptation of DNA molecules (entangled DNA fragments migrating 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 are due to non-idealities of the separation conditions (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 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 Wolfenbarger 1999), by treating them as phenotypic markers.

Although Wong et al. (2001) inferred varying co-dominance status among AFLP polymorphic sites; they based their findings on experiments that used cloned fragments as probes of southern blots. It can be assumed that presence of hybridization signal may also be affected by inter-lane mobility shifts. Southern hybridization may yield 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 on silver stained gels.

Theoretical consideration

Several criteria have been put forward for estimating nucleotide divergence. These are based upon 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%, (*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 2.1). Monomorphic bands are not considered to be useful for comparative studies as they occur in all samples, including all 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 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 in 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 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 non-specific cross-hybridization.

Non-homologous 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. 2.1) 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 the use of AFLPs to comparative studies such as genetic diversity estimates or phylogenetic analyses, the extent of sequence identity for all the comigrating bands may first need to be demonstrated.

CHAPTER THREE

Analysis of diversity of natural populations and commercial lines of *Echinacea* using AFLP.

This chapter is a reproduction of the paper published in Canadian Journal of Botany 82: 461-484 (2004), by Subbaiah M. Mechanda, Bernard R. Baum, Douglas A. Johnson and John T. Arnason. This chapter meets objectives 1 and 2 of the thesis. Genetic diversity of the natural populations of *Echinacea* was estimated using the AFLP technique. Population genetic analysis was carried out to study the genetic diversity, at the species and variety levels. The natural populations were grouped and the genetic relationships between the four species and eight varieties were established by cluster and PCO analyses. The genetic structure indices were estimated using AMOVA analysis. Discriminant analysis was used to assess if there is justification and support for the morphometric classification. Classificatory discriminant analysis was used to obtain discriminant functions for the identification of species and varieties of *Echinacea.* Phylogenetic analysis based on the AFLP phenotype was used to identify synapomorphies for the groups of natural populations of *Echinacea*, which will aid in their identification. [This chapter is a reproduction of the original paper. Editorial comments have not been included and instead have been addressed in the Discussion.]

Introduction

Echinacea (*Echinacea purpurea* (L.) Moench, in recent years, *E. angustifolia* DC.) are medicinal plants. It was traditionally used by the First Nations of the American Great Plains and Canadian Prairies for the treatment of sores, wounds and a variety of other ailments (Tyler 1993; Hobbs 1994; Borchers et al. 2000). The distribution of this North American indigenous genus ranges from eastern and central United States to southern Canada.

Echinacea is in increasing demand, and modern uses are mainly for the treatment of colds and influenza, as well as for wound healing and treatment of candidiasis (Brevoort 1996; Borchers et al. 2000). Three species are cultivated to some extent and are sold as phytomedicines: *E. purpurea*, *Echinacea pallida* (Nutt.) Nutt. and *E. angustifolia*, now considered to be a variety, *Echinacea pallida* var. *angustifolia* (DC.) Cronq. (see Binns et al. 2002a). *Echinacea purpurea* is the main commercial species. Most echinacea plants are wild crafted (i.e. harvested from the wild) and sold in the market without authentication to species. However, the pharmaceutical industry requires correct identification to guarantee the quality of echinacea products. Populations of *Echinacea* throughout their range have been severely reduced due to habitat loss and wild crafting (Foster 1991; Blumenthal 1992). Two species, *Echinacea tennesseensis* (Beadle) J. K. Small, (now considered to be a variety, *Echinacea pallida* var. *tennesseensis* (Beadle) Binns, B. R. Baum & Arnason and *Echinacea laevigata* (C. L. Boynton & Beadle), are listed as endangered species (US Fish and Wildlife Service), while many wild populations of *E. purpurea* contain very low numbers, in some cases consisting of only few individuals (Binns et al. 2002a). Hence, understanding the genetic properties and

structure of the remaining *Echinacea* populations is needed to preserve their gene pool along with the attempts to minimize damage to habitats (Blumenthal 1992).

Echinacea, the purple coneflower genus, belongs to tribe Heliantheae in the family Asteraceae. McGregor (1968) was the first to carry out a taxonomic treatment of the genus using morphological characters and chromosome numbers from wide-scale sampling of the natural populations. As a result, nine species and two varieties were described. Binns et al. (2002a) revised the systematics of *Echinacea* using live and herbarium specimens from the natural populations. Their work was based on morphometric analyses and resulted in a classification system that identified four species and eight varieties. Prior to Binns et al. (2002a), molecular methods were used to estimate phylogenetic relationships within the Tribe Heliantheae, one by chloroplast DNA restriction site analysis (Urbatsch and Jansen 1995) and the other using the internal transcribed spacer (ITS) of the rRNA genes (Urbatsch et al. 2000). Although the scope of both molecular studies was broader than the genus *Echinacea*, a number of species from *Echinacea* were included (with a single specimen per species, sensu McGregor 1968). A third molecular study, using RAPD (Random Amplified Polymorphic DNA), with the goal of estimating the genetic relationships and diversity among the three main commercial species, *E. purpurea*, *E. angustifolia* and *E. pallida*, was based on 19 seed accessions (Kapteyn et al. 2002).

Although the RAPD method is attractive for its low cost and simplicity, the more expensive amplified fragment length polymorphism (AFLP® ; Vos et al., 1995) is now preferred because of its greater reproducibility (Jones et al. 1997) and greater multiplex ratio (Powell et al. 1996; see also review by Mueller and Wolfenbarger 1999). AFLP has

become trendy and is used for (1) the identification of taxa, for example, Badr et al. (2002) for plant species, Soleimani et al. (2002) for durum wheat cultivars, Buntjer et al. (2002) for bovine species, and Breyne et al. (2000) for *Arabidopsis* ecotypes; (2) genetic diversity studies of species, for example, *Lactuca* (Koopman et al. 2001); (3) fingerprinting of cultivars (Lombard et al. 2000); (4) the estimation of phylogenetic relationships, for example, *Lactuca* species Hill et al. (1996) *Solanum* sect. *Petota* (Kardolus et al. 1998), and a great number of other papers in many plant and animal studies.

The present molecular study was undertaken in parallel to supplement and complement the morphometric diversity study by Binns et al. (2002a) using many specimens collected from populations that were included in that study. Our objectives were (1) to assess if the systematics of *Echinacea*, based on morphometrics (Binns et al. 2002a), is supported by molecular (AFLP) data; (2) to assess whether the comigrating AFLP bands have identical sequences; (3) to estimate the molecular genetic diversity of the species and varieties in Binns et al. (2002a), and the diversity within and among the natural populations; (4) to provide a means for the identification of species and varieties by employing molecular tools; and (5) to estimate the phylogenetic relationships among the species and varieties.

Materials and Methods

A flowchart describing our analyses of natural populations of *Echinacea* is presented in Figure 3.1.

Plant material

Leaves from single plants, roots, and seeds of 58 natural populations of *Echinacea* were collected from Canada and the midwestern and southeastern United States (Table 3.1). A total of 435 single plant samples of the 58 populations were used for the present study. These populations and their accession numbers are based on the taxonomic revision of *Echinacea* (Binns et al. 2002a). The seeds of some plants were germinated in a growth chamber, and the seedlings were transferred to pots in a growth chamber and grown for 1 month at 16 h of light (25 °C): 8 h dark (20 °C) cycle. For AFLP analysis, single leaves from each population were collected on ice, frozen in liquid nitrogen and stored at -80 °C. For flow cytometry, five plants were randomly chosen, and three leaves were used for ploidy measurements for each of the four species and eight varieties. For chromosome number determination, samples from the root tips of the same five plants were counted from each of the eight varieties and four species.

DNA analysis

DNA isolation

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

AFLP procedure and data scoring

AFLP was performed according to Vos et al. (1995) using the AFLP® Analysis

Fig. 3.1. Flowchart of the steps undertaken in the data acquisition and analyses of the *Echinacea* study.

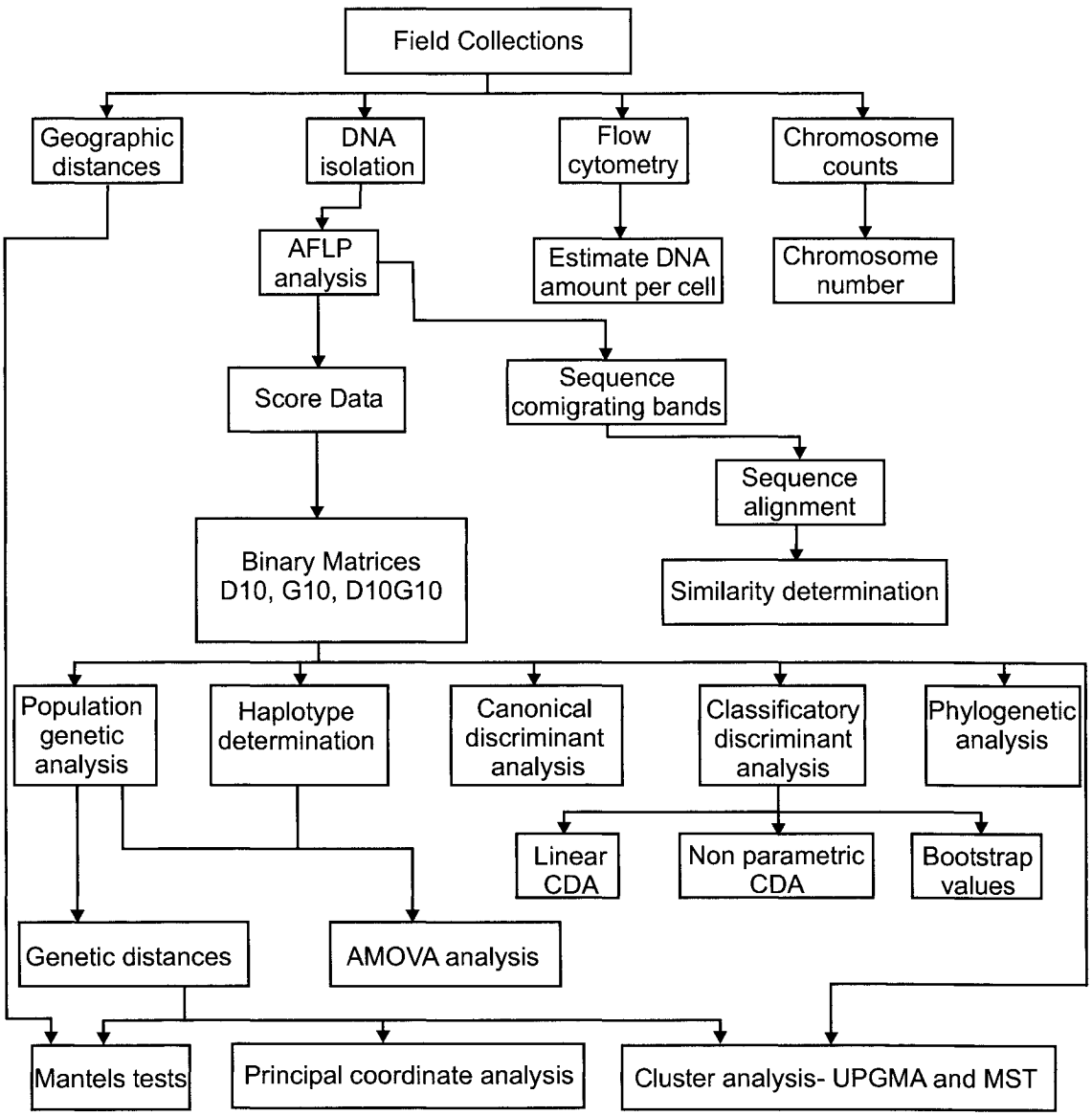

Field Collections
Geographic distances
DNA isolation
Flow cytometry
Chromosome counts
AFLP analysis
Estimate DNA amount per cell
Chromosome number
Score Data
Sequence comigrating bands
Sequence alignment
Binary Matrices D10, G10, D10G10
Similarity determination
Population genetic analysis
Haplotype determination
Canonical discriminant analysis
Classificatory discriminant analysis
Phylogenetic analysis
Linear CDA
Non parametric CDA
Bootstrap values
Genetic distances
AMOVA analysis
Mantels tests
Principal coordinate analysis
Cluster analysis- UPGMA and MST

Table 3.1. *Echinacea* species and varieties used in the present study, their accession numbers, and their collection sites.

Species	Variety	Population label	Lat. (N)	Long. (N)	State, county, and locality
Echinacea atrorubens	*atrorubens*	EAT 98-0711-1	38° 49. 1'	95° 21.47'	Kansas, Douglass Co., N of Warden on Co. Rd. 480 near cemetery
		EAT 98-0528-1	33° 59.51'	96° 17.11'	Oklahoma, Bryan Co., E of Durham on side of Hwy 70
		EAT 23881	33° 59.51'	96° 17.11'	Oklahoma, Bryan Co., near Durant, road side
		EAT 23883	34° 29.1'	96° 57.46'	Oklahoma, Murray Co., near Sulphur, two colonies,
		EAT 23887	34° 29.20'	96° 59.04'	Oklahoma, Murray Co., near Sulphur, two colonies,
	neglecta	EPN 98-0529-2	34° 19.02'	96°.52.38'	Oklahoma, Johnston Co. SW of Mill Creek on Jewel Sikes Rd.
		EPN 98-0529-1	33° 59.51'	96° 40.03'	Oklahoma, Marshall Co., 3.5 mi (1mile = 1.6 km) E of Kingston on hwy 70 at junction with side road.
		EPN 98-0530-1	34° 29.20'	96° 59.04'	Oklahoma, Murray Co. on slope overlooking veterans Lake of Chickasaw National Rec. area.

Species	Variety	Population label	Lat. (N)	Long. (N)	State, county, and locality
		EPN 23886	34° 24.28'	97° 8.32'	Oklahoma, Murray Co., near Sulphur Hillside overlooking Veterans Lake
		EPN 005	34°.19.02'	96°.52.38'	Oklahoma, Johnston Co., SW of Mill Creek on Jewel Sikes Rd
	paradoxa	EPP 98-0606-2	38°7.08'	92° 32.45'	Missouri, Camden Co., in Lake of the Ozarks State park. SE of junction 42 and 134
		EPP 98-0605-2	38° 18.45'	93° 22.46'	Missouri, Co., S of Lincoln off Hwy 65 on Road T
		EPP 0823	38° 12.09'	93° 15.21'	Missouri, Benton Co., Road off Hwy 65 just S of Lincoln
		EPP 99-0606-1	38° 18.45'	92° 32.45'	Missouri, Benton Co., off hwy 7 on road PP. South of Sedalia
		EPP 99-0606-2	38° 7.08'	92° 32.05'	Missouri, Camden Co. in Lake of the Ozarks State park. SE of Jcts. 42 and 134

Species	Variety	Population label	Lat. (N)	Long. (N)	State, county, and locality
Echinacea laevigata		EL 98-0617-1	37° 01.07'	79° 53.47'	Virginia, Franklin Co., Hwy 919 just N of Rocky Mountain. Off State Hwy 220 business
		EL 23969	36° 55.93'	79° 52.85'	North Carolina, Granville Co., near Butner, power line right of road
		EL 23964	36° 55.94'	79° 52.85'	North Carolina, Granville Co., near Butner, power line right of way and open woods
Echinacea pallida	*angustifolia*	ES-3	41° 22.00'	99° 20.03'	Arizona, Baxter Co., 3 mi S of Mountain Home on Hwy
		EA COM1	36° 23.32'	97° 14.25'	Canada, B.C, Corner Copia Farms.
		EACOM1	36° 23.32'	97° 14.25'	Canada, B.C, Corner Copia Farms.
		EA 98-0712-1	39° 15.35'	96° 34.39'	Kansas, Pottawatomie Co., near Manhattan. Overlooking Tuttle Creek lake Dam.
		EA 98-0717-1	47° 35.03'	101° 25.12'	North Dakota, McClean Co. Fort Stevenson State park, next to Fort Berthold Reservation

Species	Variety	Population label	Lat. (N)	Long. (N)	State, county, and locality
		EA 98-0716-2	47° 12.33'	101° 00.13'	North Dakota, Oliver Co., Cross Ranch State Park. Prairie roadside up slope to preserve behind fence
		EA 24058	39° 15.35'	96° 34.39'	Kansas, Pottawatomie Co., near Manhattan, Native mixed prairies
	pallida	EPA 98-0604-1	37° 00.10'	93° 09.22'	Missouri, Christian Co., S of Ozark, on Hwy 14, roadside.
	pallida	EPA 98-0604-1	37° 00.10'	93° 09.22'	Missouri, Christian Co., S of Ozark, on Hwy 14, roadside.
		EPA 98-0531-1	37° 00.15'	93° 10.15'	Oklahoma, Hughes Co., about 3miles SW of Dustin on Hwy. 9, roadside next to cow field.
		EPA009	36° 49.98'	92° 36.24'	Missouri, Douglas Co., 1 mile E on Hwy. N from Hwy. 5 S from Ava
		EPA018	39° 44.04'	90° 13.68'	Illinois, Wildlife Prairie Park, in natural tall grass prairie region

Species	Variety	Population label	Lat. (N)	Long. (N)	State, county, and locality
		ES017	36° 17.34'	92° 37.02'	Arizona, Marion Co., 1 mile W of Flopping Hwy. 412 S side
	sanguinea	ESA 98-0523-1	30°.32.20'	95° 26.04'	Texas, Cass Co., Wright Patnam Lake Park Rd. 1 mile from Jct FM 2148 on both sides of road
		ES 23962	34° 15.80'	85° 12.09'	Georgia, Floyd Co., near cave springs, glade in pine forest
		ESA 98-0523-2	33° 02.46'	94° 20.63'	Texas, Cass Co., On Hwy. 77 S of Atlanta
		ESA 23873	34° 16.04'	85° 12.09'	Louisiana, Beauregard Parish, near fields, sloping road bank.
		ESA 23874	30° 39.02'	93° 20.43'	Louisiana, Vernon Parish, near Leesville, Kistachie National Forest
		ESA 23878	34°.26.04'	85° 12.09'	Louisiana, Bienville Parish, Near Lucky. Road side at edge
		EST 98-0530-2	36° 02.48'	97° 14.34'	Oklahoma, Carter Co. 1 mile south of Fox.

Species	Variety	Population label	Lat. (N)	Long. (N)	State, county, and locality
		EST 98-0602-1	36° 43.88'	98° 19.38'	Oklahoma, Payne Co. SW of Still water off Hwy. 51 on Meridian Rd
		EST 0823-2	38° 32.03'	96° 03.33'	Oklahoma, West of Alma on county rd. S of State road 7
		EST 99-0608-2	36° 02.48'	97° 14.34'	Kansas, Lyon Co. E of Lyon State lake. Hwy. 170
	simulata	ES011	36° 28.98'	92° 19.92'	Arizona, Baxter Co., 3 miles S of Mountain Home on Hwy.
		ES 98-0608-1	36° 55.14'	90° 57.46'	Missouri, Carter Co., near VanBuren in Ozark National Forest. Hwy. Z on limestone roadside ridge
		ES 98-0608-2	39° 44.04'	90° 13.68'	Arizona, Fulton Co., just S of MO border near Mammoth Spring Hwy. 63
	tennesseensis	ET 98-0611-1	36° 02.06'	86° 21.05'	Tennessee, Wilson Co., 1 mile from Vesta Rd. off State Hwy. 231

Species	Variety	Population label	Lat. (N)	Long. (N)	State, county, and locality
		ET 98-0611-2	36° 04.08'	86° 34.42'	Tennessee, Wilson Co., on Mulberry Rd. Private Land protected by Department of Conservation
		EPC0M1	45° 52.62'	120° 47.37'	Oregon, Horizon Herbs
Echinacea purpurea		EP004	34.19.02'	96° 50.22'	Arizona, Hwy 412 (S side) E of Siloam Springs near OK
		EP010	37°.05.05'	91° 15.10'	Missouri, Shannon Co., Hwy. 108 Bay Creek
		EP019	40°.24.58'	91°.04.41'	Illinois, Near Carthage on Hwy. 136, private front lawn N
		EP 98-0616-1	36° 10.43'	78° 54.17'	North Carolina, Durham Co., N of Durham city on Roxboro rd (Hwy. 301)
		EP 99-0616-1	36° 10.43'	78° 54.17'	North Carolina, Durham Co., N of Durham city on Roxboro rd (Hwy. 301)
		ASGL-97	45° 52.38'	120° 47.346'	Washington, Trout Lake Farms, Trout Lake

Species	Variety	Population label	Lat. (N)	Long. (N)	State, county, and locality
		ASG-135	45° 52.44'	120° 47.352'	Washington, Trout Lake Farms, Trout Lake
		ASG-001	45° 52.50'	120° 47.358'	Washington, Trout Lake Farms, Trout Lake
		AS-201	45° 52.56'	120° 47.364'	Washington, Trout Lake Farms, Trout Lake
		EP COM1	45° 52,62'	120°.47.370'	Oregon, Horizon Herbs
Hybrid	*pallida* var. *angustifolia* X *atrorubens*	EA014	37° 18.78'	96° 44.70'	Kansas, Cowley Co, 0.8 mi S of Burden on Hwy. 166

Note: The classification (Binns et al. 2002a) is the revised classification of McGregor (1968).

System-I (Invitrogen Life technologies, Calif., USA), with some minor modifications as described in Baum et al. (2001). Briefly, 500 ng DNA was digested with *Eco*RI and *Mse*I, and 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., USA). The pre-amplification reactions were then diluted 1:50 with TE (Tris 10mmol/L, EDTA 1mmol/L, pH 8.0). Following a preliminary study that used 10 *Eco*RI and 10 *Mse*I primers from the kit on a few selected accessions, the following two primer sets, each with three selective nucleotides (shown in bold), were used to evaluate all the accessions:

set 1, D10 Eco-GACTGCGTACCAATTC**ACC**/Mse-GATGAGTCCTGAGTAA**CCG**,

and set 2, G10 Eco- CTGCGTACCAATTC**AGC**/Mse-GATGAGTCCTGAGTAA**CCG**.

These primers were selected based on amplification results to give an optimum number of bands (80-100 bands per gel) and the maximum number of polymorphisms detected per primer. Each *Eco*RI primer was end-labelled with ^{33}P using T4 polynucleotide kinase (Invitrogen), and selective amplification was carried out with three selective nucleotides on both the *Eco*RI and *Mse*I primer 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, annealing temperature was lowered 0.7 °C per cycle 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 gels, with the AFLP band size marker, 30-330 bp AFLP®DNA Ladder (Invitrogen), or a 100 bp DNA Ladder (Promega, Wis., USA) in one lane and a representative sample in

another lane (see Fig. 3.2). The gels were dried and autoradiographed using Kodak diagnostic X-OMAT film (Eastman Kodak, Rochester, N.Y., USA) at room temperature for 2-3 d. The representative samples were used to align different gels for band orientation, comparison, and scoring.

The autoradiograms were visually scored for the presence (1) or absence (0) of bands ranging in size from 50 to 400 bp for individual samples. A binary data matrix was compiled for each primer set.

Cloning and sequencing

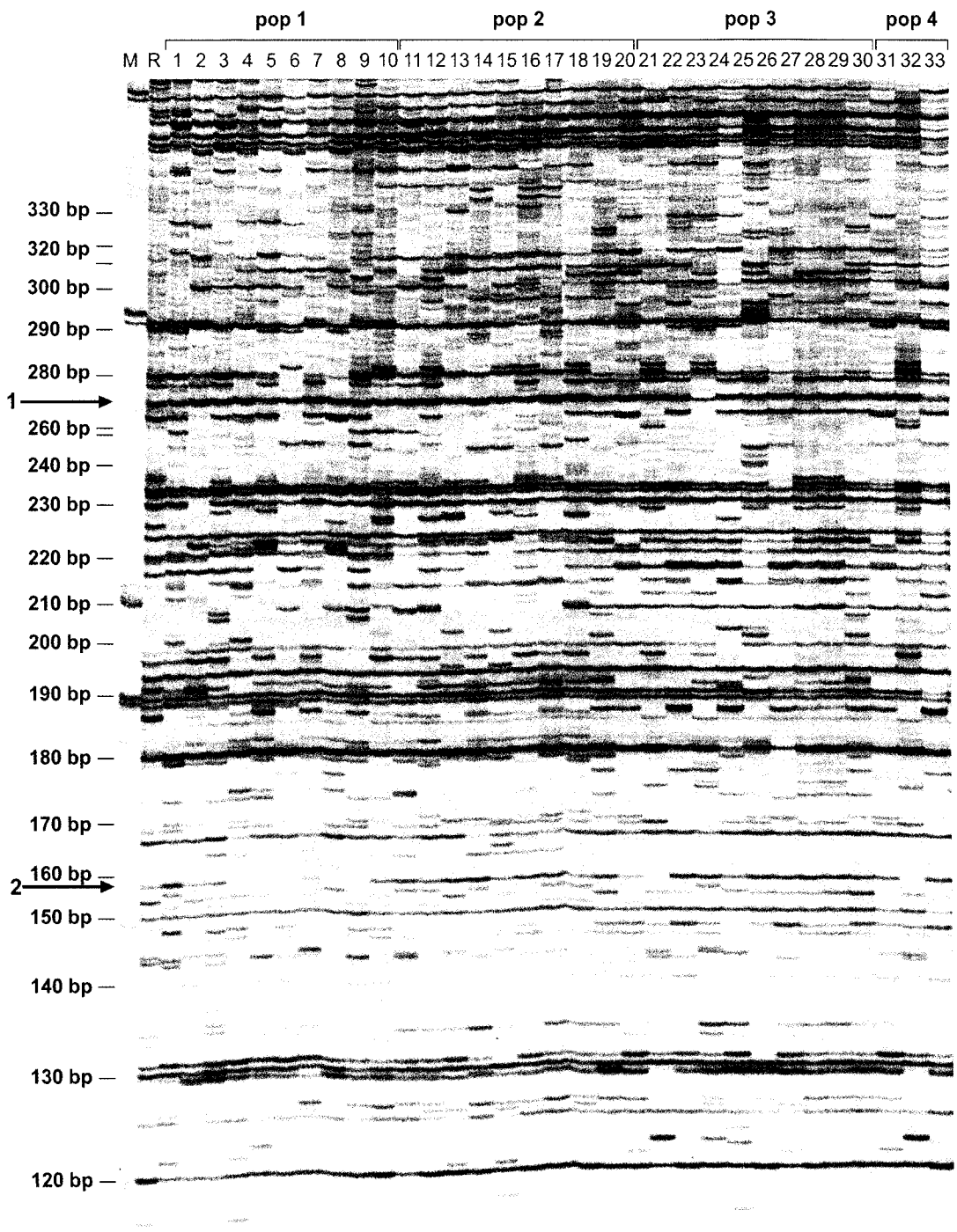
Two to six samples were used for each species and variety for a repeat of the AFLP reactions. Instead of using ^{33}P-labelled primer, the gels were electrophoresed and stained using the silver sequence DNA sequencing system (Promega). The bands, a monomorphic band and a polymorphic band, were chosen at random, excised from the gels, and eluted by the crush and soak method (Sambrook et al. 1989). The purified bands were cloned using TOPO TA cloning® kit for Sequencing (Invitrogen). Plasmid DNA was isolated using the QIAprep Miniprep (Qiagen) plasmid kit, and the purified DNA was re-suspended in water. The inserts from two to five plasmids per sample were sequenced using the ABI PRISM Dye terminator Cycle kit and run on a ABI PRISM 377 DNA sequencer (Perkin Elmer-Applied Biosystems, Foster City, Calif., USA).

Cytology

Flow Cytometry

To estimate the ploidy level, samples of the four species and eight varieties (Table 3. 4) were subjected to flow cytometry. Samples, most frequently leaf tissue, were prepared using the procedure recommended by the manufacturer of the flow cytometer,

Fig. 3.2. AFLP gel autoradiogram of four populations of *Echinacea*, using the *Eco*RI+ACC and *Mse*I +CGC primer combinations. Arrow 1 points to a monomorphic band and arrow 2 points to a polymorphic band from which the bands were excised. Lane M: 100 bp molecular weight marker; lane R: reference individual used in all the gels.


pop 1
pop 2
pop 3
pop 4
M R 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33
330 bp
320 bp
300 bp
290 bp
280 bp
1
260 bp
240 bp
230 bp
220 bp
210 bp
200 bp
190 bp
180 bp
170 bp
160 bp
2
150 bp
140 bp
130 bp
120 bp

Ploidy analyser (PA, Partec GmbH, Munster, Germany). Approximately 0.5-cm^2 leaf or other tissue material was chopped with a sharp razor blade in 0.5 mL extraction buffer (combined extraction/staining buffer of CyStain UV solution A, 15 mmol/L Hepes, 1 mmol/L EDTA, 80 mmol/L KCl, 20 mmol/L NaCl, 0.5 mmol/L spermine, 300 mmol/L sucrose, 0.2% Triton X-100; Partec, GmbH, Munster, Germany) in a plastic Petri dish. The suspension containing nuclei was filtered through 30-μm CellTrics® disposable filter (Partec) directly into the sample tube and incubated for 0.5 to 3 min, then twice the amount of staining buffer was added (CyStain UV solution B, 15 mmol/L DTE and 2 mg/L DAPI; Partec), and the sample was analysed in the Partec flow cytometer (Partec) with UV excitation. For daily optimal performance and to achieve the best system alignment of the flow cytometer, Calibration Beads UV (Partec) were used. Readings were reported as average DNA amounts from up to 10,000 cells read as counts.

Chromosome counts

Echinacea is mainly diploid (2n=22) with a few tetraploid (2n=44) entities and some triploid (2n=33) hybrid populations (McGregor 1968). To identify the putative hybrids in natural populations, chromosomes were counted for every species and variety. To maximize the number of metaphase cells, root tips were collected between 0800 and 0900 in cold water on ice. Five root tips collected from each plant were refrigerated overnight, fixed for 24 h in 95% ethanol-glacial acetic acid (3:1), and washed three times in 70% ethanol for 1 h each. The root tips were stained with Snow's stain (Snow 1963) for 2-3 d and washed 3 times in 70% ethanol for 1 h each, until the chromosomes were counted. Root tips of about 1-2 mm were squashed in 45 % glacial acetic acid with a drop

of glycerine and viewed under 1000× magnification with a Zeiss phase contrast microscope (Zeiss, Oberkochen, Germany). For each sample, five or more cells were counted.

Data analysis

Most of the following analyses were carried out on the data matrix for each individual primer set and on the adjoined matrices. A decision was made to stop analyzing more data once all the individuals proved distinct in their banding phenotypes.

Population genetic analysis

To estimate genetic variation within and among the natural populations of *Echinacea*, population genetic analysis was carried out using POPGENE (Yeh and Boyle 1997) Version 1.32 (2000). The descriptive statistics obtained, including allele frequency, gene diversity, genetic distance and F- statistics, were computed and compared with the results obtained with TFPGA (tools for population genetic analysis, Miller 1997) to cross-check with results obtained from POPGENE. To assess the relationship between species and varieties based on genetic distances obtained, cluster analysis was performed by UPGMA (unweighted pair group method with arithmetic mean, Sokal and Michener 1958), and dendrograms were constructed using NTSYS-pc (Rohlf 2000) Version 2.02 (2002).

To determine the haplotypes in each data matrix from each primer set, the simple matching coefficient (Sokal and Michener 1958) was calculated first and the resulting matrix then subjected to a cluster analysis, using NTSYSpc. The haplotypes that differ were retained in a matrix form for each primer set for analysis of molecular variance (AMOVA, Excoffier et al.1992). AMOVA was used also to obtain some estimates of

genetic structure indices. A three level hierarchy of population groups, local populations, varieties, and species, was used in AMOVA. Hierarchical analysis of variance partitioned the total variance into covariance components due to intraindividual, interindividual and inter-population differences, as detailed in Schneider et al. (2001). The purpose of the AMOVA analysis was to determine the genetic structure for various groups in the hierarchy as defined. Although AFLP is a dominant marker system, at least the way in which it was used here, it does not need to be disqualified for AMOVA as explained by Huff et al. (1993) and Peakall et al. (1995), as long as it is understood that it is the proportion of the AFLP phenotypic variance at different hierarchical levels that is calculated when the AFLP profile of each individual is treated as a haplotype. Thus, a number of indices, such as linkage disequilibrium, were not calculated since they are relevant only for co dominant markers. A minimum spanning tree (MST) was also computed among all the haplotypes. AMOVA was performed using Arlequin version 2.001 (Schneider et al. 2001).

Genetic distance and geographic distance between populations

Genetic distances between populations, with and without the putative hybrid population, were obtained as mentioned previously using POPGENE, based on Nei's genetic distance (1972). This was done separately for each primer set and the adjoined sets, and separately for the species and the varieties with and without the putative hybrid population. Geographic distances between populations were calculated with the program INVERSE (a DOS program, obtained from Victor Fraenckel, Windsway Company, 1997, SCOTIA, N.Y., USA) using the longitude and latitude of the population as input.

Principal coordinate analysis

To study and portray the genetic relationships among the species, varieties, and population in phenetic space, a principal coordinate analysis (PCO) was performed on each genetic distance matrix, using NTSYS-pc. For each case, a MST was calculated from the same distance matrix and then superimposed on the three dimensional PCO plot.

Mantel test

A Mantel test was used to test the null hypothesis that there is no association between each pair of distance data matrices of the following relationships: genetic distance obtained from each primer set with the combined primer sets, between genetic distances and between the populations and their geographic distances. The correlations and normalised Mantel Z statistic (Mantel 1967) were calculated with NTSYS-pc.

Canonical and classificatory discriminant analysis

Group determination, (i.e. classes) was preset according to morphological identification of the specimens (Binns et al. 2002a). The analyses were carried out for both the species and the ensemble of varieties. The AFLP phenotypes were expressed as binary variables with 0 or 1 value. Binary variables need not disqualify the various discriminant analyses (Rao 1952; Cochrane and Hopkins 1961; Gilbert 1968; Krzanoswski 1975). Canonical discriminant analysis (Kshirsagar 1972), an ordination method, was used to assess if there is justification to support first, the species classification, and second, the variety classification, based on morphology (Binns et al 2002a). Canonical discriminant analysis (in this case using AFLP variables) assumes prior knowledge of the grouping of the individuals and does not use the variables (in this case morphological variables) by which those individuals were grouped for the analysis.

Classificatory discriminant analysis was performed to obtain the appropriate discriminant functions to be used to identify an unknown sample based on its AFLP phenotype using a reduced set of band patterns. Two approaches were taken, a conventional (Anderson 1984; Kshirsagar 1972), or parametric, and a non-parametric method using the k-nearest neighbor method (Cover and Hart 1967), to generate a density estimate for each group and a classification function. For both methods, the same Mahalanobis distances based on the pooled covariance matrix were used, in which variables with zero variance that is, invariant variables, were deleted.

Canonical discriminant analysis

Based on the pooled covariance matrix, squared distances between class means were computed. The test of the distances between these class means, i.e. the Mahalanobis distances, were ignored, since the variables are binary in our case and therefore do not comply with the assumption of normality. Therefore, the analyses were used as a guide, without relying on the various test statistics. To find the nature of the group differences, the within canonical structure was examined, and the variables most highly associated with each of the first three axes were summarized. Plots of the first three canonical variables were made for visual examination to determine the degree of overlap between the classes, that is, the species and the varieties.

Classificatory discriminant analysis

Ten observations of the 435 specimens were haphazardly selected from the main data matrix and assembled into a separate matrix as test data, with their labels retained. Univariate test statistics from the main data matrix, especially the *F*-values and their associated probability levels for one way analysis of variance, were used to order the

variables (fragments) in descending order of importance (relative weights) in the classification. With this information, a number of consecutive runs were made including the test data, first removing one at a time first variables with non-significant F-values, then the least significant variables, until the classification results that included those of the test data became unacceptable, which is 99% correct classification based on the classification criterion from the main data matrix. Cross validation (Lachenbruch and Mickey 1968) was also available as part of the SAS program and used to assess how the discriminant function classifies the observation that it excludes (as an alternative to the test data) from the main data matrix. In both the parametric and the nonparametric (k-nearest neighbour) analyses the pooled within-group covariance matrix was used to calculate the squared distances. In the non parametric analysis, k was set to two extremes, 2 and 100 in the case of species, 2 and 60 in the case of varieties, after computing the k intervals between 2 and 100 for both species and varieties.

To obtain the confidence interval of the correct classifications for both the parametric and nonparametric analyses, 100 bootstrap runs were carried out and the mean correct classification and its standard deviation were summarized. Repeated 100 bootstrapped runs were performed by deleting the least important variables, one at a time, for each of the following runs starting from the initial 100% correct classification results, down until the 90% correct classification was reached.

All computations were done using SAS (SAS Institute Inc. 2002) running on Windows 2000. For bootstrapped runs, a SAS macro, obtained from Nancy Andes and presented at the Classification Society of North America Annual Meeting, 14-16 June

1986, was used and modified to accommodate the non-parametric analysis (available from the authors).

Sequence alignment

To determine if the comigrating bands were identical across species and varieties, the sequences of a monomorphic band and those of a polymorphic band were aligned using CLUSTAL W (Thompson et al. 1994) depicted with GENEDOC (Nicholas and Nicholas 1997) and their degrees of similarities were assessed. All the sequences for each population, variety, and species separately, were summarized by setting the percentage similarity threshold at 100%, 95% and 90% levels. In addition, to determine the similarity of the sequences with an already-sequenced gene in Genbank® (National Centre for Biotechnology Information, NCBI), the Basic Local Alignment Search Tool (BLAST, Altschul et al.1990) was used.

Phylogenetic analysis

Phylogenetic analysis was carried out using PAUP* Version 4.0 (Swofford 1999) on the 435 samples to group the populations based on the AFLP phenotypes. A haphazardly selected individual from among the *E. purpurea* individuals belonging to the functional in-group was used as the out-group for rooting. This choice of out-group was based on the results of Binns et al. (2002a) that *E. purpurea* was the sister group of all the species of *Echinacea*. A neighbor joining (NJ) tree was generated, and the distance criterion was set to search for the NJ trees with 1000 bootstraps repeats, followed by the computation of the strict consensus majority rule tree from them. Subsequently, the parsimony criterion was set to carry out a tree analysis and description of the tree. The resulting tree was condensed from 435 individuals into blocks. The tree analysis was used

to identify synapomorphies of the blocks, with the condition that these blocks would not contain the same variables in the synapomorphies with reversals.

Results

AFLP procedure and data scoring

A total of 124 bands were scored, with 62 bands for the D10 primer set and 62 bands for G10 primer set (a representative autoradiogram is provided in Fig. 3.2). Some of the samples did not produce any bands or produced only weak bands for either primer set. These were assigned question marks during data scoring. Subsequently, most samples with question marks were omitted from the analyses. Thus the number of populations was reduced from 58 to 52 and the original 577 samples were reduced to the 435 samples that were used in this study (Table 3.1). Three data matrices were prepared for the analyses: (1) the matrix resulting from the D10 primer set, (2) the matrix of the G10 primer set, and (3) the adjoined D10G10 matrix. The total number of scored values (loci) in the adjoined D10G10 matrix was 53 940, of which 40 455 were polymorphic (75.5%) and 13 485 were monomorphic (24.5%). Comparisons between the matrices D10, G10 and the adjoined D10G10, are reported under Mantel tests section.

Cloning and sequencing

A monomorphic band of 273 bp and a polymorphic band of 159 bp (Fig. 3.2) were excised and cloned from three samples of each species and variety. For the monomorphic band, a total of 79 sequences were recovered and aligned. For the polymorphic band, a total of 48 sequences were aligned. The resulting alignment statistics (Table 3.2) show that the monomorphic band is not completely identical for all the comigrating bands. At the threshold of 100% sequence similarity, the lowest value of

identical bands was found in *E. pallida* var. *angustifolia* from among the varieties and still lower from within species (Table 3.2). Percent identity increased with decreasing threshold down to 90% (Table 3.2).

In the case of the polymorphic band the comigrating bands were found to be much more dissimilar (Table 3.2). The lowest value within variety was obtained in *E. atrorubens* var. *paradoxa* (25.46%) and within species; the lowest value was obtained in *E. pallida* (23.66%) at the 100% threshold similarity. In both monomorphic and polymorphic cases the comigrating band similarities barely increased with lowering threshold down to 90%.

When calculated for all the species together, comigrating bands in the monomorphic case was 58% similar at 100% threshold similarity, but only 24% in the polymorphic bands. Lowering the threshold similarity to 90% resulted in a substantial increase of similarity for the monomorphic bands to 90%, whereas it barely increased among the comigrating polymorphic bands (Table 3.2). Similarity searches of the different sequences obtained did not result in significant matches with any known sequences in the GenBank database. More details on the results and nature of the comigrating bands are presented in a separate paper (Mechanda et al. 2004a).

Cytology

Flow cytometry

Flow cytometric analysis of the *Echinacea* species and varieties did not corroborate the expected chromosome number according to McGregor (1968) (Table 3.3). The putative triploids and tetraploids produced a relative amount of DNA content

measure of 81.22 counts compared with the diploids with readings ranging from 40.20 to 71.00 counts.

Chromosome counts

Our chromosome counts revealed only one triploid (2n=33) and one tetraploid (2n=44) in a plant identified as *E. pallida* var. *pallida*. Plants from all the other species and varieties were diploid 2n=22 (Table 3.3).

Data analysis

A first analysis of the D10 and G10 adjoined matrices indicated that each individual could be uniquely identified by its AFLP pattern or phenotype, that is, 435 individuals could be characterized by the 124 polymorphic bands. At that point we did not include more data sets for subsequent analyses.

Population genetic analysis

The gene diversity was marginally higher for all the species together at 0.4596 than for all the varieties together at 0.4506 when species that do not contain varieties were included with the varieties (Table 3.4). Among the four species, *E. purpurea* had the highest value at 0.4627 and *E. pallida* had the lowest at 0.307, whereas the remaining two species had values close to that of *E. pallida*. The three varieties of *E. atrorubens* had similarly lower rates of gene diversity. The varieties of *E. pallida* had higher values than the varieties of *E. atrorubens*, except for *E. pallida* var. *tennesseensis* which had the lowest level at 0.1815.

The number of polymorphic loci was generally higher in the species compared to the varieties within species. The species ranged from 106 to 123 whereas the varieties ranged from 54 to 113 polymorphic loci. Percent polymorphic loci were highest in

Table 3.2. Sequence identity (%) within the monomorphic and the polymorphic bands in *Echinacea*, within categories at various threshold levels (100%, 95%, and 90% similarity thresholds) in their sequence alignment.

Category	Monomorphic			Polymorphic		
	100%	95%	90%	100%	95%	90%
E. atrorubens var. *atrorubens*	91.24	95.98	95.98	30.44	33.33	33.33
E. atrorubens var. *neglecta*	89.74	94.50	94.50	33.33	33.33	33.33
E. atrorubens var. *paradoxa*	91.57	97.43	97.43	32.33	33.33	33.33
E. atrorubens	80.65	94.89	96.71	25.46	26.45	28.95
E. laevigata	98.90	98.90	98.90	31.44	33.33	33.33
E. pallida var. *angustifolia*	82.78	87.91	91.21	54.03	54.03	54.03
E. pallida var. *simulata*	88.64	88.64	88.64	42.13	42.13	42.13
E. pallida var. *tennesseensis*	94.87	94.87	94.87	36.0	36.0	36.0
E. pallida var. *pallida*	92.30	92.30	92.30	33.33	33.33	33.33
E. pallida	75.82	87.17	89.01	23.66	24.84	26.70
E. purpurea	79.48	91.57	91.57	45.00	45.00	45.00
All species	58.02	87.22	90.14	24.22	25.46	26.71

E. purpurea with 99.19% and lowest in *E. laevigata*, with 85.48% (Table 3.4). The two varieties, *E. pallida* var *tennesseensis* and *E. atrorubens* var. *paradoxa*, exhibited the lowest percent polymorphic loci. Similar gene diversity and percent polymorphic loci were found in the D10 data set alone but at a slightly higher level, whereas in the G10 data set alone, some varieties exhibited a higher level of diversity (data not shown).

For the AMOVA analysis, the number of haplotypes in the D10G10 matrix was reduced from the initial 435 to 389 after the putative hybrids and individuals with missing values were deleted. In the three-level AMOVA of four species, with 48 populations (Table 3.5), the variation was partitioned nearly equally within populations (40.53) and among populations within species (40.38). In contrast, in the three-level AMOVA of 10 taxa (i.e. the two species without varieties and the eight varieties of the other two species), and among populations within varieties (Table 3.6) the partitioned variation among taxa was slightly greater than among populations within taxa (31.02% and 27.2%, respectively), and the variation within populations was 91.8%. In other words, the variance components for populations within species were greater than that of among populations within varieties. The inclusion of the putative hybrids (the additional population, Table 3.1) did not substantially change the results (data not shown). In the AMOVA analysis of all the populations where the putative hybrid population was included within the group, the variance component among populations was 57%, greater than the 43% component within populations (Table 3.7).

Cluster analysis

UPGMA analysis using genetic distances clustered *E. atrorubens* and *E. pallida*

at the 0.11 phenon line. *E. purpurea* clustered with the former group at the 0.125 phenon line, whereas *E. laevigata* was the most genetically distant species from the group (Fig. 3.3). However, the varieties of the same species (*E. purpurea* and *E. laevigata* without varieties were included in the computation of the distances) did not cluster together (Fig 3.4).

Genetic distance and geographic distance between populations, and Mantel tests

Pair-wise geographic distances were computed only for *E. atrorubens* and *E. pallida* populations (Table 3.1), since the few populations of *E. purpurea* that were collected had originated from commercial lines (Table 3.1). Material of *E. laevigata* was insufficient for this purpose (Table 3.1). Three matrices of pairwise genetic distances were summarized for all populations, for the populations of *E. atrorubens*, and for the populations of *E. pallida* for each primer set and for the adjoined matrix of the two (not shown), each to be subjected to a Mantel test with the geographic distances. Similarly, for each primer set separately, the genetic distances between the four species, with or without the putative hybrid, and those between the 10 taxa, with or without the putative hybrid, and the adjoined matrix of the D10 and G10 primer sets, were calculated (not shown) according to the methods described.

Mantel tests between the different distance matrices (Table 3.8) led to the rejection of the hypothesis that there was no association among the different combinations of primer sets for each category in all cases except two. These two cases are the test of the D10 primer set with the G10 set in 10 taxa and the D10 set with the D10G10 set in the four species (Table 3.8). Thus, in the majority of the cases there was an association between the combinations of the data sets. This association was

accompanied by a high correlation among the data sets, with the exception of the D10 and the G10 sets of the 10 taxa with and without the putative hybrid and with all the populations. It is noteworthy that the D10 distance data with the adjoined D10G10 data were highly correlated (0.996), but they were not associated as a result of the Mantel test (Table 3.8).

Mantel tests between geographic distances and genetic distances yielded very weak correlations and in most cases indicated no association between geographic distribution and genetic distances (Table 3.9). The hypothesis that there was lack of association between the genetic distances of the D10 primer set and the geographic distances was rejected in all the populations and in all cases with *E. pallida*. In each of the four cases that suggested some association between the genetic distances and the attendant categories, the correlation was low to very low (Table 3.9).

Principal coordinate analysis

PCO analysis of the genetic distance matrix of the four species (Fig. 3.5) on the D10G10 data set pointed to *E. pallida* as a pivotal species, from which *E. purpurea* and *E. atrorubens* are nearly equidistant (Fig. 3.5, note the distortion of the rendition in three dimensions), whereas *E. laevigata* is the most distant. PCO of the varieties, including the two species lacking varieties, did not support coherence of species from the varieties (Fig. 3.6). Nonetheless, based on the MST, the two ordinations identified *E. pallida* as a pivot species, from which variation radiated to various varieties and (or) species. Finally, in the clustering and ordination of the varieties *E. atrorubens* var. *atrorubens* was isolated from the two other conspecific varieties.

Table 3.3. Ploidy analysis of *Echinacea* species and varieties by chromosome counting and flow cytometry.

McGregor's (1968) classification	Binns et al. (2002) classification	Chromosome number = 2n current work	Chromosome number = 2n (McGregor, 1968)	Flow cytometry reading[a]
E. purpurea	*E. purpurea*	22	22	54.66
E. angustifolia var. *angustifolia*	*E. pallida* var. *angustifolia*	22	22	40.20
E. pallida	*E. pallida* var. *pallida*	33,44	44	81.22
E. sanguinea	*E. pallida* var. *sanguinea*	22	22	43.66
E. tennesseensis	*E. pallida* var. *tennesseensis*	22	ND	47
E. atrorubens	*E. atrorubens* var. *atrorubens*	22	22	68.42
E. paradoxa var. *paradoxa*	*E. atrorubens* var. *paradoxa*	22	22	68.33
E. paradoxa var. *neglecta*	*E. atrorubens* var. *neglecta.*	22	22	71
E. simulata	*E. pallida* var. *simulata*	22	22	47.2
E. laevigata	*E. laevigata*	22	22	42.77
E. angustifolia var. *strigosa*	*E. pallida* var. *angustifolia*	ND	22,44	40.22

Note: Analysis based on five replicates, with five to eight cells per replicate. ND, not determined.

[a] Average DNA amounts from up to 10,000 cells read as counts. Numbers in bold are new counts.

Table 3.4. Single population descriptive statistics for the species and varieties of *Echinacea,* for the D10G10 adjoined matrix.

Species, varieties	Sample size	Nei's (1973) gene diversity, H (±SD)	No. of polymorphic loci	Percent polymorphic loci
E. atrorubens var. *atrorubens*	38	0.2484±0.1905	88	70.97
E. atrorubens var. *neglecta*	48	0.2582±0.2055	88	70.97
E. atrorubens var. *paradoxa*	41	0.2143±0.2217	62	50.00
E. atrorubens	127	0.3831±0.1305	121	97.58
E. laevigata	16	0.3733±0.1684	106	85.48
E. pallida var. *angustifolia*	52	0.4180±0.1356	115	92.74
E. pallida var. *pallida*	34	0.3766±0.1470	112	90.32
E. pallida var. *sanguinea*	67	0.3916±0.1445	113	91.13
E. pallida var. *simulata*	27	0.3517±0.1680	109	87.90
E. pallida var. *tenneesseensis*	18	0.1815±0.2161	54	43.55
E. pallida	198	0.3703±0.1535	119	95.97
E. purpurea	83	0.4627±0.0674	123	99.19
All species	423	0.4596±0.0691	123	99.19
All varieties [a]	423	0.4506±0.0750	123	99.19

[a] Includes *E. laevigata* and *E. purpurea*, which do not have varieties.

Table 3.5. Summary of analysis of molecular variance (AMOVA) for the four species of *Echinacea* and their populations (primer sets D10 and G10 AFLP data matrices were combined).

Source of variation	df	Sum of squares	Variance components	Percent of total variation
Among species	3	1372.994	4.40638*	19.09
Among populations within species	47	3770.682	9.32053*	40.38
Within populations	338	3161.291	9.35293*	40.53
Total	388	8304.967	23.07985	100.00

Note:*, Significant ($P < 0.05$) values based on 1,023 permutations. df = degrees of freedom

Table 3.6. Summary of analysis of molecular variance (AMOVA) for the 10 taxa (i.e. the two species without varieties and the eight varieties of the other two species) of *Echinacea* (data matrices of AFLP primer sets D10 and G10 were combined).

Source of variation	df	Sum of squares	Variance components	Percentage of variation
Among varieties	9	2883.414	6.94489*	31.02
Among populations within varieties	41	2260.262	6.08823*	27.20
Within populations	338	3161.291	9.35293*	41.78
Total	388	8304.967	22.38605	100.00

Note:*, Significant ($P < 0.05$) values based on 1,023 permutations.

df = degrees of freedom

Table 3.7. Summary of analysis of molecular variance (AMOVA) for all the populations of *Echinacea* in the study (data matrices of AFLP primer sets D10 and G10 were combined).

Source of variation	df	Sum of squares	Variance components	Percentage of variation
Among populations	51	5227.452	12.31057*	56.67
Within populations	342	3218.891	9.41196*	43.33
Total	393	8446.343	21.72253	100.00

Note:*, Significant ($P < 0.05$) values based on 1,023 permutations. df = degrees of freedom

Table 3.8. Summary of Mantel tests, with the genetic distance matrices obtained from the data of the primer sets (D10, G10, and D10G10).

	G10 genetic distance with geographic distance			D10 genetic distance with geographic distance			D10G10 genetic distance with geographic distance		
Category	r	t	P	r	t	P	r	t	P
All populations	-0.01754	-0.2236	0.4326	0.08159	1.9004	0.0350*	0.05761	1.0049	0.1778
Echinacea atrorubens	-0.10386	-1.0857	0.1279	-0.20380	-1.9046	0.0390*	-0.17603	-1.7498	0.0509
Echinacea pallida	0.25713	4.1690	0.0020**	0.22149	3.7557	0.0020**	0.26164	4.3725	0.0020**

Note: *P*, probability of random Z is ≥ observed Z *, $P < 0.05$; **, $P < 0.01$. r, correlation coefficient; t, t values.

Results were considered significant at $P < 0.05$.

Canonical discriminant analysis

Species positions on the canonical discriminant analysis plot based on the D10G10 data set was similar to that in the PCO (Fig. 3.7). The AFLP variables were arranged according to their order of "importance", that is, the magnitude of their *F*-values (Table 3.10). Generally, the D10 variables appear to have greater importance than the G10 variables in the D10G10 set; however a great number of G10 variables still affect the placement of the individual samples on the ordination plot (Fig. 3.7). An examination of the pooled within canonical structure of the D10G10 data identified the first six variables with the greatest *F*-values, all belonging to the D10 primer set (Table 3.12) with greatest weight on axis 2. Of the eight variables with greatest weight on axis 1, two belong to the G10 primer set and the rest to the D10 primer set. Wilks' λ statistic was 0.001 (P>F was <0.001), that is, the null hypothesis that the means of at least two species are equal was rejected. The canonical correlation coefficients of the three axes were 0.966, 0.961 and 0.874 respectively, indicating an overall goodness of fit of the canonical variates to their species group (Pimentel 1979: 223). When individual samples were grouped by variety, including the two species not divided into varieties, the three-dimensional plot exhibited four well-defined clusters of points (Fig. 3.8): (1) *E. atrorubens* var. *paradoxa*, (2) *E. atrorubens* var. *neglecta*, (3) *E. purpurea*, and (4) the remaining varieties, including the species *E. laevigata*. Although the results of Wilks' λ statistic was 0.0 (*P>F* was <0.001), that is, the null hypothesis that the means of at least two varieties are equal was rejected; and although the canonical correlation coefficients of the first three axes were 0.988, 0.987, and 0.967, respectively, indicating an overall goodness of fit of the canonical

Table 3.9. Summary of Mantel tests with the geographic and the genetic distance matrices obtained from the data of the primer sets (D10, G10 and D10G10).

	D10+G10			D10+D10G10			G10+D10G10		
Category	r	t	P	r	t	P	r	t	P
4 species	0.97779	1.8152	0.0300*	0.99636	1.8913	0.0519	0.99112	1.8026	0.0370*
4 species and hybrid	0.80463	1.8196	0.0300*	0.95732	2.204	0.0140*	0.94096	1.9777	0.0090**
10 "varieties"	0.39426	1.6285	0.569	0.91728	3.9151	0.0020**	0.7256	2.799	0.0030**
10 "varieties and hybrid"[a]	0.41016	1.8604	0.0410*	0.89790	4.202	0.0020**	0.76797	3.116	0.0020**
All populations	0.45307	11.2311	0.0020**	0.89899	25.3026	0.0020**	0.7936	15.2314	0.0020**

Note: *P*, probability of random Z is ≥ observed Z *, $P < 0.05$; **, $P < 0.01$. Results were considered significant at $P < 0.05$. r, correlation coefficient; t, t values. [a] Category includes "varieties", that is, the two species without varieties and the eight varieties of the other two species.

Fig. 3.3. Genetic relationships of the species of *Echinacea* based on the AFLP phenotypes. The dendrogram was constructed using the Dice coefficient of similarity (Dice 1945) and UPGMA (unweighted pair group method with arithmetic mean) clustering.

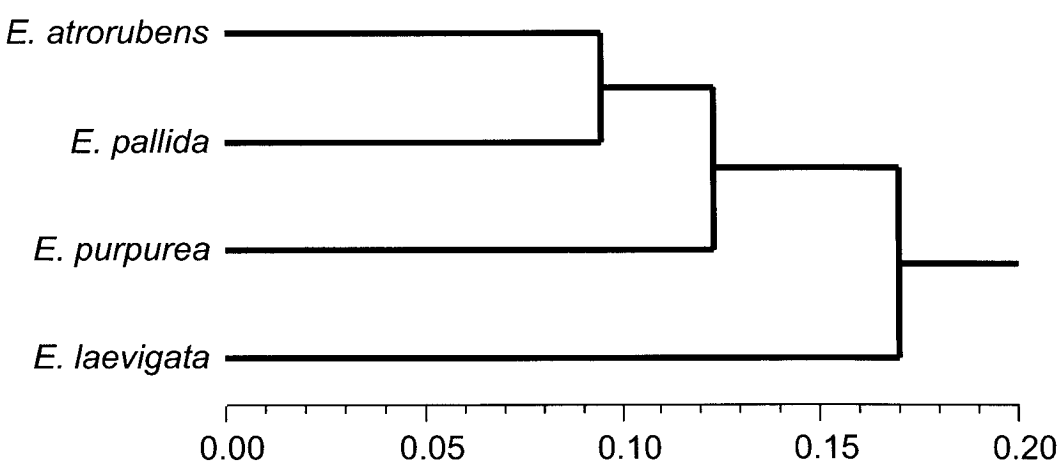

E. atrorubens
E. pallida
E. purpurea
E. laevigata
0.00
0.05
0.10
0.15
0.20

Fig. 3.4. Genetic relationships of the varieties of *Echinacea* based on the AFLP phenotypes. The dendrogram was constructed using the Dice coefficient of similarity (Dice 1945) and UPGMA (unweighted pair group method with arithmetic mean) clustering.

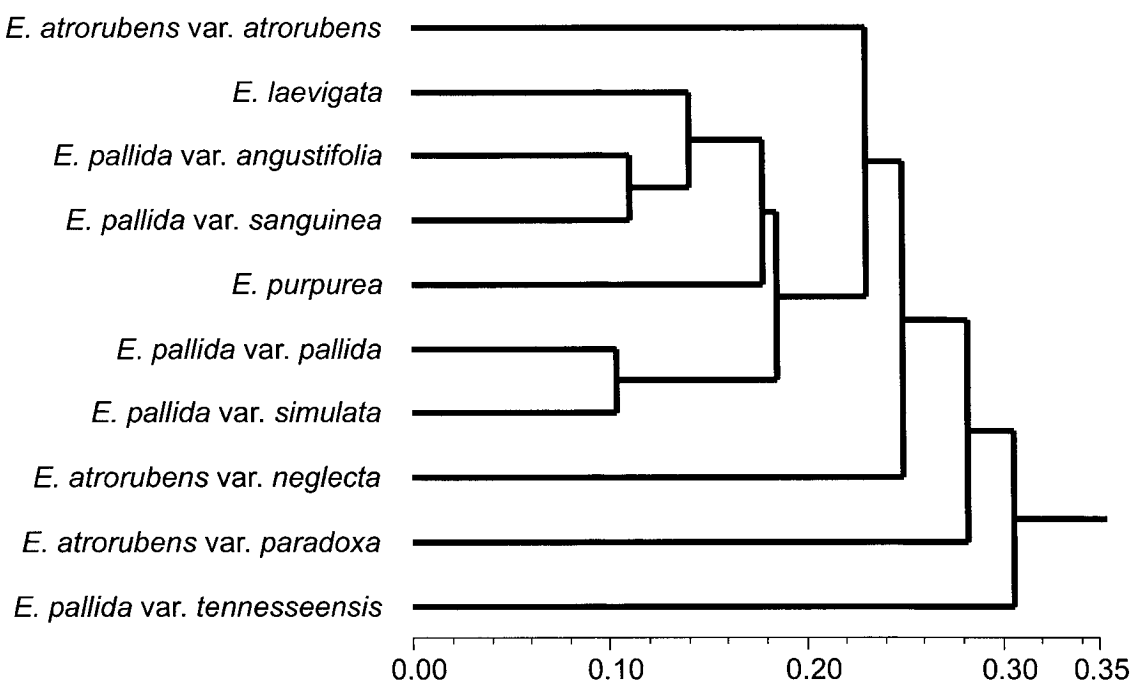

E. atrorubens var. atrorubens
E. laevigata
E. pallida var. angustifolia
E. pallida var. sanguinea
E. purpurea
E. pallida var. pallida
E. pallida var. simulata
E. atrorubens var. neglecta
E. atrorubens var. paradoxa
E. pallida var. tennesseensis
0.00
0.10
0.20
0.30
0.35

Fig. 3.5.Three dimensional plot of the principle coordinates analysis of the AFLP data of the *Echinacea* species. Variation explained by axis 1 is 69.93%, axis 2, 29.53%, and axis 3, 1.54%. MST (minimum spanning tree) superimposed, connecting the species points.

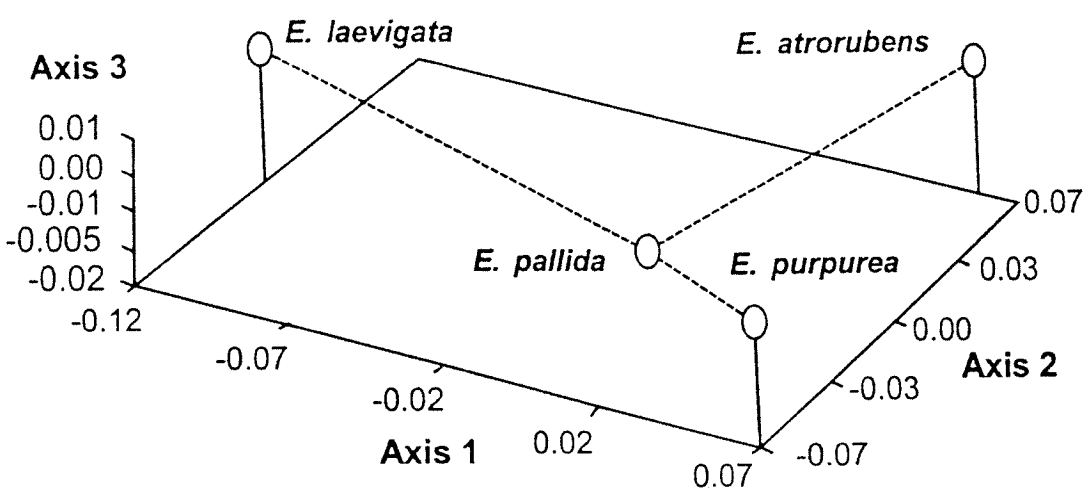

E. laevigata
E. atrorubens
Axis 3
0.01
0.00
-0.01
-0.005
-0.02
E. pallida
E. purpurea
0.07
0.03
0.00
-0.03
-0.07
Axis 2
-0.12
-0.07
-0.02
0.02
0.07
Axis 1

Fig. 3.6. Three dimensional plot of the principle coordinates analysis of the AFLP data for the *Echinacea* varieties including *Echinacea laevigata* and *Echinacea purpurea*, which do not have varieties. Variation explained by axis 1 is 42.82%; axis 2, 22.89%, and axis 3, 13.99%. MST (minimum spanning tree) superimposed, connecting the variety points.

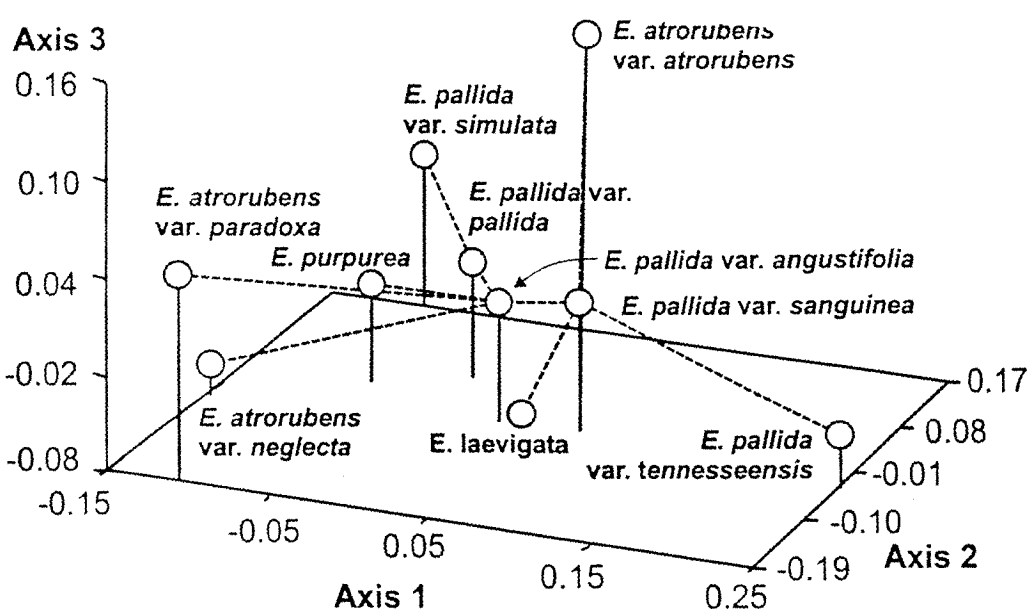

Axis 3
0.16
0.10
0.04
-0.02
-0.08
E. atrorubens var. atrorubens
E. pallida var. simulata
E. pallida var. pallida
E. atrorubens var. paradoxa
E. purpurea
E. pallida var. angustifolia
E. pallida var. sanguinea
E. atrorubens var. neglecta
E. laevigata
E. pallida var. tennesseensis
-0.15
-0.05
0.05
0.15
0.25
Axis 1
0.17
0.08
-0.01
-0.10
-0.19
Axis 2

Table 3.10. Canonical discriminant analysis of the four *Echinacea* species.

Variable	F-Value	Canonical axis 1	Canonical axis 2	Canonical axis 3
D10_144	190.10	-0.20	0.33	-0.12
D10_137	184.33	0.20	0.33	-0.10
D10_108	149.67	0.06	0.35	0.02
D10_138	116.97	0.10	0.27	-0.18
D10_132	83.81	0.13	0.22	-0.06
D10_184	80.52	-0.16	0.20	0.03
D10_131	80.15	0.24	-0.05	0.05
D10_178	69.13	0.10	0.21	0.03
G10_140	51.80	0.17	0.11	0.02
D10_119	51.16	0.19	-0.02	0.01
D10_134	51.01	0.16	0.11	-0.06
G10_097	50.58	-0.04	0.19	0.07
D10_228	49.95	0.19	0.01	-0.02
D10_179	49.37	0.05	0.19	0.04
D10_113	48.31	-0.06	0.14	-0.22
G10_069	45.55	0.17	0.01	0.10
D10_110	43.76	-0.09	0.16	0.03
D10_135	42.47	-0.01	0.18	-0.08
D10_159	42.22	0.13	0.13	0.01
D10_122	42.21	0.08	0.17	0.07
G10_248	39.31	0.16	-0.00	-0.11

Variable	F-Value	Canonical axis 1	Canonical axis 2	Canonical axis 3
D10_148	39.20	0.17	0.02	0.02
D10_123	37.87	0.16	-0.02	0.12
D10_221	36.63	0.16	0.04	0.01
D10_180	33.82	0.15	-0.03	0.08
D10_176	33.05	0.00	0.17	0.03
D10_121	31.82	0.13	0.07	0.06
D10_203	30.42	-0.10	0.10	-0.10
D10_169	27.44	0.11	0.09	0.04
D10_130	26.43	-0.13	0.02	0.07
D10_160	23.65	0.13	0.02	0.05
D10_156	23.30	0.10	0.09	0.05
D10_190	20.77	0.12	-0.02	-0.08
G10_113	20.10	-0.02	0.13	-0.01
D10_117	18.47	0.11	0.03	-0.01
G10_095	18.13	-0.06	0.10	0.02
G10_150	17.54	0.02	-0.03	0.07
D10_182	17.37	0.05	-0.09	0.10
G10_145	16.64	0.03	0.11	0.01
G10_143	13.66	0.08	0.04	0.07
D10_140	13.45	0.06	0.07	-0.09
D10_230	13.12	-0.01	-0.09	-0.08
D10_164	12.42	0.05	-0.09	0.01

Variable	F-Value	Canonical axis 1	Canonical axis 2	Canonical axis 3
D10_172	11.82	-0.00	0.10	0.03
G10_186	11.47	-0.06	0.07	-0.03
D10_216	11.33	-0.02	-0.08	0.02
D10_188	11.26	0.07	-0.02	-0.10
G10_106	11.00	-0.01	0.09	-0.00
G10_120	10.85	0.03	-0.07	0.06
G10_141	10.56	-0.07	0.04	-0.09
G10_228	10.53	-0.07	0.01	-0.07
G10_114	10.40	0.08	-0.04	0.01
D10_212	10.23	-0.03	-0.01	0.04
G10_128	10.21	0.03	0.06	-0.00
G10_188	10.16	0.08	-0.03	0.05
G10_132	10.10	-0.04	0.08	0.01
D10_170	9.35	-0.01	-0.06	0.07
D10_211	9.14	0.07	0.03	0.03
D10_152	8.74	-0.07	0.01	-0.08
G10_179	8.24	0.02	-0.06	0.02
G10_090	8.20	0.03	0.06	0.06
D10_150	8.17	-0.01	0.08	0.01
D10_115	8.15	-0.04	0.06	-0.04
G10_210	7.74	-0.04	0.02	0.08
D10_154	7.70	0.06	-0.05	0.02

Variable	F-Value	Canonical axis 1	Canonical axis 2	Canonical axis 3
G10_125	7.60	-0.02	0.04	-0.12
G10_080	7.12	-0.03	0.04	-0.10
G10_148	6.77	-0.05	0.05	0.00
G10_155	6.77	-0.04	0.06	-0.01
G10_260	6.67	-0.06	0.04	-0.02
G10_142	6.60	0.04	0.02	-0.00
G10_199	6.53	-0.02	0.01	0.09
G10_225	6.44	-0.02	0.07	-0.04
D10_205	6.21	-0.03	-0.06	0.06

Note: Variables arranged in descending order of magnitude of their *F*- values.

Pooled within canonical structure, variables of the first three axes provided.

$P > F$ was < 0.001 for all variables.

Table 3.11. Linear discriminant function based on the 10 variables with the highest *F*-values for the identification of *Echinacea* species.

Variable	Discriminant function 1	Discriminant function 2	Discriminant function 3	Discriminant function 4	EAT23881[a] (*Echinacea atrorubens*)	EA014[a] (*Echinacea atrorubens*)
Constant	-10.81245	-11.88005	-16.90104	-9.03487		
D10-144	11.70903	4.75299	9.49310	2.43188	1	1
D10-137	-2.67348	1.32700	3.88900	0.63938	0	1
D10-108	6.52589	11.43440	11.50515	2.04990	0	1
D10-138	1.75664	-1.96446	3.69608	-2.96640	1	1
D10-131	0.99365	2.85766	4.75618	10.74973	0	1
D10-184	7.87340	5.43386	1.72070	-1.03755	1	1
D10-132	-2.34357	-2.59267	-1.53488	-6.27877	0	1
D10-178	-6.47970	0.38531	-2.67320	-1.39134	0	1
D10-134	-0.52247	2.24383	2.42074	1.21010	0	1
D10-119	3.58087	4.93930	5.56465	7.69010	1	0
Tally for each discriminant function[b]						
Echinacea atrorubens	**14.11532**	1.28164	3.56821	2.91684		
Echinacea pallida	6.02694	11.99488	**16.3717**	3.62741		

Note: Discriminant function 1: *Echinacea atrorubens*; 2: *Echinacea laevigata*; 3: *Echinacea pallida*; 4: *Echinacea purpurea*. In the two right-most columns are AFLP

phenotypes of two samples derived from two populations (Table 1) taken from the test data set.

[a] 0, absence of phenotype in the population; 1, presence of phenotype in the population.

[b] Tally was calculated for all phenotypes present in the sample populations, minus the control. Highest values for the identified species are in bold.

Table 3.12. Mean and standard deviation (SD) of bootstrap results of nonparametric discriminant analyses of *Echinacea* species and varieties.

Category	K values	Data set type	Mean	SD
Species	2	Training set	99.04	0.49
	2	Test	93.36	1.93
	60	Training set	97.32	0.74
	60	Test	94.72	1.54
Varieties	2	Training set	99.37	1.08
	2	Test	91.77	2.30
	100	Training set	98.51	1.89
	100	Test	89.66	2.59

variates to their varietal group; clearly, the plot of varieties does not reflect their taxonomic disposition.

Classificatory discriminant analysis

Parametric analysis (linear model)

Using the linear discriminant function, the total correct classification for the four species was 96.67%, cross-validated by 97.22% in the training data and 92.59% in the test data of 10 individual samples. After eliminating the least important variables one at a time, these values gradually declined down to 90.51%, 92.82%, and 81.48%, respectively, at which point the analysis was stopped, leaving us with the 68 most important variables, based on the F values (not shown). The same analysis for the eight varieties yielded 99.13% total correct classification, cross-validated by 93.33% in the training data, and 93.18% for the test data. After eliminating the least important variables one at a time, these values gradually declined to 97.46%, 92.67%, and 88.63%, respectively, with the model retaining the 74 most important variables, based on their F-values (not shown). Because the linear model did not meet the criteria of 99% correct classification results of the test data (outlined in the Methods section), no bootstrapping was performed. However, the linear model can still be used for the purpose of identification, as in the example below.

Non-parametric analysis

A summary of the results of *k*-nearest neighbor analysis, subjected to bootstrap analysis, is presented in Table 3.11. The training data was above 99% mean correct classification based on 100 bootstraps. Gradually increasing the *k* parameter to extreme values, 60 in the case of species and 100 in the case of varieties, yielded 97.32% correct

classification for the species and 98.51% for the varieties. As expected, the test data always yielded slightly lower classification results. Thus, it might be possible to identify an individual *Echinacea* plant to a species or a variety using discriminant functions (data not shown).

An example of using the discriminant function based on linear model for the identification of a specimen from among the test data is provided in Table 3.12. This particular discriminant function was obtained from the 10 most important variables, that is, those with the highest *F*-values. The classification results with these 10 variables are indeed lower than those above, that is, 91.45% for *E. atrorubens*, 83.33% for *E. laevigata*, 88.36% for *E. pallida* and 98.63% for *E. purpurea*, and the overall correct classification results is 82.94% without having obtained bootstrap results.

To use any discriminant function, the user first obtains the AFLP phenotype (in this example 10 specific bands), then scores for presence/absence of bands (Table 3.12). Next, the user takes the constant, then multiplies each variable's score (in our case, 0 or 1) by the corresponding coefficient, retaining its sign (whether positive or negative), and sums all products in the column for each species. The column of the species with the highest value identifies the species. In this instance, two samples were taken from the test data set, one sample identified as *E. atrorubens* and one as a putative hybrid, both based on morphology (Table 3.1). According to the sum of each column, *E. atrorubens* was identified correctly and the putative hybrid was assigned to *E. pallida* (Table 3.12).

Phylogenetic analysis

A summary of NJ phylogram of the 435 samples is presented in Fig. 3.9. Information on the major clades is provided in Table 3.13. Clearly, the results support

Fig. 3.7. Three-dimensional canonical discriminant plot of the populations of *Echinacea* grouped into the four species from the AFLP phenotypes. Variation explained by axis 1 is 47.76%, axis 2, 41.36%, and axis 3, 10.88%.

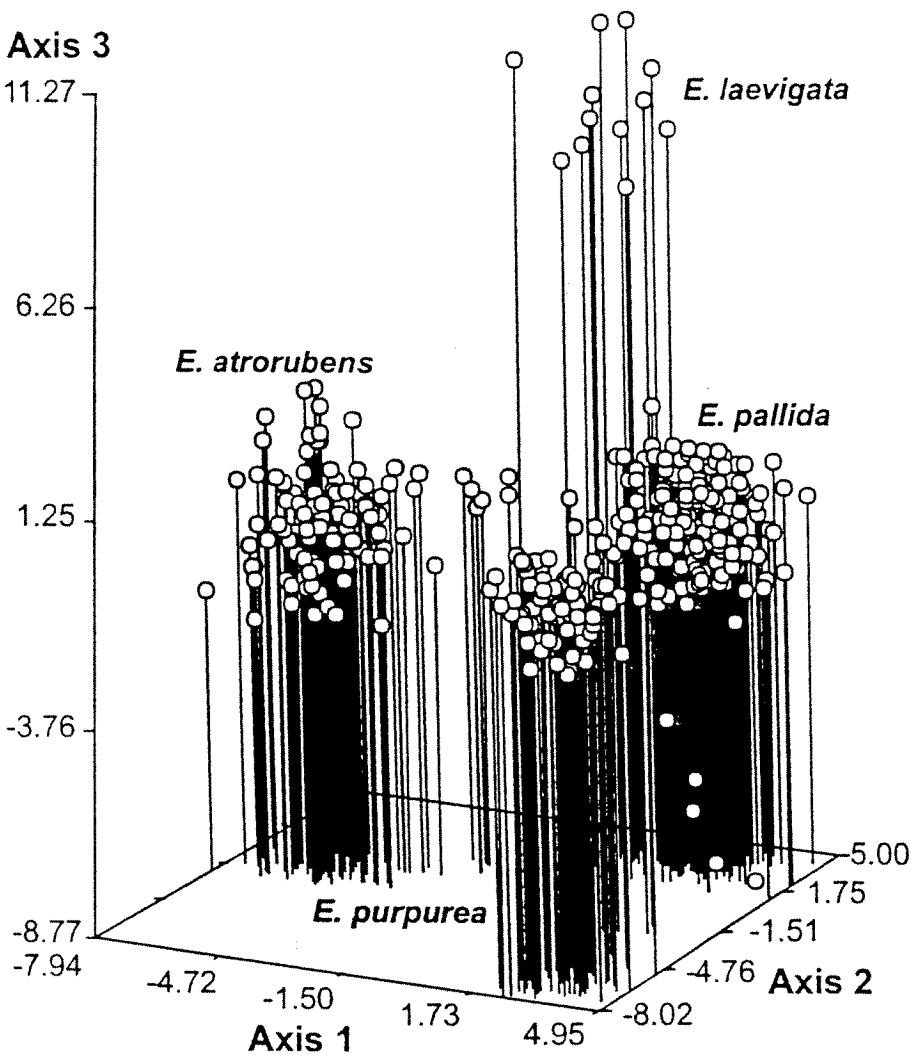
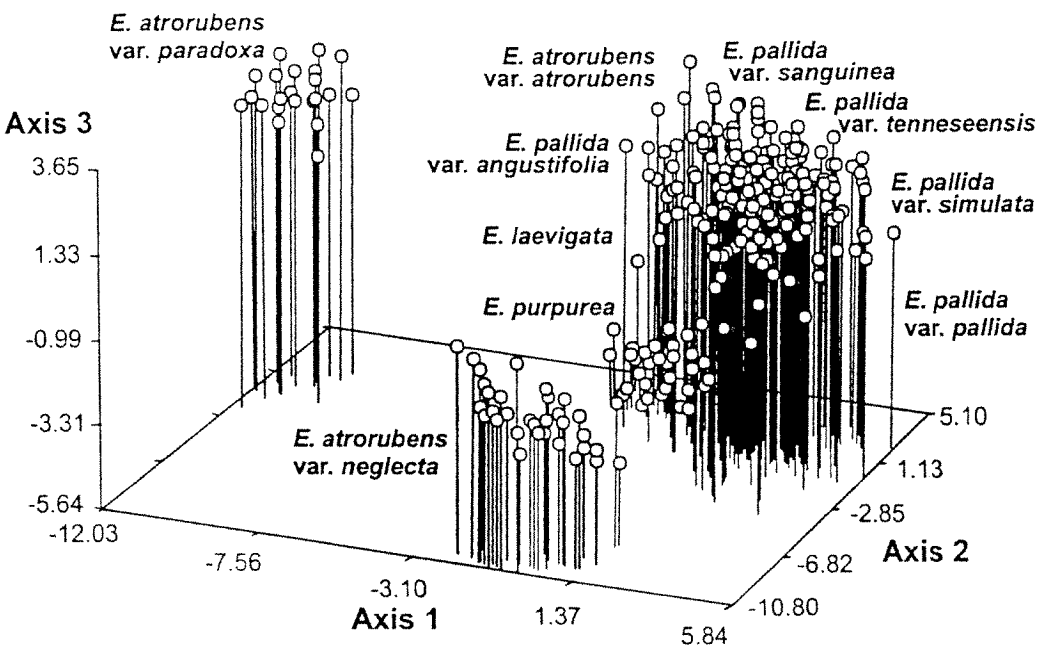

Axis 3
11.27
6.26
1.25
-3.76
-8.77
E. laevigata
E. atrorubens
E. pallida
E. purpurea
-7.94
-4.72
-1.50
1.73
4.95
Axis 1
5.00
1.75
-1.51
-4.76
-8.02
Axis 2

Fig. 3.8. Three-dimensional canonical discriminant plot of the populations of *Echinacea* grouped into varieties including *Echinacea laevigata* and *Echinacea purpurea*, which do not have varieties, based on the AFLP phenotypes. Variation explained by axis 1 is 33.39%, axis 2, 31.90%, axis 3, 11.72%. Each circle represents an individual.


E. atrorubens
var. paradoxa
E. atrorubens
var. atrorubens
E. pallida
var. sanguinea
E. pallida
var. tenneseensis
E. pallida
var. angustifolia
E. pallida
var. simulata
E. laevigata
E. purpurea
E. pallida
var. pallida
E. atrorubens
var. neglecta
Axis 3
3.65
1.33
-0.99
-3.31
-5.64
-12.03
-7.56
-3.10
1.37
5.84
Axis 1
5.10
1.13
-2.85
-6.82
-10.80
Axis 2

E. purpurea as a group, with 82 out of 84 individuals in the AFLP phenotype data (Fig. 3.9 box 13, Table 3.13). Three variables were found to support the *E. purpurea* clade. Two individuals labeled *E. purpurea* were placed among other varieties in box 9. The results also clearly support *E. laevigata* as a distinct group, since all the 16 individuals fell in this clade (Fig. 3.9 box 4, Table 3.13) but without noticeable synapomorphies. The two other species, *E. atrorubens* and *E. pallida* are not supported by the AFLP phenotype data. For instance, some individuals (6 of 38) belonging to *E. atrorubens* var. *atrorubens* (Fig. 3.9 box 1, Table 3.13) are found in other boxes. Although box 2 is supported by a synapomorphy, it contains individuals belonging to varieties of *E. pallida*, whereas different individuals of the same varieties occur in other boxes. Boxes 5 and 6 contain a similar proportion of individuals labeled as *E. pallida* var *angustifolia* and var. *sanguinea*. The variety *E. pallida* var. *tennesseensis* is supported by all individuals in block 7 with a synapomorphy. Box 8 contains slightly more than half of the individuals of *E. pallida* var. *angustifolia* with the putative hybrid, supported by two synapomorphies (Fig. 3.9). Box 9 is a mixture of individuals of three varieties of *E. pallida*. All 48 individuals labeled *E. atrorubens* var. *neglecta* are supported by a synapomorphy, and the 41 individuals labeled *E. atrorubens* var. *paradox*a are supported by six synapomorphies (Fig. 3.9, boxes 10 and 11). Of 53 individuals only five of *E. pallida* var. *pallida* were isolated (Fig. 3.9, box 12, Table 3.13).

Evidently, the phylogenetic analysis of the AFLP phenotypes supports only two of the four species and four of the eight varieties of the two unsupported species defined previously by morphology. Generally, phylogenetic analysis yielded a tree with substantial homoplasy (Fig. 3.9). However, none of the bootstrap analyses, using the NJ

or the heuristic search, supported any of the groups of individuals according to their assignment to taxonomic entities. However, an examination of the tree indicated that individuals belonging to a population were found together, with a few exceptions (results not shown). For instance, a single plant from population EAT 98-0711-1 was found with all the plants of EAT 98-0528-1, all belonging to *E. atrorubens* var. *atrorubens* (Tables 3.1, 3.13); one plant from EPN 98-0529-2, belonging to *E. atrorubens* var. *neglecta*, was found with plants of population EPP 0823, belonging to *E. atrorubens* var. *paradoxa* (Tables 3.1, 3.13); all eight plants except one from population ES-3 of *E. pallida* var. *angustifolia* were scattered among the domesticated collections of *E. purpurea* (Tables 3.1, 3.13).

Discussion

AFLP results provide some support for the taxonomy of Binns et al. (2002a) at the species level but no general support at the varietal level. However, we also found that comigrating AFLP fragments, especially those that are polymorphic across individuals, were not identical (Table 3.2). This raises a number of questions concerning the use of AFLPs, namely, how then is it possible that there was support for the species in view of the lack of identity of comigrating bands? If comigrating bands are not identical, can they still be used to estimate genetic diversity, to make phylogenetic inferences, and to identify species or subspecific taxa within a genus? To answer these questions we will first consider the potential theoretical objections to the use of AFLP data for comparative biology and propose a solution that provides some justification for their use for specific purposes including the abovementioned.

AFLP fragments are anonymous and are identified by their length as determined by migration on a gel, not by their sequence composition. The nonidentical fragments of equal length will mistakenly be scored as identical (Koopman et al. 2001), leading to errors. Lack of homologies between comigrating AFLP bands have been reported in *Echinacea* (Mechanda et al. 2004a) and in garlic (Ipek and Simon 2003). Clark and Lanigan (1993) discussed the criteria that need to be met for RAPDs to be useful in estimating nucleotide divergence. AFLPs, like RAPDs, are dominant markers and some criteria discussed by Clark and Lanigan apply here. (1) Primer selection must not be biased in favor of those that reveal the most polymorphisms; the AFLP primers we used are not biased; (2) for diploids, a population sample must be examined to determine band frequencies; as seen in Table 3.1, we have examined samples of approximately 10 individuals per population among 58 populations; (3) true nucleotide divergence should not exceed approximately 10%; we found that among the monomorphic bands of samples across species and across varieties, the sequence similarity was greater than 90%. In sharp contrast, the divergence of the polymorphic bands was much greater, that is, only 33% - 54% of the sequences met the 90% threshold of similarity for the varieties, and this similarity ranged from 26%-33% among species (Table 3.2). As expected, for all the species, only 26% of the polymorphic bands had identical sequences at the 90% threshold similarity. Furthermore, various critics doubt the appropriateness of the data for phylogenetic inference, especially parsimony methods (Backeljau et al. 1995). This criticism is based on the fact that RAPD is a dominant marker system. Fragments of

Fig. 3.9. Condensed phylogram of *Echinacea* individuals obtained from AFLP phenotypes of the combined data, that is, the D10G10 matrix (see text). Phylogram derived from 435 individuals belonging to 52 populations (Table 3.1). Parsimony statistics based on tree analysis of the phylogram under the parsimony criterion: length 4596 changes; consistency index = 0.0268; homoplasy index = 0.9732; retention index = 0.6674. See Table 3.13 for box contents. "Synapomorphies", that is, AFLP bands associated with the subtending clusters, with their notation as primer combination with fragment size.

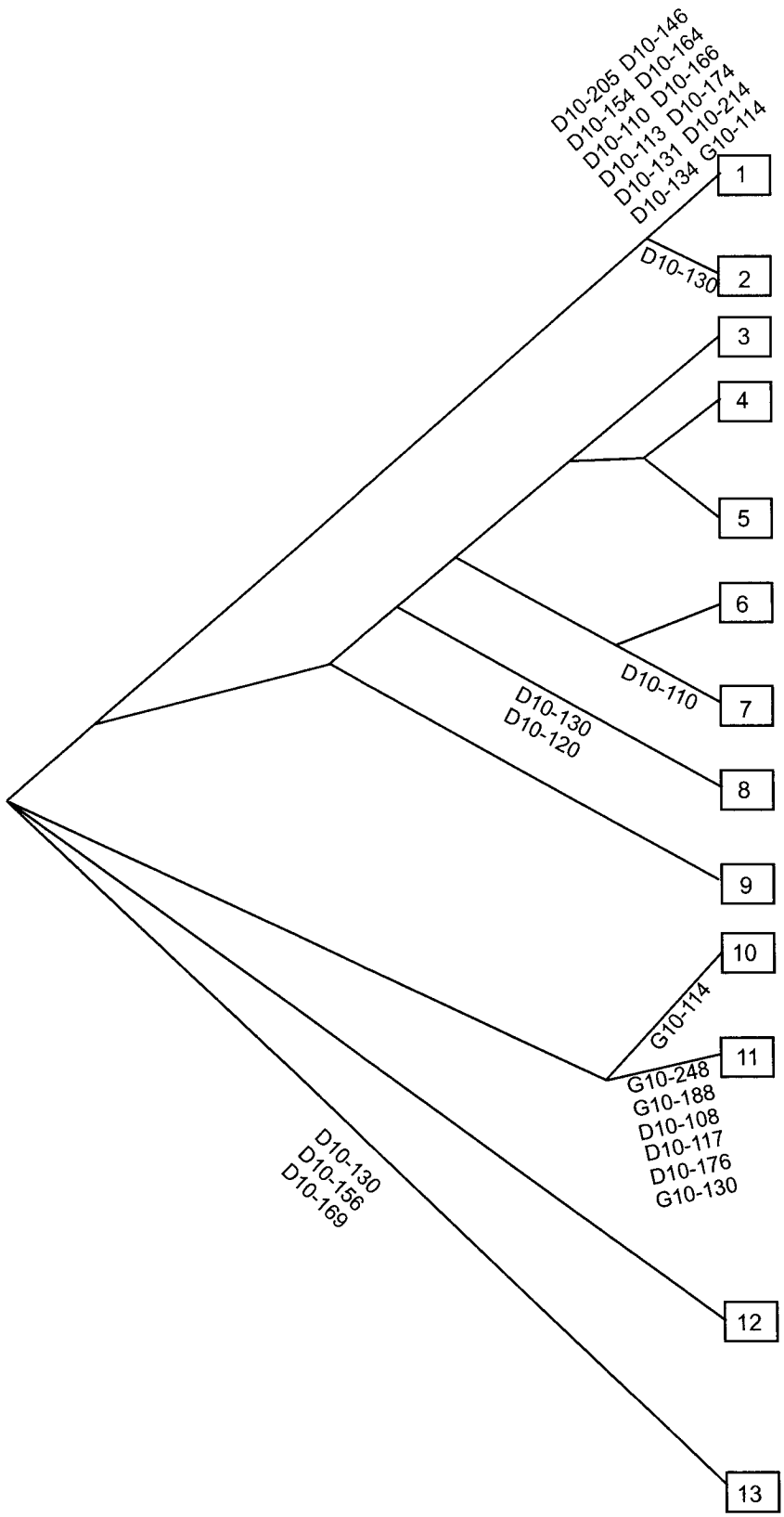

D10-205 D10-146
D10-154 D10-164
D10-110 D10-166
D10-113 D10-174
D10-131 D10-214
D10-134 G10-114
1
D10-130
2
3
4
5
6
D10-110
7
D10-130
D10-120
8
9
10
G10-114
11
G10-248
G10-188
D10-108
D10-117
D10-176
G10-130
D10-130
D10-156
D10-169
12
13

Table 3.13. The taxon and (or) entities of the condensed phylogram with the individual distribution in the populations of *Echinacea*.

Box	Taxon and (or) entities	Number of individuals/total individuals	Populations
1	*E. atrorubens* var. *atrorubens*	32/38.	EAT 98-0711-1, EAT 23881, EAT 23883, EAT 23884, EAT 98-0528
2	*E. pallida* var. *pallida*	10/34	ES017
	E. pallida var. *sanguinea*	8/67	ES 23962
	E. pallida var. *simulata*	20/28	ES 98-0608-2, ES011, ES 98-0608-1
3	*E. atrorubens* var. *atrorubens*	6/38	EAT 23883
	E. pallida var. *sanguinea*	19/67	ESA 23878, ESA 23874, ESA 98-0523, ESA 238783
	E. pallida var. *angustifolia*	8/53	EA COM1
4	*E. laevigata*	16/16	EL 23969, EL 23964
5	*E. pallida* var. *angustifolia*	9/53	EA COM1
	E. pallida var. *sanguinea*	6/67	ESA 0523-2, ESA 23873, ESA 23874
6	*E. pallida* var. *sanguinea*	34/67	EST 98-0606-1, EST 98-0602-1, EST 0823-2
	E. pallida var. *angustifolia*	1/53.	EA-COM1
7	*E. pallida* var. *tennesseensis*	18/18	ET 98-0611-1, ET 98-0611-2
8	*E. pallida* var. *angustifolia*	20/53.	EA 98-0712-1, EA 98-0717-1, EA 98-0716-2
	Hybrid: *E. pallida* var.	8/8.	EA014

	angustifolia × *E. atrorubens* var. *atrorubens*		
9	*E. pallida* var. *angustifolia*	8/53	EA 24058
	E. pallida var. *pallida*	19/34.	EPA 98-0604-1, EPA 009, EPA 018.
	E. pallida var. *simulata*	8/28	EP002
	E. purpurea	2/84.	EP COM1
10	*E. atrorubens* var. *neglecta*	48/48	EAT 98-0711-1, EAT 98-0528-2, EAT 28831, EAT 28833, EAT 23884
11	*E. atrorubens* var. *paradoxa*	41/41	EPP 98-0606-2, EPP 99-0606-2, EPP 0823, EPP 99-0606-1
12	*E. pallida* var. *pallida*	5/34	EPA 98-0604-1, EPA 009
13	*E. purpurea*	82/84	EP004, EP019, Trout Lake Samples.
	E. pallida var. *angustifolia.*	7/53	ES-1, ES-2. ES-3

similar length amplified from different *Drosophila* species for example, are not always derived from corresponding loci, and not all fragments in an amplification pattern are independent (Zande and van de Bijlsma 1995). These criticisms also apply to the AFLP marker system. The view that AFLP data were appropriate for phylogenetic inference was based on the empirical determination that comigrating bands are identical in terms of sequence homology [*i.e.,* identity] in closely related species of *Hordeum* (El-Rabey et al. 2002). However, the most critical view regarding the use of AFLP data for the estimation of population genetic parameters was based on the observation that polymorphic markers displayed various degrees of co-dominance, and therefore the assumption of dominance was unfounded (Wong et al. 2001).

In light of the above shortcomings of the AFLP marker system, including the result that many comigrating bands are not identical, we decided to treat the AFLP banding patterns as phenotypes as opposed to genotypes. Thus, the results of this study, and probably other AFLP studies, and their interpretations must be viewed in this fashion.

Echinacea species are outcrossing, and they hybridize to various degrees (McGregor 1968). For both the morphometric study (Binns et al. 2002a) and the present one, we attempted to obtain material from putative hybrid populations. The difficulty of identifying hybrid populations is due to the great amount of morphological variability (McGregor 1968; Binns et al. 2002a) and plasticity among and within populations (Binns et. al. 2002a). To verify the presence of triploid hybrids and ploidy level as reported by McGregor (1968) we employed flow cytometry, a technique successfully used by

Koopman (2000) and Koopman and De Jong (1996) to determine chromosome numbers in morphologically similar lettuce species. With the limited sample investigated in this study, we were unable to determine unequivocal hybridity of specimens in the field. Nonetheless, one putative hybrid population (Table 3.1) was identified solely on morphological basis (Binns et al. 2002a). The AFLP phenotype of the putative hybrid clustered with those of *E. pallida*, which was in agreement with its suspected parents.

Population genetic analysis

Based on the 74.5% total polymorphic bands, the gene diversity within species is more or less the same as within varieties, except for the rare case of *E. pallida* var. *tennesseensis*, with 43.5% polymorphic loci, which is found only in Tennessee and is almost extinct. The single example close to the former with respect to gene diversity is *E. atrorubens* var. *paradoxa* with yellow flowers. It had 50% polymorphic loci and was found in isolation only in Ozarkian uplands of western Missouri and adjacent Arkansas (Binns et al. 2002a). The percent variation within populations is roughly 40% (Tables 3.5 and 3.6), similar to that among populations within species. However, the percent variation among populations is only 27% within varieties (Table 3.6). Clearly, the variation among populations within varieties is smaller than expected because of the geographical proximity of the populations belonging to a particular variety. In a similar fashion, the variation among populations within species is reduced to 19%. The variation among populations within species was of the same magnitude as that within populations, clearly indicating differentiation among populations. This partition of variation is different from other cases, also using AFLPs, where the variation among cultivars was the same as within cultivars, as in the case of *Coffea arabica* (Steiger et al. 2002), or

where the variation within populations was greater than among populations as in *Poa annua* from Western Oregon (Mengistu et al. 2000). However, Kapteyn et al. (2002) found a much greater percent of variation within populations, ranging from 78% to 98%, than among species (2% to 21%) in three commercially relevant *Echinacea* species, using RAPDs. What Kapteyn et al. (2002) called populations were seed material obtained from nurseries and germplasm from a few locations containing possible admixtures. This variation is definitely not reflected in their principal coordinate plot (Fig. 2, in Kapteyn et al. 2002). In contrast, the spread of points in the canonical discriminant plot (Fig. 7) is wider.

Cluster analysis, principal coordinate analysis, Mantel tests

Cluster analysis and PCO of the AFLP phenotypes supported the relationship of two of the four species found by morphometric analysis (Binns et al. 2002a), but only supported *E. purpurea* and *E. laevigata* as species. Support was found neither for relationships of McGregor's (1968) species and varieties, nor for the varieties as determined by Binns et al. (2002a), that is, the varieties did not cluster within the species to which they belong. AFLP phenotype data have commonly been used as comparative data to support classification at the genus (Droogenbroeck et al. 2002) and species levels (Berg et al. 2002; Mace et al. 1999a, 1999b; Tomkins et al. 2001; Badr et al. 2002). However, in other cases there was no such corroboration, such as at the ecotype level (Breyne et al. 1999), at the subspecies level (Hedren et al. 2001), at the variety level (Roldán-Ruiz et al. 2001; Hedren et al. 2001), at the species level (Koopman et al. 2001), and at the section level (Wong et al. 2002). However, Ogden and Thorpe (2002) found AFLPs to be a rapid method for discriminating taxa at progressively finer evolutionary

levels in the case of Caribbean *Anolis* lizards. The data generated by AFLP analysis depend on the choice of restriction enzymes and primer pairs and there are no set rules for comparative purposes. AFLPs are thought to be spread throughout the genome (e.g. Nilsson et al. 1999). If fragments were unequally distributed over the range studied (Hedren et al. 2001), the number of fragments shared by coincidence would be higher.

In the present study, generally, the correlation among the different data sets was high at the species level, and much less so between the D10 and G10 data sets for the comparison among varieties (Table 3.8). Nonetheless, in most cases there was a significant tendency for the association between the data sets in various combinations. There was, however no association (and an extremely low correlation) between the geographic distances and the genetic distances (regarded here as based on phenotype data) between all the populations, except for the D10 primer set where the association was significant (Table 3.9). For *E. pallida*, Mantel tests were all significant and the associated correlations between the various distance matrices were higher (Table 3.9). The populations of *Echinacea* with their relatively high gene diversity (Tables 3.4-3.6) differ genetically from each other, independently of their geographical origin. It is not known if this is due to active genetic differentiation or to introgression.

Discriminant analyses

The rationale and thus the validity for using binary data, i.e. the AFLP scores, were presented in the Materials and Methods section with appropriate references. Discriminant analysis has been used recently with binary molecular data, e.g. RADPs (Semagn et al. 2000) and AFLPs (Guan et al. 2002).

Canonical discriminant analysis

The three-dimensional plot (Fig. 3.7) provided a good summary that the AFLP phenotype data supports the species defined by morphometrics (Binns et al. 2002a). In contrast, the AFLP data generally did not support the morphological classification into varieties by either McGregor (1968) or by Binns et al. (2002a). However, the three-dimensional plot of the varieties as groups (Fig. 3.8) does raise several interesting questions. Although *E. purpurea,* and to a lesser extent *E. laevigata* are separate from the cluster containing all *E. pallida* varieties, *E. laevigata* and *E. atrorubens* var. *atrorubens* are not distinct from the *E. pallida* cluster on the plot. Furthermore, the three varieties of *E. atrorubens* by themselves are clearly rendered distinct on the plot. Clearly, the squared Mahanalobis distances between the two varieties, *E. atrorubens* var. *paradoxa* and *E. atrorubens* var. *neglecta*, are greater compared to the distances among all the other varieties; however, their gene diversity is the lowest with the exception of *E. pallida* var *tennesseensis* (Table 3.4). Mahalanobis distances and genetic distances are computed differently, that is, the former from (statistical) variances. The generally similar results depicted by the PCO plot (Fig. 3.6), is based on the genetic distances (Nei 1972) among the varieties.

Classificatory discriminant analysis

Since this investigation did not yield either species-specific or variety-specific bands, we resorted to classificatory discriminant analyses as a means for elaborating an identification scheme. Banding patterns unique to each individual in the data were reported by Coart et al. (2002) for AFLPs in oak species and by Quagliaro et al. (2001) for AFLPs in sunflower. Although the identification results (the correct classification

results of the analysis) of the non-parametric method used in this investigation yielded the best results (Table 3.11), we are not aware of a way to implement it for the practical identification of specimens of *Echinacea*. Instead we have provided linear discriminant functions for the identification of *Echinacea* specimens, based on a minimum number of AFLP bands required for achieving at least 80% accuracy (Table 3.12). Identification by discriminant function might be useful for the identification of samples which do not possess characters required for morphological identification with conventional keys as in Binns et al. (2002a).

Phylogenetic analysis

AFLP data have been used to reconstruct phylogeny in *Echinacea* (Kim and Still 2004) and other plants (e.g., Raamsdonk et al 2000; Hedren et al. 2001; Koopman et al. 2001; Gobert et al. 2002 and Hodkinson et al. 2002), bovines (Buntjer et al. 2002), snakes (Giannasi et al. 2001), bacteria (Keim et al. 1997), and yeast (Ganter and Barros Lopes 2000). All these studies were based on the assumption that the comigrating bands are identical in sequence similarity in closely related species (Wong et al, 2001). In the present study, numerous co-migrating bands, especially from among the polymorphic ones that were sequenced, did not prove to be identical. Therefore, we treated the AFLP data as phenotypic data.

Like other phenotype data, AFLP data can be subjected to phenetic and phylogenetic analyses. Both phenetic and phylogenetic analyses require that characters (variables) are the "same", that is, homologous in the organisms being analysed. Comigrating bands may not have similar sequences (Mechanda et al. 2004a). However, even if they did, they still may not be comparable. Due to the nature of AFLPs, the

"same" fragments may have originated from different regions of the genome and are thus paralogous. The analysis of morphological characters presents similar problems, for instance, the character "blue petals" in two species may be due to different chemistries arising from different genes. In practice, both phenetic and phylogenetic analyses are performed on data where characters are treated the same at a particularly defined level in the organisms being compared. Thus, AFLP comigrating bands can be analysed in the same way based upon gel mobility (size in base pairs). Although we carried out both phenetic (data not reported) and phylogenetic analyses in the *Echinacea* study, we used only the results of the phylogenetic analysis to attempt to find taxon-specific bands for the groups (identified in the analysis as synapomorphies). Some methods such as UPGMA, although they find clusters, do not provide a means to detect taxon-specific bands. Thus in this work we resorted first to finding a NJ tree (considered by some as phenetic) followed by submitting this NJ tree to a tree analysis under the parsimony criterion.

Echinacea purpurea and *E. laevigata* were each clearly identified each as a group on the tree (Fig. 3.9, boxes 13 and 4 respectively). However, no synapomorphies could be found for *E. laevigata*. At the species level, individuals of *E. atrorubens* and *E. pallida* were distributed into different boxes on the tree. At the variety level, *E. atrorubens* var. *atrorubens*, *E. atrorubens* var. *neglecta*, *E. atrorubens* var. *paradoxa* and *E. pallida* var. *tennesseensis* were clearly identified as groups on the tree (Fig. 3.9 boxes 1, 10, 11 and 7 respectively). Synapomorphies were identified for all these varieties except for six individual plants of *E. atrorubens* var. *atrorubens* (box 3). A question remains as to whether all the groups identified on the tree (Fig. 3.9, Table 13) are suggestive of

taxonomic entities. Although some of these groups are not supported by synapomorphies, it is quite possible that they could be established with other AFLP primer combinations in addition to the ones used in the present work. Indeed, if one wants to screen the entire genome for polymorphisms, then a restriction enzyme that recognizes sequences that are randomly and uniformly distributed should be used (Breyne et al. 1999), such as *Sac*I or *Pst*I. The enzyme that we employed may be biased by AT content and towards non-coding regions (Bonnema et al. 2002). In addition, it is possible, that AFLP phenotypes stem from many parts of the genome, especially from the repetitive DNA sequences (Reamon-Buttner et al. 1999). Other possible explanations for the lack of general support for taxonomic identities may be introgression among the entities and (or) phenotypic plasticity.

We also need to address those few instances where an individual plant collected from a local population was positioned among individuals from another population on the NJ tree. There are several possible explanations for this observation, including (1) these individuals represent a haplotype occurring in both populations containing at least hundreds of individuals, (2) introgression, probably not uncommon within the outcrossing *Echinacea* species, has taken place, (3) migration of individuals between geographically close populations has occurred in the past. Furthermore, the observation that seven individuals, identified as *E. pallida* var. *angustifolia* from one population ES-3, were placed among the domesticated *E. purpurea* (Tables 1, 13) requires explanation. These individuals were scattered among collections from different origins. In addition to the former possibilities there is a chance of misidentification, since the other individuals from the same population were placed together with the other populations of *E. pallida*

although this species is not placed in a single group. It can hardly be a mix-up because *E. purpurea* has fibrous roots, easily distinguishable from the other species featured with tap-roots. Would it be possible that the phenotypic plasticity was also expressed in the AFLP phenotype?

In conclusion, since we were unable to find species-specific or variety-specific AFLP fragments, we used discriminant functions for identification. Possibly other primers, in addition to the ones employed in this study and probably from coding regions, might yield better and specific diagnostics. Nonetheless, the AFLP results lent support to some of the species defined by Binns et al. (2002a) by morphometrics. The present study has also shown that single individuals can be identified by their AFLP fingerprints.

CHAPTER FOUR

Direct shoot regeneration from leaf segments of mature plants of *Echinacea purpurea* (L.) Moench.

This chapter is a reproduction of the paper published in the Journal, In Vitro Cellular and Developmental Biology – Plant. 39: 505-509. (2003) by Subbaiah M. Mechanda, Bernard R. Baum, Douglas, A. Johnson, and John T. Arnason. This chapter meets part of objective 4 of the thesis, the development of tissue culture methods for multiplication of the species and varieties of *Echinacea*. Chemically elite plants, nearly extinct or endangered species can be multiplied and reintroduced back into the natural habitat using tissue culture methods. Only *Echinacea purpurea*, which is the most commercially cultivated species for pharmaceutical purposes, was regenerated directly from leaves of mature plants without going through the callus phase. Large number of uniform plants were regenerated and rooted successfully. The rooted plants were transplanted to pots, where they flowered and produced seeds normally. We did not attempt to reintroduce the tissue cultured plants into their natural habitats. However, our attempts to apply the same tissue culture procedures to other species and varieties of *Echinacea*, failed to produce shoots. This method has the potential for large-scale production of uniform *Echinacea* plants that have been authenticated for desired chemistry.

INTRODUCTION

Echinacea purpurea is a medicinal plant which was traditionally used by the First Nations of the American Great Plains and Canadian prairies for the treatment of respiratory ailments, sores, wounds, and a variety of other ailments (Tyler, 1993; Hobbs, 1994). Modern uses of *E. purpurea* preparations are mainly for the treatment of colds and influenza, but also for wound healing and treatment of candidasis (Bauer, 1999). Other species of *Echinacea* are of interest to the herbal industry (e.g., Bauer et al., 1990; Bauer and Foster, 1991; Binns et al. 2002b). According to an early taxonomic treatment by McGregor (1968), there are nine species of *Echinacea* in North America, of which three (*E. angustifolia, E. pallida and E. purpurea*) are normally cultivated and used for commercial production (Cox and Urbatsch, 1990; Hobbs, 1994). A recent morphometric analysis of *Echinacea* has led to a revision of the genus (Binns et al., 2002a), in which only four species and eight varieties of *Echinacea* are supported. Many of the wild populations of *Echinacea* throughout their range have been severely reduced due to habitat loss or wild crafting. One species, *E. tennesseensis*, is listed as an endangered species and *E. purpurea* has a low number of wild populations. This report is part of a study on tissue culture regeneration of all the four species of *Echinacea* and some of the varieties of commercial interest.

For both commercial production of *Echinacea* varieties and rehabilitation of wild populations, the development of efficient *in vitro* production of large quantities of plants by micro propagation is needed. In addition, regeneration of mature shoots from callus and suspension cultures is a critical requirement for the application of in vitro selection,

breeding for highly active ingredients, active ingredient detection and enhancement, and genetic engineering of *Echinacea*.

Amplified restriction fragment polymorphism (AFLP®) polymorphic markers have been used to predict phytochemical markers in cultivated mature *Echinacea* germplasm (Baum et al., 2001). A tissue culture method for direct shoot regeneration from the mature plants with the predicted phytochemical markers would shorten the time for propagation of these plants in large quantities and avoid the dependency on seeds that often are not homogenous due to out- crossing in the field.

Echinacea micropropagation has been possible in *E. purpurea* and *E. angustifolia* (Gockel et al., 1992; Choffe et al., 2000; Coker and Camper, 2000). Choffee et al., (2000) used *E. purpurea* petiole explants from 2- mo. old sterile seedlings to induce shoots using MS medium (Murashige and Skoog, 1962) containing 6-benzyladenine (BA) in combination with indole-3-acetic acid (IAA). Coker and Camper (2000) also used hypocotyl explants from sterile seedlings of *E. purpurea* to induce shoots using MS medium containing 1 mg l^{-1} each of α -naphthaleneacetic acid (NAA) and kinetin. Gockel et al., (1992) reported an efficient method for propagation of *E. angustifolia* using sterile seedlings using Woody Plant Medium (WPM; Lloyd and McCowan, 1981) supplemented with 3% sucrose and 1 mg l^{-1} BA.

Although our main aim was to regenerate plants directly from mature leaves, we also planned to use callus from leaves to test for chemical content in breeding lines of *Echinacea* to meet the demand by the herbal industry. Moreover, as an alternative to direct regeneration, our objective was to be able to produce shoots from callus.

No regeneration system has been established to produce shoots directly from mature plants of *Echinacea*. This communication describes for the first time a procedure for high-efficiency direct shoot regeneration from leaf explants of mature *E. purpurea* plants. The method produces shoots directly with high efficiency in 30 d, and rooted plants can be obtained in 60 d.

MATERIALS AND METHODS

Seed and Plant material sterilization. E. purpurea seeds (50-60) collected from natural populations in the mid-western USA were sterilized by shaking the seeds with 50 ml of 20% aqueous solution of Javex (bleach, 6% sodium hypo chlorite) in 250 ml beakers for 20 min., followed by shaking overnight on a rotary shaker in Plant Preservation Mixture (PPM) ^{TM} (10 ml l^{-1} in sterile water; Plant Cell Technology, Washington DC). Following washing three times in sterile water the seeds were placed in Petri plates (100x15mm) on sterile filter papers (Whatman® International Ltd., Maidstone, UK) saturated with 1ml l^{-1} ethylene (Etherel™) (Sigma-Aldrich Canada Limited, Oakville, Ontario) in distilled sterile water. The Petri plates were sealed with Parafilm® (American National Can™, Greenwich, CT) and incubated in a growth chamber for 14 d at 4°C, then the temperature was gradually raised over 5 d to 25°C. This is necessary for the stratification of seeds before germination. Germination occurred in 10 d after transfer to 25°C. The seedlings were transferred to Magenta boxes (Sigma) containing 50 ml solid MS (Murashige and Skoog 1962) medium, pH 5.5 (MS basal medium with sucrose and agar; Sigma). After 2 wk, they were transferred into pots containing 5:3:1 vermiculite: promix: large-grained quartz sand, and were transferred to a growth chamber (25°C, 16 h day; 20°C, 8 h night). Fertilizer of 20:20:20 (N: P: K), 1

ounce diluted in 3 gallons of water, was applied weekly and no pesticides or fungicides were used.

Tissue culture. Preliminary experiments were conducted in small-scale trial-and-error basis with different concentrations of auxins and cytokinins. Based on the response to callus induction and shoot induction, the following experiments were set up with the concentrations described below.

For multiple shoot production, approximately 1 cm^2 leaf explants from 6-mo.-old mature donor plants (grown in a growth room 25°C, 16 h day and 20°C, 8 h night, 30-60% humidity) were surface-sterilized by shaking the explants with 10% Javex (bleach, 6% sodium hypochlorite) for 10 min, followed by three washings with sterile water. The leaf explants (10-12 per Petri dish with 25 ml solid media) were subcultured into WPM (Sigma) and left for 3 d in WPM with no growth regulators, to assess contamination, with the intention of discarding any contaminated Petri dishes. They were then sub-cultured into WPM containing different concentrations of BA with 5% coconut milk (Sigma; filter sterilized). Each of the growth regulator treatments was in triplicate containing 10-12 leaf segments from different plants; the experiments were carried out twice. WPM powder (Sigma) was dissolved in water as per manufacturer instructions, 3% (w/v) sucrose was added, and the pH was adjusted to 5.5 and solidified with 0.85 % (w/v) agar (high gel strength; Sigma). The growth regulators and coconut milk were purchased as filter sterilized solutions and were added to the media after autoclaving. Callus was also induced on WPM media containing different concentrations of BA and NAA. The callus was then sub-cultured onto shoot-inducing media containing different concentrations of BA and 5% coconut milk.

Shoots arising from both callus and direct regeneration were rooted in WPM or MS medium containing 0.5-1.0 mg l^{-1} indole-3-butyric acid (IBA). These rooted plants were transplanted to soil and kept under intermittent mist for 2 d before normal growth in the growth chamber with 16 h light and 8 h dark cycles at 25±2° C. Student's *t*-tests were carried out on the three replicate results.

RESULTS

The seeds were surface-sterilized before germination, to eliminate fungi and bacteria with sodium hypochlorite treatment and sterile water washing. However, fungi could not be eliminated completely by this method. Hence, PPM at a concentration of 10mg l^{-1} was used to inhibit fungal and bacterial contamination of seeds. The leaves were sterilized with out PPM to avoid stunted growth of the shoots. Sterile mature leaf segments were used to induce callus and shoots. The results of the preliminary small-scale trial and error experiments are not provided.

The optimum concentration of growth regulators in WPM medium for induction of callus was 0.001 μM (0.5 mg l^{-1}) NAA and 6.0 μM, (2.5 mg l^{-1}) BA (Table 4.1). However, varying amounts of callus with varying levels of visual friability were formed at almost all the concentrations. Roots were formed at higher concentrations of NAA (data not shown).

Shoots were induced directly from the leaf segments of mature plants on WPM media containing different concentrations of BA and 5% coconut milk. The combination of BA and 5% coconut milk induced shoots at concentrations from 2.5 to 7.5mg l^{-1} BA. However, the optimum concentration producing shoots was 6.0 μM (2.5 mg l^{-1}) BA (Table 4.2). The shoots were excised from the explants and roots were induced on WPM

media containing different concentrations of IBA. The lowest concentration inducing 100% rooting was found to be 2 μM (1 mg l^{-1}) IBA with most of the concentrations supporting varying amounts of rooting (Table 4.3).

The rooted plants were transferred to pots after 2 d in a mist chamber, covered with polyethylene sheets. The sheets were removed and the mist was stopped for acclimatization of the plants in a growth chamber (Fig. 4.1). The rooted plants were transplanted into pots in a growth chamber. The plants appeared normal and flowered at 25^0C (16 h day and 8 h night) in a growth chamber.

Discussion

In vitro regeneration of *E. purpurea* and *E. angustifolia* from aseptically grown seedlings has been reported with very little information on other explants or other species of *Echinacea* (Gockel et al., 1992; Coker and Camper 2000; Choffe et al., 2000). Choffe et al. (2000) reported direct shoot regeneration from seedlings of *E. purpurea* using MS medium with IAA as the auxin and BA or thiadiazuron (TDZ) as cytokinin. Coker and Camper (2000) also used MS medium with NAA and kinetin to induce shoots from sterile seedlings of *E. purpurea*. Choffe et al. (2000) reported that BA was the most effective inducer of shoot regeneration from juvenile seedlings. In addition, TDZ and IAA increased the frequency of regeneration with MS medium. WPM was first used by Gockel et al. (1992) for tissue culture using juvenile seedlings of *E. angustifolia*. We did not try replacing NAA with IAA and BA with kinetin because this has been tried with juvenile seedlings (Choffe et al. 2000; Cocker and Camper 2000). We felt that it may not be appropriate for mature leaves as the growth regulator composition changes during growth and development. Coconut milk has been successfully used before for direct

Table 4.1. Callus produced from leaf explants of *Echinacea purpurea* after 1 mo. on WPM supplemented with 0.5 mg l^{-1} (0.001 μM) NAA and different concentrations of BA.

Table 4.2. Multiple shoots produced from leaf explants of *Echinacea purpurea* after 1 mo. on WPM supplemented with BA and 5% coconut milk.

BA concentration		Total number of explants producing shoots	Mean percentage of explants producing shoots	Mean number of shoots per responding explants	Mean number of shoots × responding explants	Mean number of shoots per explant	Std. Error
mg l^{-1}	μM						
0.0	0.0	0/72	0	0	0	0	0
0.5	1.2	0/72	0	0	0	0	0
1.0	2.4	0/60	0	0	0	0	0
1.5	3.6	0/72	0	0	0	0	0
2.0	4.8	10/60	16.66	8	80	1.3*	3.33
2.5	6.0	48/72	66.66	15	720	10**	3.33
3.0	7.2	20/60	33.33	11	220	3.7**	3.33
5.0	12	20/60	33.33	10	200	3.3*	6.66
7.5	18	30/60	50.0	7	210	3.5*	10.0
10.0	2.4	0/72	0	0	0	0	0
12.5	30	0/60	0	0	0	0	0
15.0	36	0/72	0	0	0	0	0

Data are means of two experiments, each with three replicates.

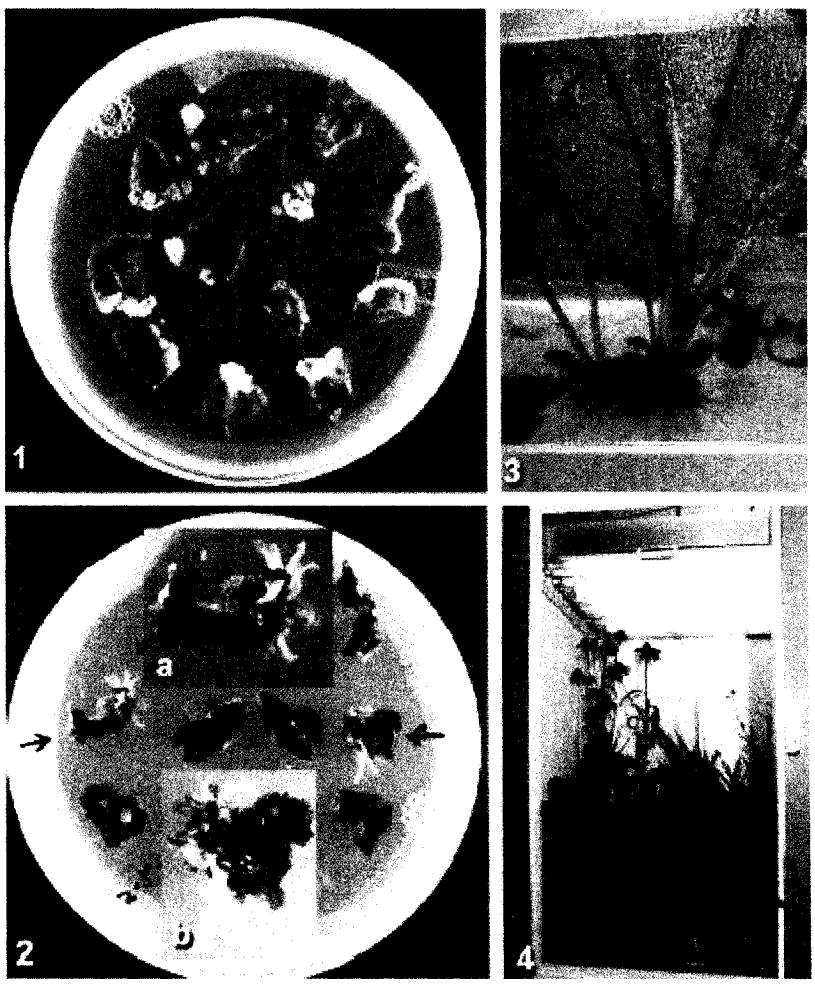
* $P < 5\%$; ** $P < 1\%$.

Table 4.3. Roots produced from the shoots of regenerated *Echinacea purpurea* after 1 mo. on WPM supplemented with IBA.

IBA concentration		Total number of explants producing roots	Mean percentage of shoot producing roots	Std. Error
mg l^{-1}	μM			
0.0	0	1	3.33	3.33
0.5	1	25	83.33*	8.81
1.0	2	30	100.00**	0.00
1.5	3	30	100.00**	0.00
2.0	4	30	100.00**	0.00

Data are means of two experiments, each with three replicates. * $P < 5\%$, ** $P < 1\%$.

Fig. 4.1. Regeneration of *Echinacea purpurea* from mature leaf segments. 1. Callus. 2. Shoot regeneration directly from the leaf segment. 3. Rooting of the regenerated shoots. 4. Potted plants flowering in a growth chamber.


1
3
a
2
b
4

regeneration of multiple shoots in *Eucalyptus teriticornis* (Subbaiah and Minocha 1990). However, there are no reports of tissue culture methods to produce multiple shoots directly from leaf segments of mature *Echinacea purpurea* plants.

Direct multiple shoot production from leaf explants is critical in order to produce large number of uniform plants from the elite lines of mature *E. purpurea* plants. This will also aid in the selection of elite lines of *E. purpurea* plants which have been grown in the field and tested for the desired chemistry for mass propagation of uniform plants. The callus induction stage can be overcome by direct induction of shoots. In the present work, by using WPM and adding 5% coconut milk with BA, we found that the frequency of regeneration was very high - as many as 12-20 shoots could be produced from segments of leaves of mature plants. Coker and Camper (2000) used a combination of NAA and kinetin at 1 mg l^{-1} and regenerated callus from only juvenile seedlings using MS media. They had 85% success rate with juvenile seedlings, whereas in the present study 66.7% regeneration efficiency was achieved using explants from mature leaves.

The high efficiency regeneration method from mature plants of *E. purpurea* is useful for the evaluation of native populations for desired chemicals. In addition, this will help in developing regeneration protocols for the near extinct varieties and other species of *Echinacea* such as *E. tennesseensis*, which may be reintroduced and maintained as natural populations. The commercially cultivated species *E. angustifolia* and *E. pallida* are currently being investigated in our laboratory for their regeneration capacity.

CHAPTER FIVE

General Discussion and conclusions

The standardization and authentication of *Echinacea* products depend on the correct identification of the *Echinacea* species and varieties which are cultivated or wild crafted. The herbal industry sells *Echinacea* supplements as tinctures, capsules, extracts and tablets. *Echinacea* roots and upper parts are bought by the companies from *Echinacea* farmers and wild crafters. Identification of the source and species is critical at this point, as it would be difficult to remedy the misidentification once the plants are processed. Traditionally, well characterized chemical markers are the most popular method for the identification and standardization of *Echinacea* species and products (Bauer 1999). However, the reliability of these chemical markers depends on the external environment, stage of development of the plant and processing conditions (Bauer 1999, Arnason 1999). A DNA based system for the identification and authentication of *Echinacea* species would be more useful and reliable than the chemical methods. The sensitivity, repeatability, reliability and the running costs of developing and standardizing DNA based methods is critical for the adoption of these methods. The markers developed in this study will further facilitate the correct identification for breeding better cultivars from the selected natural populations, for commercial production and conservation of the natural germplasm. The tissue culture methods developed for *E. purpurea* may be appropriate for the other species and varieties of *Echinacea*.

Collection of *Echinacea* samples

As part of the germplasm enhancement project, a germplasm collection of the natural populations was established for the study of chemical and genetic diversity of

Echinacea and the development of molecular markers for identification of species and varieties. The natural populations were collected during field trips by Dr. S. E. Binns, Mr. S. M. Mechanda, Dr. B. R. Baum and Dr. J. T. Arnason. The samples were collected based on the geographic locations given in McGregor (1968) with GPS (Global Positioning System), used to mark longitude and latitude as indicated in Fig (3.1). Rhizomes, leaves, flower heads and seeds were collected. The leaves that were used for DNA isolation were collected on ice, and shipped by courier overnight on ice and immediately frozen on arrival in liquid nitrogen prior to storage at -80 °C. The seeds and rhizomes were germinated in the greenhouse and leaves collected and frozen as before. Seeds from USDA germplasm collections were also germinated in the greenhouse and leaves collected and frozen as before. This was a shared responsibility between Dr. S. E. Binns and Mr. S. M. Mechanda. DNA was isolated after we had all the samples, frozen and stored at -80 °C.

Flow cytometry and chromosome counts

The chromosome numbers for the species and varieties of *Echinacea* has been reported (McGregor 1968); however, the amount of DNA per nucleus (C-value) was not. Flow cytometry has been used to determine the relative DNA amounts, even when there is a single chromosome difference between plants (Partec, Germany). In a pilot study, we counted the number of chromosomes and determined the relative DNA content per nucleus, focusing upon a subset of the 435 samples used in this study, to identify putative hybrids in the natural populations. An internal standard such as chicken erythrocytes (2.33 pg) is often used, when the C- value is not known, for example a study on DNA content variation in the Rosaceae (Dickson et al. 1992). As we were interested in the

relative amounts of DNA between samples, we used as a substitute both diploid and tetraploid *Brassica* for which we had C-values as our standard, along with Calibration Beads UV (Partec). There was corroboration between the chromosome counts of McGregor (1968) and the chromosome counts from this study (Table 3.3). However, the relative DNA amounts did not correspond to the number of chromosomes. For example, all the *E. pallida* varieties, with 22 chromosomes have DNA amounts in the range of 40 to 47 counts (Table 3.3). Whereas all the *E. atrorubens* varieties despite having only 22 chromosomes have DNA amounts in the range of 68-71 counts (Table 3.3). Based upon previous work in which flow cytometry was used in the *Rosaceae* species identification (Dickson et al.1992) or in Agrostis, again for species identification (Bonos et al. 2002), closely related species such as *Echinacea* may still be expected to show differences in relative DNA amounts which may be used for their identification. The complexity of the genome may be caused by gradual increase in gene number due to retention of duplicate genes and abrupt increases in the abundance of mobile genetic elements (Lynch and Conery 2003). Large variation in nuclear DNA within genera (Dickson et al. 1992) of Rosaceae has been observed, but the cause has not been understood very well. In addition the presence of polyploids in plants may add to the complexity of the genome (Soltis and Soltis 2000).

Two models of the origin of polyploids in plants have been put forward (Soltis and Soltis, 2000; Lee and Chen, 2001). In the first model, unreduced male and female gametes from two diploid species fuse to form polyploids or two autotetraploids can hybridize to from polyploids. In the second model, the two diploids hybridize to form allotetraploid and in the F1 hybrid the chromosomes double to form a polyploid. Both the

models could be operating in *Echinacea* to produce polyploid. Since *E. purpurea*, *E. laevigata* and some of the varieties such as *E. atrorubens var. paradoxa* and *E. pallida var. tennesseensis* are geographically isolated they may not be involved in formation of polyploids. *E. pallida* varieties and *E. atrorubens* varieties may have hybridized to form polyploids as they overlap geographically. However, identification of these hybrids is a tedious task as the chromosomes are very small.

AFLP and comigrating bands

AFLP technique was used to identify polymorphisms in the natural populations and commercial lines of *Echinacea*. What makes AFLP robust are as follows: it is highly repeatable and uses relatively small amounts of DNA (Mueller, and Wolfenbarger, 1999; Savelkoul et al. 1999). AFLP is universally applicable, and has very high discrimination between samples. Further this technique does not require sequence information *a priori* and has the highest multiplex ratio in terms of the number of loci sampled and genome wide coverage (Savelkoul et al. 1999; Mueller and Wolfenbarger, 1999). AFLP markers have been used in mapping genomes, positional cloning, varietal identification, population genetic studies, phylogenetic methods and germplasm management (Savelkoul et al. 1999; Mueller and Wolfenbarger, 1999). However, when applied to genetic diversity estimates or phylogenetic relationships, it has several limitations.

The AFLP data were manually scored in this study. Each and every gel had a sample that was designated as standard and a 30-330 bp AFLP DNA ladder. These were used to align the gels and only the bands which were well resolved across the gels and clear were included for data scoring. A binary matrix with band present (1) and band absent (0) were constructed and used for further analyses.

Over 160 papers on AFLP, published in 2004, have been identified using the PubMed site of NCBI. The majority of the papers treated the AFLP data as genotypic data in that the data were scored with the assumption that comigrating bands were similar in sequence. Conclusions on genetic diversity or identification of samples were drawn based on this assumption. A minority of papers considered the AFLP data as phenotypic data based on the lack of sequence information for comigrating bands and used the data to derive phylogenetic relationships and identification of the individuals.

Even though majority of the fragments originate from genomic DNA, the scores were based on the presence or absence of fragments and their mobility on the gel, which is definitely a phenotype. This conclusion also applies to RAPDs. There is not enough sequence information about the AFLP fragments and their specific locus of origin is not known. This means the AFLP bands have to be isolated, cloned and sequenced. When using a RAPD or AFLP results, the underlying assumption is that comigrating bands identified in different individuals, strains, populations, varieties or species (Riesberg 1996; Robinson and Harris 1997; Ipek and Simon 2003) are homologous. Thus in closely related species such as those of *Lactuca* and Barley, AFLP fragments are assumed to be related by descent, have the same or similar sequence and originate from the same locus within these genomes (Kim et al. 2004). As AFLP fragments are generally not isolated and sequenced, their sequences are not known. AFLP data are analyzed phenetically, but rarely by cladistic analysis, such as was the case for *Lactuca* species (Koopman et al. 2001). However, these authors found the topology of the phylogenetic tree and cladistic tree were similar indicating to them that comigrating bands are similar. It is not possible or practical to know the origin and sequence of all the AFLP fragments. Though the

fragments seem to be unique and orthologous and have been used for phylogenetic analysis in closely related species (Kim et al. 2004), our results in *Echinacea* contradict this notion. Various critics doubt the appropriateness of the AFLP data for phylogenetic inference, especially parsimony methods (Backeljau et al. 1995, Robinson and Harris, 1997). In addition, due to the view that AFLP fragments have alleles (Wong et al. 2001) and are dominant in nature, AFLP data have to be used with caution.

Comigrating band sequence similarity has been investigated with the RAPD and AFLP bands. Co-migrating RAPD bands were 92 % similar in maize (Beaumont et al. 1996). Riesberg (1996) reported that 91% of the RAPD co-migrating bands were similar. However, at the inter-specific level in *Brassica* species the pairwise similarity was only 80 % (Thormann et al. 1994). Rouppe van der Voort et al. (1997) sequenced 20 putative similar AFLP markers and 19 of them were shown to be nearly identical. Cerevera et al. (2001) cloned several fragments of identical length that were related and showed that each of them had the same sequence, but did notice several non visible bands co-amplifying during cloning. Ipeck and Simon (2003) sequenced 79 AFLP fragments from garlic accessions used for genetic diversity studies and found 95% of the comigrating bands were similar in sequences in closely related garlic clones, however, unrelated clones were not identical, which may be the case in the natural populations of *Echinacea* where there is no clonal identity of the samples.

Homology is similarity in DNA or protein sequences between two or more taxa, which is inherited from a common ancestor. Since molecular homology is a reflection of species or gene phylogeny, there are more kinds of homologous relationships between molecular sequences than in morphology (Patterson 1988). AFLP fragments are

anonymous and are identified by their length as determined by migration on a gel, not by their sequence composition. The non-identical fragments of equal length will mistakenly be scored as identical (Koopman et al. 2001), leading to errors in similarity measures and phylogenetic relationships. Lack of similarities between co-migrating AFLP bands have been reported in *Echinacea* (Mechanda et al. 2004a) and in garlic (Ipek and Simon 2003). A total of 79 fragments from a monomorphic band, 273 bp and present in all the four species and eight varieties of *Echinacea*, were cloned and sequenced. All the fragments turned out to be the same size as the expected size (273 bp); however, the sequence identity was never 100%.

Similarly, a polymorphic band of 159 bp which was present in some samples and absent in others across the 4 species and 8 varieties of *Echinacea* was chosen from atleast three samples per population, cloned and sequenced. A total of 48 clones from the polymorphic band were sequenced. Most of them (25) of them were of the expected size. However, a significant number of clones contained inserts of 74 bp (1 clone), 91 bp (11 clones), 108 bp (6 clones) and 125 bp (5 clones). Members of these size classes did not show significant sequence similarity with members of other size classes. The sequence identity was very low for the polymorphic fragments when compared to the monomorphic fragment (Table 2.1). This was true even for members within a size class, for example, when the 25 sequences with a size of 159 bp were aligned by themselves, suggesting that polymorphic fragments were very dissimilar in their sequences. Sequence similarity may also depend on the nature of the relatedness of the genus. So far, only the sequences of comigrating bands from either isogenic or near isogenic plants have been reported in the literature. But the samples analyzed in this study are derived from natural

populations of *Echinacea*. *Echinacea* is an open pollinated species with introgressants, which may explain the sequence dissimilarity of the comigrating bands. However, sequencing large number of different size fragments may be necessary to confirm the suggestion that monomorphic bands are more similar compared to polymorphic bands.

Amplification for 36 cycles during AFLP is a standard procedure for the AFLP kits (Invitrogen, 2004; Kim and Still, 2004; Vos et al.1995). PCR amplification of the gel isolated fragment is a critical step to convert the single stranded AFLP fragment into double stranded fragment for cloning. Further, 23 cycles of amplification for each isolated fragment after 36 cycles during AFLP is also a standard procedure for cloning fragments from the AFLP gels (Peters et al. 2001; Rouppe van der Voort et al. 1997). Peters et al. (2001) have validated this procedure for *Arabidopsis*. They reamplified the AFLP fragments for 30 cycles, checked the correct fragment sizes using agarose gel electrophoresis and sequenced the products directly without cloning. The sequences of the AFLP fragments were identical to the genomic sequences predicted by *in silico* analysis, thus verify the fidelity of AFLP.

Two lines of evidence argue against the hypothesis that the 74, 91, 108 and 125 bp fragments were derived from the 159 bp fragment by PCR during the AFLP and cloning processes. First, re-amplification of the cloned 159 bp fragment using the same conditions and primers, but with a ^{33}P end labeled primer yielded only a single 159 bp fragment as detected by autoradiography (data not shown). Secondly, when all the polymorphic sequences were aligned, there were no subsets of sequences contained within the smaller fragments which could be derived from the 159 bp fragment (Fig 2.5), even allowing for deletions introduced by recombination. The sequences were

completely different except for the primer sequences on both the ends, added during the PCR reaction. Thus the smaller fragments were not derived from the larger fragment by PCR during amplification or by homologous recombination in the host cell during cloning. At this time the precise origin of the smaller fragments remains a mystery.

About 30 AFLP gels were run during this study to identify the polymorphic bands, from populations of 4 species and 8 varieties of *Echinacea*. The large fragments (350- 500 bp range) in the top portion of the AFLP gel were not very well resolved and were difficult to score. The bottom portion of the AFLP gel with small fragments (less than 50 bp) was fuzzy and hard to score. So we decided to score fragments in the range of 50- 300 bp fragments which are in the middle portion of the gel. Hence we feel that for practical reasons the middle portion of the gel may be the best area for scoring the AFLP patterns. Furthermore, monomorphic fragments sequenced from the middle portion of the gel were more similar compared to the polymorphic fragments from the lower portion of the gel. However, to further confirm the suggestion that middle portion of the gel is more reliable than the lower portion of the gel, bands of different size need to be sequenced, rather than just scored.

Implications for analysis of AFLP data and genetic diversity estimation

Based on our results demonstrating that the comigrating AFLP fragments of *Echinacea* are not similar in sequence, we decided to treat the AFLP data as phenotypic data rather than genotypic data. Since the comigrating fragments were not 100% identical, plus it is not practical to sequence all of the AFLP fragments, it was decided that the comigrating bands of the same size be scored as homologous based on their mobility in the AFLP gel.

Non-homologous bands or homoplasy might bias genetic distance estimates between taxa (Lamboy 1994), and diversity measurements (Robinson and Harris 1999). Using non-homologous and non-independent AFLP fragment data will increase homoplasy (Bremer 1991). However, for the purpose of identification, the use of AFLP banding patterns as phenotypic characters is deemed appropriate by us. Thus AFLP data can be used for estimating genetic diversity and identification purposes.

The sequence information for the comigrating bands so far, is mostly from closely related genotypes such as in mapping populations of Potato (Rouppe van der Voort et al. 1997), closely related cultivars in Garlic (Ipeck and Simon 2002), and just one ecotype of *A. thaliana* (Peters et al. 2001) from a single geographic location. Most of these studies have samples from isogenic or near isogenic populations or from closely related cultivars. However, the sequences in the present study were obtained from natural populations of *Echinacea*. In general, it is expected that sequence similarity of the comigrating bands depends on the species studied, whether it is a self pollinating or cross pollinating genus and the number of species and varieties studied. Thus, closer the two samples are, the greater the likelihood that the bands are orthologous.

Population genetic analysis

Population genetic analysis was carried out on the AFLP data to determine the amount of genetic diversity among and within the natural populations of *Echinacea.* Smaller population size and isolation of the species or the variety has an effect on the descriptive statistics as is evident in *E. pallida* var. *tennesseensis* which is very isolated and in the endangered list, has low gene diversity, lower number of polymorphisms and percent polymorphic loci (Table 3.4). However, *E. laevigata*, species without varieties,

still shows relatively greater gene diversity, number of polymorphic loci and percent polymorphic loci. Even *E. purpurea*, which is commercially cultivated, shows high degree of polymorphism (Table 3.4), which is an indication of the high level of introgression which has been observed in *Echinacea* (McGregor 1968). The overall high percent polymorphic loci in the natural populations may be due to their different geographic locations and the introgression between and within populations, which is typical of all obligate outcrossing species such as *Echinacea*. A much greater percent variation within populations was reported for three commercially relevant *Echinacea* species using RAPD data (Kapteyn et al. 2002). They also found accessions of *E. pallida* (*E. pallida* var. *pallida*) and *E. angustifolia* (*E. pallida* var. *angustifolia*) had variations which were highly significant, but not the accessions of *E. purpurea*. The nine species and three varieties of *Echinacea* were highly polymorphic (80%), when AFLP data was used to determine the genetic diversity (Kim et al. 2004). The results form both these studies agree with the high rates of polymorphisms observed in this study, except for the geographically isolated species with very few populations, where the rates were marginally lower, may be due to limited gene flow. Large populations usually represent large gene pools and greater potential for measuring genetic diversity. However, the geographic range, distribution of populations and resources for collecting and analysis, determines the size of the sampling in a study such as the genetic diversity of *Echinacea*.

AMOVA was carried out to determine the genetic structure and to partition the variance that was present at different hierarchical levels, such as population groups, local populations, varieties and species of *Echinacea*. The variation in native *Echinacea* populations was partitioned almost equally within and among populations within species

(Table 3.5). This may reflect continuous gene flow throughout the range and introgression within and among the species. *E. purpurea* and *E. laevigata* do not have varieties; they are very isolated and have very few populations. There is more room for introgression and hybridization when there are more varieties in a species, for example *E. pallida* has five varieties and *E. atrorubens* has three varieties, and there are putative hybrids between these two species and their varieties (Table 1.1). However, the variation among the species and varieties was greater among the populations than within, the variation was very high within populations (Table 3.6). The geographical proximity of the populations belonging to particular variety may contribute to the variation among the populations. When all the population variance was compared there was more variation among populations than within populations (Table 3.7). Kapteyn et al. (2002) found high percentages of variance within populations of the three commercially important species, *E. purpurea*, *E. pallida* (*E. pallida* var. *pallida*) and *E. angustifolia* (*E. pallida* var. *angustifolia*). But the percentage of variance among population was consistently lower in all of them. So our data is consistent of what is expected of typical obligate outcrossing species such as *Echinacea* and is in agreement with the previous studies.

Cluster analysis and principal coordinate analysis of the populations

The populations and the individuals from each population were grouped using the Nei's genetic distance matrix into species and varieties. The four groups matched the four species of Binns et al. (2002a). But the eight groups, which were the varieties, did not match the eight morphological varieties of Binns et al. (2002a), as varieties within a species did not group together. Though the four species matched the species, their relationships were different compared with morphological species relationships

established by the morphological classification of Binns et al (2002a). *E. laevigata* was the most distant species and was closer to *E. purpurea* based on the AFLP data (Fig 3.3). However, according to morphological classification, *E. purpurea* was the most distant species, but was still closer to *E. laevigata*. Morphologically they have similar phenotypes, but still are separate species.

The four species were clearly separated in both cluster analysis and PCO, but not the varieties. *E. laevigata* was the most distant of the four species. The varieties did not match the morphological classification, but were consistent with the cluster analysis (Fig 2.4 and Fig 2.6). In PCO of both species and varieties, *E. pallida* is the most pivotal species, from which other species and varieties seem to have radiated. *E. atrorubens* var. *atrorubens* was isolated from the other conspecific varieties in both cluster and PCO analyses indicating that it is a separate group, even though it is one of the three varieties of *E. atrorubens* species.

Mantel test

Mantel test was used to test the association between various distance matrices. The relationships between the genetic distances obtained from primer sets, D10, G10, and the combined D10G10 data set were also established by this analysis. This will give an estimate of how the primer set matrices correlate among them and with the combined data set. Further, the relationships between the geographic distances of the populations, based on the location of the various populations and their genetic distances were also established.

There was high correlation and association between the primer sets and the combined data set in general. Only in two cases the correlations were low despite

association between them. They were, the D10 primer set with the G10 primer set when all the species, varieties and the hybrids were considered and the D10 primer set with the combined data set at the four species level. D10 primer is different from the G10 primer set just by one nucleotide and may target slightly different regions of the genome. However, there were very low correlations and no association between the genetic distances of the populations and their geographic distribution. This may be attributed to the high degree of variation seen within and among populations of *E. pallida* and *E. atrorubens.* There is very high introgression among and within these populations, with known hybrids between these two species (McGregor 1968; Binns et al. 2002a).

Correspondance analysis

Correspondance analysis requires that the data be in a two-way frequency cross tabulation table format, it is neither appropriate nor possible to analyse the binary AFLP data (Benzecri 1980; Benzecri 1992; Greenacre 1993; StatSoft Inc., 1984-2003), because most individual samples are unique haplotypes. Furthermore, our aim was identification of the species and to see if there is justification for some of the classes found with the morphological classification in a previous publication. Any type of frequency data derived from AFLP data of the samples from a population will not yield any kind of identification scheme.

Canonical discriminant analysis

The AFLP data was treated as phenotypic data after the non homologous nature of the sequences from the comigrating AFLP bands was verified. Canonical discriminant analysis and classificatory discriminant analysis were carried out to see 1) if there is justification to support for morphological classification of Binns et al. (2002a). 2) to see

if the species and varieties can be classified using the AFLP data. Phylogenetic analysis on the AFLP phenotypic data was carried out to find the synapomorphies associated with the groups for the purpose of identification of the species and varieties using AFLP data (Fig 2.10). These two analyses are based on the AFLP phenotype, which is separate from the cluster and PCO analyses; hence the results may not be compared. Canonical discriminant analysis was used to assess if there is justification for the species and the varieties of *Echinacea*, which were preset for the analysis according to the morphological classification of the specimens (Binns et al. 2002a). This analysis assumes the groupings from the morphological groupings, but uses only the AFLP variables for analysis. The AFLP binary variables or any other binary variables can be used for analysis, even though they do not satisfy all the assumptions (Krzanowski 1975; Lachenbruch 1975)

The assumptions in discriminant analysis (Klecka 1984) are that 1) The variables are in two or more groups 2) Atleast two variables cases per group 3) any number of variables, provided it is less than total number of cases minus two 4) discriminating variables are measured at the interval level 5) discriminant variables may not be a linear combination of other discriminating variables 6) covariance matrices for each group must be approximately equal, unless special formulas are used 7) each group has been drawn from a population with a multivariate normal distribution on the discriminating variables. Because our data did not meet all the assumptions, namely that the group has been drawn from a population with multivariate normal distribution of the discriminating variables, discriminant analysis was used as a guide only, and the test statistics were not used for the analysis.

Discriminant analysis of AFLP data has been used to identify the source of *E. coli* contamination (Guan et al. 2002) and they have also used other methods such as immunological verification to corroborate the AFLP data. Beef cattle were grouped based for their marbling score by discriminant analysis as high and low groups based on their AFLP fragment data (Tsuji et al. 2004). Discriminant analysis has been shown to be a robust technique which can tolerate some deviations from these assumptions and not all aspects of the analysis require these assumptions (Lachenbruch 1975; Klecka 1984). Only when there is small sample size does one need to be more careful about satisfying the assumptions, with large data such as the one generated in this study, the tests of significance can be ignored if the data violates the assumption (Krzanowski 1975; Lachenbruch 1975). Problems with the data set, such as large amounts of missing data or variables which are highly correlated, may pose a problem in the discriminant analysis (Klecka 1984). In addition, a variable with zero standard deviation within one or more groups, grossly different group sizes and outliers may cause a problem in the discriminant analysis, which was not the case in our data set.

The four groups (Fig 3.7) matched the four groups established by morphometric analysis and corroborated the four species established by morphology (Binns et al. 2002a). D10 primer set with 6 variables with the greatest F-values was important in forming the groups, with few variables from the G10 primer set (Table 3.11). However, the groups for varieties (Fig 3.8), despite high canonical correlations did not match the eight varieties of morphological classification (Binns et al. 2002a).

Classificatory discriminant analysis

Classificatory discriminant analysis was carried out to obtain the discriminant functions,

based on the AFLP fragments, to be used for the correct identification of an unknown specimen to the species and varieties of *Echinacea*. Since we did not find species specific or variety specific bands which could be used for identification purposes, we resorted to this analysis. Parametric analysis and non parametric analyses were used to find discriminant functions for the correct identification specimen to a species. Parametric analysis is appropriate when the assumptions, that the data are approximately normal within class distributions are used for the analysis. But in nonparametric analysis, the within class distributions need not be normal or can be different from multivariate normal distribution, as was in the case with our data and need not meet the assumptions for analysis.

Parametric analysis was useful in classifying the test sample with more than 90% correct classification, but required 68 variables which are the AFLP fragments for classification for the species and 74 variables for the varieties. But parametric analysis failed to meet the 99% correct classification criteria, which was the goal set at the start of the project. However, it still can be used for the purpose of identification of a specimen to a specific group with a lower degree of accuracy.

Nonparametric analysis based on the k-nearest neighbor analysis yielded only 97% correct classification for the species and 98% for the varieties. This was better than the parametric analyses. Thus, using one or both the methods it is possible to identify an individual plant to a species or variety using the discriminant functions.

Phylogenetic analysis

Phylogenetic analysis was carried out as a device only, to group the 435 individuals based on their AFLP phenotype and primarily to identify the synapomorphies

associated with the groups. These group specific synapomorphies are the AFLP fragments which can be used to identify the groups. We were interested in the bands, as they can be cloned, sequenced and sequence characterized. Specific PCR primers can be designed based on the sequence of the fragments and amplified by PCR, for use in identification of species or varieties. This way, unknown samples could be identified to the group.

The original Phylogram is 20 pages long containing all the 435 samples used in this study and is included in the appendix of the thesis, but a condensed phylogram is shown in Fig. 3.10. The choice of functional out-group (FOG) as an out-group for rooting trees follows the operational rules set (FIG/FOG) by Watrous and Wheeler (1981) in parsimony analysis was justified as *E. purpurea* was found to be the sister group of all the species of *Echinacea* in morphological classification. This has been used for testing relationships in the family Lacertidae using mitochondrial 12S and 16S genes and phylogenetic analysis by other such as Fu and Murphy (1997).

Two clear groups of individuals from *E. purpurea* and *E. laevigata* were apparent in the phylogram. Synapomorphies were associated with only *E. purpurea*. The other two species formed groups, but the individuals from the same species were not limited to the same group. Only four of the eight varieties, *E. atrorubens* var. *atrorubens*, *E. atrorubens* var. *neglecta*, *E. atrorubens* var. *paradoxa* and *E. pallida* var. *tennesseensis* formed groups in the phylogram and were associated with synapomorphies. Whether these groups are taxonomic entities will need further investigation.

Using AFLP data, Kim et al. (2004) identified two groups, one consisting of *E. purpurea*, *E. sanguinea* and *E. simulata* and the other group containing the rest of the

species and varieties of *Echinacea* tested. The division into two groups had high bootstrap support. However, when they portrayed the relationship of these groups by a different method such as PCA, *E. purpurea* clustered separately from all other taxa. The groups did not support the four species and eight varieties of *Echinacea* of Binns et al. (2002a).They also found that *Echinacea* taxa are closely related; with the average within population genetic distance only 0.029 whereas, the average pairwise distance between two populations was almost three times higher. They point out that small genetic distances and low bootstrap values for several groups within groups may indicate the recent divergence of *Echinacea*.

In contrast to Kim et al. (2004) we found four groups which support the four morphological species of Binns et al. (2002a) namely, *E. purpurea*, *E. pallida*, *E. atrorubens* and *E. laevigata*. We differed from Binns et al (2002a) in that *E. laevigata* was the most distant species, whereas *E. purpurea* was the most distant in their study. With respect to varieties we did not find support for the eight varieties of Binns et al. (2002a).

There are several possible reasons for the differences between our results and the results of Kim et al. (2004) including: 1) sample size. Kim et al. used only 39 plants representing nine species and three varieties (sensu McGregor, 1968), compared to the 435 plants collected and analyzed in our study; 2) most of their samples were obtained as seed from germplasm collections and identified based upon the taxonomy of McGregor, 1968. We used the updated version of Binns et al. (2002a). Thus, there may have been misidentification or confusion in the taxonomy of *Echinacea*; 3) possible differences in laboratory practices in the application of AFLP such as purity of DNA, ensuring

complete DNA digestion, ligation of the adaptors and data scoring may vary affecting the results; 4) the use of neighbor-joining method to construct a dendrogram of individual samples with the assumption of comigrating band homology. This assumption is not be valid appropriate, as our results showed that the comigrating bands were not similar.

Both NJ analysis and UPGMA analysis can be used for analysis of AFLP data. However, both types of analysis may not be suitable for establishing phylogenetic relationships. Neighbor joining methods also work by clustering and do not assume a molecular clock. NJ methods are computationally rapid and the approximation is considered to be quite good for grouping large number of taxa (Felsenstein, 2004). However, NJ methods test only single events and do not estimate back mutations. This method is not suitable for AFLP fragments which are not homologous and non independent and underestimates long distances. Thus the use of NJ methods depends on the kind of data and the goal of the analysis. We used phylogenetic analysis of the phenotypic data as a device to identify the synapomorphies associated with the groups for identification only, but not for phylogenetic analysis.

UPGMA is a group average method with a hierarchical monotonic clustering scheme and has a conserving effect on space (Sneath and Sokal, 1973). Even though UPGMA assumes a molecular clock and rate of constancy among the lineages, the method can be used for studying relationships between species and varieties within a genus. However, phylogenetic analysis may yield misleading results if the distances actually reflect a substantially nonclock-like tree (Felsenstein, 2004). UPGMA was used to study the relationships between the species, and the varieties, separately in our study.

BA concentration		Mean number of explants producing callus/total number of explants	Mean percentage of explants producing callus	Std. Error
mg l^{-1}	µM			
0.0	0.0	0/72	0	0
0.5	1.2	0/60	0	0
1.0	2.4	0/72	0	0
2.0	4.8	24/72	33.33**	3.33
2.5	6.0	32/60	53.33**	6.66
5.0	12.0	22/60	36.67**	3.33
7.5	18.0	20/60	33.33*	8.81

Data are means of two experiments, each with three replicates. * $P < 5\%$; ** $P < 1\%$.

Tissue culture and Ecology of *Echinacea*

The major source country for *Echinacea* is the United States of America with the natural populations distributed in the central and south-western USA. Since the natural populations are dwindling due to wild crafting and loss of habitat, commercial cultivation is the choice to meet the requirements of the herbal industry in North America and Europe. *Echinacea* is collected by the herbal industry by wild crafting, controlled wild crafting or semi-cultivation, and commercial cultivation. Quality, consistency, and efficacy, of wild crafted material are a problem and standardization is especially important, when it comes to clinical trials. *E. purpurea* is the best known species for commercial cultivation in Europe and USA, with some cultivation of *E. pallida* var. *pallida* and *E. pallida* var. *angustifolia*. Commercial cultivation of *Echinacea* is mostly located in North-Western United States, and western Canada (Letchamo et al. 2002). Austria, Germany, Russia, New Zealand, Ukraine, Yugoslavia and Republic of South Africa are growing *E. purpurea* and *E. pallida* var. *pallida* on commercial scales. Non traditional *Echinacea* growing countries such as Africa, Asia, Latin America and Middle East are introducing commercial cultivation.

Commercial cultivation of *Echinacea* will help in preserving the wild populations of *Echinacea*. However, monoculture in cultivated *Echinacea* may become a problem and result in the loss of genetic diversity. Genetic diversity in commercial *E. purpurea* was very high as observed in this study. There is vast amount of variation and genetic diversity in the natural populations of *Echinacea*. Hence a germplasm collection of all the genetically diverse *Echinacea* species and varieties is a good idea for genetic selection and improvement of *Echinacea*. Cultivars with higher root and shoot yield, ease of

mechanical harvesting, uniform growth, flowering, seed ripening, good leaf to stem ratio are necessary for commercial cultivation (Letchamo et al. 2002). In addition higher content of active ingredients such as cichoric acid, isobutyl amides, polysaccharides and essential oils are also desirable in the commercial cultivars (Letchamo et al. 2002). Hence, we have to study all the variations in the natural populations and preserve the germplasm after careful evaluation. AFLP methods are ideal for detecting large amounts of variation present within and among individuals, populations, species and varieties.

The tissue culture methods developed in this study for *E. purpurea* may be used for production of uniform *Echinacea* plants for commercial cultivation. This would ensure the quality, consistency and efficacy of the products. The uniform plants can also be used in the selection and breeding efforts for better cultivar development. Hydroponic cultivation of *E. purpurea* is seen as a faster way of producing potent roots, than field grown roots, as it takes three years for the roots to fully develop for harvesting. In addition germinating seeds and transplanting seedlings is laborious and expensive. Thus tissue culture rooted plants can be used for commercial cultivation either in the field or by hydroponics.

Conclusions

Comigrating, AFLP monomorphic bands are similar, whereas the polymorphic bands were not similar in *Echinacea* species and varieties. However, due the identical mobility of the AFLP comigrating bands, they were considered homologous and treated as phenotypic data.

No species-specific or variety-specific bands were found for the identification of *Echinacea* species or varieties. The four species morphological classification of

Echinacea was corroborated by the AFLP data, but not the eight varieties of *Echinacea*. Only two of the four species, and three of the eight varieties, were clearly identified as groups by phylogenetic analysis, with synapomorphies. Mass propagation of *Echinacea*, using tissue culture of natural populations can be used for production of uniform plants, from selected elite lines for commercial cultivation, hydroponic culture or for reintroduction into the natural habitat of the nearly extinct species or varieties.

Future work

Practical methods for species identification are critical for the herbal industry. Further work is required to establish these methods and make them practical. Determining ploidy levels of new germplasm can assist in species determination (Bonos et al. 2002). Use of the laser flow cytometry for determining the ploidy level of the entire *Echinacea* germplasm may indicate the extent of introgression in the natural populations. In addition chromosome counts of the entire germplasm would be useful in determining the correlation of the chromosome numbers to the nuclear DNA content.

AFLP data has been used to resolve species in both eukaryotes and prokaryotes (Savelkoul et al. 1999). DNA sequencing has been used successfully to identify individual samples and draw species boundaries in populations (Stoeckle, 2003; Hebert et al. 2004). However, in the case of *Echinacea* each individual plant had a unique AFLP profile. Thus it is hard to draw species boundaries based only on AFLP data. Even though individual samples can be grouped using the AFLP data, these groups may not represent species. The species must be independently researched using other methods such as morphology, ploidy levels and crossing experiments. The whole biological

species concept may have to be re-examined if AFLP data were to be used for drawing species boundaries depending on taxonomic level and category.

The problems in data scoring, analysis and the potential bias that may be introduced due the codominant nature of AFLP combined with lack of sequence identity limits their use for population genetic and evolutionary studies. Sequencing all of the comigrating monomorphic and polymorphic bands from all the species and varieties of *Echinacea* would help better understand the nature of the comigrating bands. Further, abundant markers such as SNPs (Single Nucleotide Polymorphisms) could be used as simple genetic markers, which are found in the vicinity of every possible gene in the genome, for genetic diversity estimation (Osman et al. 2003). The high level of polymorphism observed in *Echinacea* facilitates identification of SNPs. AFLP fragment derived SNPs can be used to increase the number of polymorphisms and to better cover the genome (Rafalski 2002).

High-throughput sequencing of multiple orthologous loci across species and varieties, and complete understanding of the processes influencing nucleotide diversity are critical. Association of the diversity with function of genes and phenotypes in the natural populations will yield useful alleles for plant breeding and crop improvement. Association approaches can be combined with high throughput genomics to characterize the genomes (Buckler et al. 2002). Application of expressed sequence tags in plant biology is gaining strength with large amounts of sequence data available for non-model organisms and whole genome sequencing may be the future (Crawford and Mort, 2003). Sequencing, one gene or a set of genes across the species and varieties of *Echinacea* will help understand the process of speciation and reveal the diversity of *Echinacea* which is

not possible now. As we seek uniform and practical methods of species identification, DNA bar coding may be used in conservation biology (Stoeckle 2003).

Hebert et al. (2004) have identified species of birds taken from museum samples which were tentatively identified before the molecular species classification. They have used a 10X average intraspecific difference as a threshold for sequence difference for drawing species boundaries based on a 684-bp region of the mitochondrial cytochrome oxidase I gene. However, in plants the same gene cannot be used due to lack of differences in the sequence between species. Finding a suitable gene in plants, which are highly conserved to make universal primers for PCR amplification and still divergent enough for species identification is a real problem.

Establishing tissue culture methods for the regeneration and multiplication of all the species and varieties of *Echinacea* is necessary for commercial production as well as for conserving the natural populations.

LITERATURE CITED

Adam, K. L. 2002. *Echinacea* as an alternative crop. Pages 1-11. http://www.attra.ncat.org/attra-pub/PDF/echinaca.pdf

Aggarwal, R. K., V. V. Shenoy, J. Ramadevi, R. Rajkumar and L. Singh. 2002. Molecular characterization of some Indian Basmati and other elite rice genotypes using fluorescent-AFLP. Theor. Appl. Genet. 105: 680-690.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215: 403-410.

Anderson, T. W. 1984. An introduction to multivariate statistical analysis, 2nd edition. John Wiley & Sons, NewYork. USA.

Arnason, J. T. 1999. North American Raw Materials: the case for an isobutylamide standardized product. Echinacea: Past, Present and Future. International Conference proceedings. Trout Lake Farm. Washington, USA, unpaginated

Awang, D. V. C. 1999. Immune stimulants and antiviral botanicals: Echinacea and ginseng. P.450-456. In: J. Janick (ed.), Perspectives on new crops and new uses. ASHS Press, Alexandria, VA, USA.

Backeljau, T., L. D. Bruyn, H. D. Wolf, K. Jordaens, S. V. Dongen, R. Verhagen, and B. Winnepenninckx. 1995. Random amplified polymorphic DNA (RAPD) and parsimony methods. Cladistics. 11: 119-130.

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

Barrett, B. 2003. Medicinal properties of Echinacea: a critical review. Phytomedicine.10: 66-86.

Bauer, R. and S. Foster. 1991. Analysis of alkamides and caffeic acid derivatives from *E. simulata* and *E. paradoxa* roots. Planta Medica. 57: 447-448.

Bauer, R., P. Remiger, and E. Alstat. 1990. Alkamides and caffeic acid derivatives from the roots of *Echinacea tennesseensis*. Planta Medica. 56: 533-534;

Bauer, R. 1998. Standardization of *Echinacea purpurea* expressed juice with reference to cichoric acid and alkamides. J. Herbs Spices Med. Plants loc. cit.

Bauer, R. 1999. *Echinacea* chemistry and analytical issues. AHPA International *Echinacea* symposium proceedings. American Herbal Product Association (AHPA), Silver Spring, Maryland, USA, unpaginated.

Bauer, R. and S. Foster. 1991. Analysis of alkamides and caffeic acid derivatives from *E. simulata* and *E. paradoxa* roots. Planta Medica. 57: 447-448.

Bauer, R. and H. Wagner. 1991. Echinacea species as potential immunostimulatory drugs. p. 253-321. In: H. Wagner and N. R. Farnsworth (eds.), Economic and medicinal plant research. Vol. 5. Academic Press, New York.

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

Baum, B. R., S. E. Binns, S. Mechanda, and J. T. Arnason. 1999. The *Echinacea* Germplasm Enhancement Project. In proceedings of the International *Echinacea* Symposium, June 3-5, 1999, Kansas City, MO. AHPA (American Herbal Products Association). Maryland. USA. unpaginated.

Beaumont, V. H., J. Mantet, T. R. Rocheford, and J. M. Widholm. 1996. Comparison of RAPD and RFLP markers for mapping F-2 generations in maize (*Zea mays* L.). Theor. Appl. Genet. 93(4): 606-612.

Benzecri, J. P. 1980. L'analyse des données tome 2 : l'analyse des correspondances, Paris : Bordas.

Benzecri, J. P. 1992. Correspondence Analysis Handbook, New-York: Dekker, P. 1994, L'analyse factorielle, Paris: PUF

Berg, R. G., G. J. van der, Bryan, A. Del Rio, and D. M. Spooner. 2002. Reduction of species in the wild potato Solanum section Petota series Longipedicellata: AFLP, RAPD and chloroplast SSR data. Theor. Appl. Genet. 105: 1109-1114.

Binns, S. E. 2001. The taxonomy, phytochemistry and biological activity of the genus *Echinacea*. *Asteraceae*. PhD Thesis submitted to the University of Ottawa, Ottawa. Canada.

Binns, S. E., B. R. Baum, S. Mechanda, and J. T. Arnason. 1999. The *Echinacea* Germplasm Enhancement Project. In proceedings of the International *Echinacea* Symposium, June 3-5, 1999, Kansas City, MO. AHPA (American Herbal Products Association). Maryland. USA. Unpaginated.

Binns, S. E., B. R. Baum and J. T. Arnason. 2002a. A taxonomic revision of the genus *Echinacea* (Heliantheae: Asteraceae). Systematic Botany. 27(3): 610-632.

Binns, S. E., J. F. Livesey, J. T. Arnason and B. R. Baum. 2002b. Phytochemical variation in Echinacea from roots and flowerheads of wild and cultivated populations. J. Agric. Food Chem. 50: 3673-87.

Blumenthal, M. 1992. Conserving wild medicines. Whole Foods. June 1992: 20-21.

Blumenthal, M., A. Goldberg, and J. Brinckmann. 2000. Herbal Medicine. Expanded commission. Echinacea monographs. In: American botanical council press. Austin. Texas. USA

Bonnema, G., P. Berg, and P. Lindhout. 2002. AFLPs mark different genomic regions compared with RFLPs: a case study in tomato. Genome. 45: 217- 221.

Bonos, S. A., K. A. Plumley, and W. A. Meyer. 2002. Ploidy determination in *Agrostis* using flow cytometry and morphological traits. Crop Sci. 42: 192-196.

Borchers, A. T., C. L. Keen, J. S. Stern, and M. E. Gershwin. 2000. Inflammation and Native American medicine: the role of botanicals. Am. J. Clin. Nutr. 72: 339- 47.

Brahmasandra, S. N., D. T. Burke, C. H. Mastrangelo, and M. A. Burns. 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? Pl. Syst. Evol. 175: 39-54.

Brevoort, T. 1996. US botanical market, an overview. Herbalgram. 36: 49-57.

Breyne, P., D. Rombaut, A. Van, M. Van, and T. Gerats. 1999. AFLP analysis of genetic diversity within and between *Arabidopsis thaliana* ecotypes. Mol. Gen. Genet. 261: 627- 634.

Briskin, D. P. 2000. Medicinal plants and phytomedicines, linking Plant biochemistry and Physiology to Human Health. Plant Physiol. 124: 507-514.

Buckler IV, E. S. and J. M. Thornsberry. 2002. Plant molecular diversity and applications to genomes. Current Opinions in Plant Biology. 5: 107-111.

Buntjer, J. B., M. Otsen, I. J. Nijman, M. T. Kuiper, and J. A. Lenstra. 2002. Phylogeny of the bovine species based on AFLP fingerprinting. Heredity. 88: 46-51.

Carr, J., M. Xu, J. W. Dudley and S. Korban. 2003. AFLP analysis of genetic variability in New Guinea *impatiens*. Theor. Appl. Genet. 106: 1509-1516.

Cerevera, M. T., J. A. Cabezas, J. C. Sancha, F. Martinez de Toda, and J. M. Martinez-Zapater. 1998. Application of AFLPs to the characterization of *Vitis vinifera* genetic resources. A case study with accessions from Rioja (Spain). Theor. Appl. Genet. 97: 51-59.

Cerevera, M. T., V. Strome, B. Ivens, J. Gusmao, B. H. Liu, V. Hostyn, J. V. Slycken, M. Van Montagu and W. Boerjan. 2001. Dense genetic linkage maps of three *Populus* species based on AFLP and microsatellite markers. Genetics. 158: 787-809.

Chalhoub, B. A., S. Thibault, V. Laucou, C. Rameau, H. Hofte, and R. Cousin. 1997. Silver staining and recovery of AFLP amplification products on large denaturing polyacrylamide gels. Biotechniques. 22(2): 216-218.

Choffe, K. L., J. M. R. Victor, S. J. Murch, and P. K. Saxena. 2000. *In vitro* regeneration of *Echinacea purpurea* L. Direct somatic embryogenesis and indirect shoot organogenesis in petiole culture. In Vitro Cell Dev Biol. Plant. 36: 30-36.

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

Coart, E., V. Lamote, M. De Loose, E. Van Bock Stable, P. Lootens, and I. Rodan Ruiz. 2002. AFLP markers demonstrate local genetic differentiation between two

indigenous oak species *Quercus robus* L. and *Quercus petraea* (Matt) (Liebl) in Finnish populations. Theor. Appl.Genet. 108: 1-8.

Cochrane, W. G. and C. E. Hopkins. 1961. Some classification problems with multivariate qualitative data. Biometrics. 17: 10-32.

Coker, P. S., and N. D. Camper. 2000. *In vitro* culture of *Echinacea purpurea* L. J. Herbs Spices Med. Plants. 7:1-7

Cover, T. M. and P. E. Hart. 1967. Nearest neighbor pattern classification. IEEE Transactions on information theory. IT-13: 21-27.

Cox, P. B. and L. E. Urbatsch. 1990. A phylogenetic analysis of the coneflower genera. Systematic Botany. 15: 394-402.

Crawford, D. J., and M. E. Mort. 2003. New Trends in Plant Systematics. Taxon. 52: 3-7

Crespel, L., M. Chirollet, E. Durel, D. Zhang, J. Meynet, and S. Gudin. 2002. Mapping qualitative and quantitative phenotypic traits in Rosa using AFLP markers. Theor. Appl. Genet. 105: 1207- 1214.

De Rick, J., E. Calsyn, I. Everaert, E. Van Bockataele, and M. De Loose. 2001. AFLP based alternatives for the assessment of distinctness, uniformity and stability of sugar beet varieties. Theor. Appl. Genet. 103: 1254-1265.

Despres, L., L. Gielly, B. Redoutet and Taberlet, P. 2003. Using AFLP to resolve phylogenetic relationships in a morphologically diversified plant species complex when nuclear and chloroplast sequences to fail to reveal variability. Mol. Phylogenet. Evol. 27: 185-196.

Dice, L. R. 1945. Measures of the amount of ecologic association between species. Ecology. 26: 297-302.

Dickson, E. E., K. Armuganathan, S, Kresovich and J. Doyle. 1992. Nuclear DNA content variation within the Rosaceae. Amer. J. Bot. 79: 1081-1086.

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

Droogenbroeck, B. V., P. Breyne, P. Goetghebeur, E. Romeijn-Peeters, T. Kyndt, and G. Gheysen. 2002. AFLP analysis of genetic relationships among papaya and its wild relatives (Caricaceae) from Ecuador. Theor. Appl. Genet. 105: 1-8.

El-Rabey, H. A., A. Badr, R. Scher-Pregl, W. Martin and F. Salamini. 2002. Speciation and species separation in *Hordeum* L. (Poaceae) resolved by discontinuous molecular markers. Plant Biol. 4: 1-19.

Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 131: 479-491.

Excoffier, L. 2000. Analysis of Population Subdivision: In Handbook of Statistical Genetics. D. Balding, M. Bishop and C. Cannings. (eds.) Wiley & Sons, Ltd.

Foster, S. 1991. *Echinacea*: nature's immune enhancer. Rochester. VT. Healing Arts Press. USA.

Fu, J. and R. W. Murphy. 1997. Phylogeny of Chinese *Orelalax* and the use of functional outgroups to select among multiple equally parsimonious trees. Asiatic Herpetol. Res. 7: 38-43.

Ganter, P. F. and M. de Barros Lopes. 2000. The use of anonymous DNA markers in assessing worldwide relatedness in yeast species *Pichia kluyveri* Bedford and Kudrjavzev. Can. J. Microbiol. 46: 967-980.

Giannasi, N., R. S. Thorpe, and A. Malhotra. 2001. The use of amplified fragment length polymorphism in determining species trees at fine taxonomic levels: analysis of a medically important snake, *Trimeresurus albolabris*. Mol. Ecol. 10: 419-26.

Gilbert, E. S. 1968. On discrimination using qualitative variables. J. Am. Stat. Assoc. 63: 1399-1412.

Gobert, V., S. Moja, M. Colson, and P. Tabarlet. 2002. Hybridization in the section Mentha (Lamiaceae) inferred from AFLP markers. American J. of Botany. 89: 2017-2023.

Gockel, C., C. H. Wawrosch, C. H. Leonhardt, and B. Kopp. 1992. Micro propagation of *Echinacea angustifolia*. Planta Medica. 58: supplement issue 1A: 626.

Greenacre, M. J. 1993. Correspondence Analysis in Practice, London: Academic Press.

Guan, S., R. Xu, S. Chen, J. Odumeru. and C. Gyles. 2002. Development of a procedure for discriminating among *E. coli* isolates from animal and human sources. Appl. Environ. Microbiol. 68(6): 2690- 2698.

Hedren, M., M. F. Fay, and M. W. Chase. 2001. Amplified fragment length polymorphisms (AFLP) reveal details of polyploid evolution in *Dactylorhiza* (Orchidaceae). Amer. J. Bot. 88: 1868-1880.

Henthorne, F. 2000. On the propagation and cultivation of *Echinacea*. Pages 1-5. http://thetillery.bizland.com/Echinacea propagation. html

Hill, M., H. Witsenboer, M. Zabeau, P. Vos, R. Kesseli, and R. Michelmore. 1996. PCR based fingerprinting using AFLP as a tool for studying genetic relationships in *Lactuca spp*. Theor. Appl. Genet. 93: 1202-1210.

Hoarau, J. Y., L. Grivet, B. Offmann, L. M. Raboin, J. P. Diorflar, J. Payet, M. Hellmann, A. D'Hont, and J. C. Glaszmann. 2002. Genetic dissection of modern sugarcane cultivars (*Saccharum* spp.) II. Detection of QTLs for yield components. Theor Appl Genet. 105: 1027-1037.

Hobbs C. 1994. *Echinacea*. A literature review. Botany, history, chemistry, pharmacology, toxicology, and clinical uses. Herbalgram. 30: 33-48.

Hodkinson, T. R., M. W. Chase and S. A. Renvoize. 2002. Characterization of a genetic resource collection for Miscanthus (Sacharinae, Andropogoneae, Poaceae) using AFLP and ISSR PCR. Ann. Bot. (Lond.). 89: 627-36.

Huff, D. R., R. Peakall, and P. E. Smouse. 1993. RAPD variation within and among natural populations of out crossing buffalograss [*Buchloë dactyloides* (Nutt.) Engelm.] Theor. Appl. Genet. 86: 927-934.

Innan, H., R. Terauchi, G. Kahl, and F. Tajima. 1999. A method for estimating nucleotide diversity from AFLP data. Genetics. 151: 1157-1164.

Ipek, M. and P. Simon. 2003. Sequence characterization pf polymorphic AFLP fragments in garlic (*Allium sativum* L.). Plant & Animal Genome XI. Final Abstracts Guide: 245. Poster abstract. P685.

Jones, C. J., K. J. Edwards, S. Castaglione, M. O. Winfield, F. Sala, C. van de Wiel, G. Bredemeijer, B. Vosman, M. Matthes, A. Daly, R. Brettschneider, P. Bettini, M. Buiatti, E. Maestri, A. Malcevschi, N. Marmiroli, R. Aert, G. Volckaert, J. Rueda,

R. Linacero, A. Vasquez, and A. Karp. 1997. Reproducibility testing of RAPD, AFLP, and SSR markers in plants by a network of European laboratories. Molecular Breeding. 3: 381-390.

Joshi, M. and U. Dhar. 2003. In vitro propagation of *Saussurea obvallata* (DC.) Edgew. An endangered ethno religious medicinal herb of Himalaya. Plant Cell Reports. 21: 22-29.

Kapteyn, J., P. B. Goldsbrough. and J. E. Simon. 2002. Genetic relationships and diversity of commercially relevant *Echinacea* species. Theor. Appl. Genet. 105: 369-376.

Kardolus, J. P., H. J. van Eck, and G. Van Den Berg. 1998. The potential of AFLP's in biosystematics: a first application in *Solanum* taxonomy (Solanaceae). Plant Syst. Evol. 210: 87-103.

Karp, A., K. J. Edwards, M. Bruford, S. Funk and B. Vosman. 1997. Molecular technologies for bio diversity evaluation: opportunities and challenges. Nature Biotech. 15: 625-628.

Keim, P., A. Kalif, J. Schupp, K. Hill, S. E. Travis, K. Richmond, M. Adair, H. Jones, C. R. Kuske, and P. Jackson. 1997. Molecular evolution and diversity in *Bacillus anthracis* as detected by amplified fragment length polymorphism markers. J. Bacteriol. 179: 818-824.

Kigler, B. 2002. *Echinacea*. American Fam. Physician. 67: 77-80.

Kim, M. S., P. H. Moore, F. Zee, M. M. Fitch, D. L. Steiger, R. M. Manshardt, R. E. Paull, R. A. Drew, T. Sekioka and R. Ming. 2002. Genetic diversity of *Cariaca papaya* as revealed by AFLP markers. Genome. 45: 503-512.

Kim, D. H and D. W. Still. 2003. Genetic diversity and phenetic relationships among *Echinacea* species. Plant & Animal Genome XI. Final abstracts guide: 132. Poster abstract. P229.

Kim, D. H., D. Heber and D. W. Still. 2004. Genetic diversity of *Echinacea* species based upon amplified fragment length polymorphism markers. Genome. 47: 102-111.

Klecka, W. R. 1984. Discriminant analysis. Series: Quantitative applications in the Social Sciences. SAGE university paper 19. Sage publications.

Knorr, C., H. H. Cheng, and J. B. Dodgson. 2001. DNA cloning and sequence analysis of chicken AFLP. Anim Genet. 32(3): 156-9.

Koopman, W. J. M. 2000. Identifying lettuce species (*Lactuca* subsect. *Lactuca*, Asteraceae): a practical application of flow cytometry. Euphytica, 116: 151-159.

Koopman, W. J. M. and J. H. De Jong. 1996. A numerical analysis of karyotypes and DNA amounts in lettuce cultivars and species (*Lactuca* subsect. *Lactuca*, Compositae). Acta Bot. Neerl. 45: 211-222.

Koopman, W. J. M., M. J. Zevenbergen, and R. G. Van Den Berg. 2001. Species relationships in *Lactuca* S. L. (Lactuceae, Asteraceae) inferred from AFLP fingerprinting. Am. J. Bot. 88(10): 1881-1887.

Koroch; A., H. R. Juliani, J. Kapteyn, and J. E. Simon. 2002. *In vitro* regeneration of *Echinacea purpurea* from leaf explants. Plant Cell, Tissue, and Organ Culture. 69(1): 79-83.

Krzanowski, W. J. 1975. Discrimination and classification using both binary and continuous variables. J. Am. Stat. Assoc. 70: 782-790.

Kshirsagar, A. M. 1972. Multivariate analysis. Marcel Dekker, New York.

Lachenbruch, P. A. and M. A. Mickey. 1968. Estimation of error rates in discriminant analysis. Technometrics. 10: 1-10.

Lakshmanann, P., M. Danesh, and A. Taji. 2002. Production of four commercially cultivated *Echinacea* species by different methods of *in vitro* regeneration. The J. of Hort. Sci. and Biotech. 77:158 -- 163

Lamboy, W. 1994. Computing genetic distance similarity coefficient from RAPD data, the effects of PCR artifacts. PCR methods and applications. 4: 31-37.

Lee, H. S. and Chen, Z. J. 2001. Protein-coding genes are epigenetically regulated in Arabidopsis polyploids. Proc. Natl. Acad. Sci. 98: 6753-6758.

Lefebvre, V., B. Goffinet, J. C. Chauvet, B. Caromel, P. Signoret, R. Brand, and A. Pallois. 2001. Evaluation of genetic distances between Pepper inbred lines for cultivar protection purposes; comparison of AFLP, RAPD and phenotypic data. Theor. Appl. Genet. 102: 741-750.

Letchamo, W., L. V. Polydeonny, N. O. Gladisheva, T. J. Arnason, J. Livesey and D. V. C. Awang. 2002 Factors affecting Echinacea quality. Trends in new crops and new uses. Pages, 514-521. J. Janick and A. Whipkey (eds.). ASHS Press, Alexandria, VA.

Letchamo, W., J. Livesey, T. J. Arnason, C. Bergeron and V. S. Krutilina. 1999. Cichoric acid and Isobutyl amide content in *Echinacea purpure*a as influenced by flower developmental stages. pp. 494-498. In: Perspectives on new crops and new uses. J. Janick (eds.). ASHS Press. Alexandria. Va. USA.

Levieille, G. and G. Wilson. 2002. In vitro propagation and iridoid analysis of the medicinal species *Harpagophytum procumbens* and *H. zeyheri*. Plant Cell Reports. 20: 1-11.

Lombard, V., C. P. Baril, and D. Zhang. 2000. Genetic relationships and fingerprinting of rapeseed cultivars using AFLP. Consequences for varietal registration. Crop Sci. 23: 84-85.

Lloyd, G., and B. McCowan. 1981. Woody Plant Medium for micro propagation. Int. Propagation Soc. Proceedings. 30: 421-427.

Lynch, M., and J. S. Conery. 2003. The origins of genome complexity. Science. 302: 1401-1404.

Mace, E. S., R. N. Lester, and C. G. Gebhardt. 1999a. AFLP analysis of genetic relationships among the cultivated eggplant *Solanum melongena* L. and wild relatives (Solanaceae). Theor. Appl. Genet. 99: 626-633.

Mace, E. S., R. N. Lester, and C. G. Gebhardt. 1999b. AFLP analysis of genetic relationships in the tribe Daturaea (Solanaceae). Theor. Appl. Genet. 99: 634-641.

Makunga, N. P., A. K. Jager, and J. van Staden. 2003. Micropropagation of *Thapsia garganica* a medicinal plant. Plant Cell Reports. 21: 1-14.

Mantel, N. 1967. The detection of disease clustering and a generalised regression approach. Cancer Research. 27: 209-220.

McGregor, R. L. 1968. The taxonomy of the genus *Echinacea* (Compositae). Univ. of Kansas Sci. Bul. 48: 113-142.

McGregor, C. E., R. van Treuren, R. Hoekstra, and Th. J. L. van Hintum. 2002. Analysis of the wild potato germplasm of the series *Acaulia* with AFLPs: implications for *ex situ* conservation. Theor. Appl. Genet. 104: 146-156.

McKeown, K. A. 1999. A review of the taxonomy of the genus *Echinacea*. Perspectives on new crops and new uses. J. Janick (ed.) ASHS Press, Alexandria. VA. USA.

McLenachan, P. A., K. Stockler, R. C. Winkworth, K. McBreen, S. Zauner, and P. J. Lockhart. 2000. Markers derived from amplified fragment length polymorphism gels for plant ecology and evolution studies. Mol. Ecol. 9: 899-903.

Mechanda, S. M., B. R. Baum, D. A. Johnson, and J. T. Arnason. 2004a. Sequence assessment of comigrating AFLP bands in *Echinacea* – implications for comparative studies. Genome. 47: 15-25.

Mechanda, S. M., B. R. Baum, D. A. Johnson, and J. T. Arnason. 2003. Direct shoot regeneration from leaf segments of mature plants of *Echinacea purpurea* (L.) Moench. In vitro cellular and developmental biology-Plant. 39: 505-510.

Mechanda, S. M., B. R. Baum, D. A. Johnson, and J. T. Arnason. 2004b. Analysis of diversity of natural populations and commercial lines of *Echinacea*, using AFLP. Canadian J. Bot. 82: 461-484.

Meksem, K., E. Ruben, D. Hyten, K. Triwitayakorn, and D. A. Lightfoot. 2001. Conversion of AFLP bands into high-throughput DNA markers. Mol. Genet. Genomics. 265(2): 207-14.

Mengistu, I. W., G. W. Muller-Warrant, and R. E. Barker. 2000. Genetic diversity of *Poa annua* in western Oregon grass seed crops. Theor. Appl. Genet. 101: 70-79.

Mignouna, D., A. Mank, N. Ellis, N. Van Den Bosch, R. Asiedu, M. Abang, and J. Peleman. 2002. A genetic linkage map of water yam (*Dioscorea alata* L.) based on AFLP markers and QTL analysis for anthracnose resistance. Theor. Appl. Genet. 105: 726-735.

Miller, M. P. 1997. Tools for Population Genetic Analysis. (TFPGA). ver. 1.3. A Windows program for the analysis of allozyme and molecular population genetic data. Dept of Biol. Sci. Northern Arizona University, Flagstaff. USA (accessed 2003).

Mueller, U. G., and L. L. Wolfenbarger. 1999. AFLP genotyping and fingerprinting. TREE. 14: 389- 394.

Murashige, T. and F. A. Skoog. 1962. A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol. Plant. 15: 473-497.

Nei, M. 1972. Genetic distance between populations. Am. Nat. 106: 283-292.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proc. Nat. Acad. Sci. USA. 70: 3321-3323.

Nicholas, K. B. and H. B. Jr. Nicholas. 1997. GENEDOC: a tool for editing and annotating multiple sequence alignments. Multiple sequence alignment & shading utility, Version 2.6.002. Distributed by the authors http://www.psc.edu/biomed/genedoc

Nilsson, N. O., M. Hansen, A. H. Panagopoulos, S. Tuvesson, M. Ehlde, I. M., Christiansson, M. R. Rading, and T Kraft. 1999. QTL analysis of *Cercospora* leaf spot resistance in sugar beet. Plant Breeding. 118: 327-334.

Ogden, R. and R. S. Thorpe. 2002. The usefulness of amplified length polymorphism markers for taxon discrimination across graduated fine evolutionary levels in Caribbean Anolis lizards. Mol. Ecol. 11: 437-45.

Osman, A., B. Jordan, P. A. Lessard, N. Muhammad, M. R. Haron, N. M. Rifffin, A. J. Sinskey, C. Rha and D. E. Housman. 2003. Genetic diversity of *Eurycoma longifolia* inferred from single nucleotide polymorphisms. Plant Physiol. 131: 1294-1301

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

Peakall, R., P. E. Smouse, and D. R. Huff. 1995. Evolutionary implications of allozyme and RAPD variation in diploid populations of dioeocious buffalo grass *Buchloë dactyloides*. Mol. Ecol. 4: 135-147.

Percival, S. S. 2000. Use of *Echinacea* in medicine. Biochem. Pharmacol. 60: 155-158.

Peters, J. L., H. Constandt, P. Neyt, G. Cnops, J. Zethof, M. Zabeau, and T. Gerats. 2001. A physical amplified fragment-length polymorphism map of *Arabidopsis*. Plant Physiol. 127: 1579-1589.

Pimentel, R. A. 1979. Morphometrics. The multivariate analysis of biological data. Kendall/Hunt, Dubuque.

Potokina, E., F. R. Blattner, T. Alexandrova and K. Bachmann. 2002. AFLP diversity in the common vetch (*Vicia sativa* L.) on the world scale. Theor. Appl. Genet. 110: 1-6.

Powell, W., M. Moragante, C. Andre, M. Hanafey, J. Vogel, S. Tingey, and A. Rafalski. 1996. The comparison of RFLP, RAPD, AFLP and SSR markers for germplasm analysis. Mol. Breeding. 2(3): 225-238.

Prim, R. C. 1957. Shortest connection networks and some generalizations. Bell Sys. Tech. J. 36. 1389-1401.

Quagliaro, G., M. Vischi, M. Tyrka, and A. M. Olivieri. 2001. Identification of wild and cultivated Sunflower for breeding purposes by AFLP markers. J. Heredity. 92(1): 38-42.

Raamsdonk, L. W. D., M. Vrielink van Ginkel, and C. Kik. 2000. Phylogeny reconstruction and hybrid analysis in *Allium* subgenus Rhizirideum. Theor. Appl. Genet. 100: 1000-1009.

Rafalski, A. 2002. Applications of the single nucleotide polymorphism in crop genetics. Curr. Opin. Plant Biol. 5: 94-100.

Rao, C. R. 1952. Advanced statistical methods in biometric research. Wiley, New York. USA.

Reamon Buttner, S. M., T. Schmidt, and C. Jung. 1999. AFLPs represent highly repetitive sequences in *Asparagus officinalis* L. Chromosome Res. 7(4): 297-304.

Ren, N. and M. P. Timko. 2001. AFLP analysis of genetic polymorphism and evolutionary relationships among cultivated and wild *Nicotiana* species. Genome. 44: 559-571.

Rieseberg, L. H. 1996. Homology among RAPD fragments in interspecific comparisons. Mol. Ecol. 5: 99-105.

Robinson, J. P. and S. A. Harris. 1999. Amplified Fragment Length Polymorphisms and Microsatellites: A phylogenetic perspective. Chapter 12. In: Which DNA Marker for Which Purpose? 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. URL: http://webdoc.sub.gwdg.de/ebook/y/1999/whichmarker/index.htm

Rohlf, F. J. 2000. NTSYSpc. Numerical Taxonomy and Multivariate Analysis System. Ver. 2.02 (2002). Exeter Software. Setauket, N. Y. USA.

Rouppe Van Der Voort, J. N. A. M. P. Van Zandvoort, H. J. Van Eck, R. T. Folkertsma, R. C. B. Hutten, J. Draaistra, F. J. Gommes, E. Jacobsen, J. Helder, and J. Bakker. 1999. Use of allele specificity of comigrating AFLP markers to align genetic maps from different potato genotypes. Mol. Gen. Genet. 255: 438-447.

Roldán-Ruiz, I., van Euwijk, F. A., Gilliland, T. J., Dubreuil, P., Dillmann, C., Lallemand, J., De Loose, M., and Baril, C. P. 2001. A comparative study of molecular and morphological methods of describing relationships between perennial ryegrass (*Lolium perenne* L.) varieties. Theor. Appl. Genet. 103: 1138-1150.

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

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

Saitou, N. and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4: 406-425.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory press. Cold Spring Harbor. NY. 3 volumes.

SAS Institute. 2002. SAS/STAT® user's guide. Version 8.2, SAS Institute. Cary NC. USA.

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

Schneider, S., D. Roessli, and L. Excoffier. 2001. Arlequin: software for population genetics data analysis. ver 2.001. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva, Switzerland. URL:http://anthro.unige.ch/arlequin.

Semagn, K., A. Bjornstad, B. Stedje, and E. Bekele. 2000. Comparison of multivariate methods for the analysis of genetic resources and adaptation in *Phytolacca dodecandra* using RAPD. Theor Appl. Genet. 101: 1145-1154.

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

Snow, R. 1963. Alcoholic hydrochloric acid-carmine as a stain for chromosomes in root tips squash preparations. Stain Technol. 38: 9-13.

Sokal, R. R. and C. D. Michener. 1958. A statistical method for evaluating systematic relationships. Univ. Kansas Sci. Bull. 38: 1409-1438.

Soleimani, V. D., B. R. Baum, and D. A. Johnson. 2002a. AFLP and pedigree-based genetic diversity estimates in modern cultivars of durum wheat [*Triticum turgidum* L. subsp. *durum* (Desf.) Husn.]. Theor. Appl. Genet. 104: 350-357.

Soleimani, V. D., B. R. Baum, and D. A. Johnson. 2002b. Identification of Canadian durum wheat [*Triticum turgidum* L. subsp. Durum (Desf.) Husn.] cultivars using AFLP and their STS markers. Can. J. Pl. Sci. 82: 35-41.

Soltis, P. S. and Soltis, D. E. 2000. The role of genetic and genomic attributes in the success of polyploids. Proc. Natl. Acad. Sci. 97:7051-7057.

Steiger, B. L., C. Nagai, P. H. Moore, C. W. Morden, R. V. Osgood, and R. Ming. 2002. AFLP analysis of genetic diversity within and among *Coffea arabica* cultivars. Theor. Appl. Genet. 105: 209-215.

Stoeckle, M. 2003. Taxonomy, DNA, and the bar code of life. BioScience. 53: 2-3.

Strommer, J., J. Peters, J. Zethof, P. De Keukeleire, and T. Gerats. 2002. AFLP maps of *Petunia hybrida*: building maps when markers cluster. Theor. Appl. Genet. 105: 1000-1009

Subbaiah, M. M. and S. C. Minocha. 1990. Shoot regeneration from stem and leaf callus of *Eucalyptus teriticornis*. Plant Cell Rep. 9: 370-373.

Swafford, D. L. 1999. PAUP: phylogenetic analysis using parsimony: Smithsonian Institution, Washington, D. C. USA.

Thompson, J. D., D. G. Higgins, and T. J. Gibbons. 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.

Thormann, C. E., M. E. Ferreira, L. E. A. Carmago, J. G. Tivang and T. C. Osborn.1994 Comparison of RFLP and RAPD markers to estimate genetic relationships within and among cruciferous species; Theor. Appl. Genet. 88: 973–980.

Tomkins, J. P., T. C. Wood, L. S. Barnes, A. Wetman, and R. A. Wing. 2001. Evaluation of genetic variation in the daylily (*Hemerocallis spp.*) Using AFLP markers. Theor. Appl. Genet. 102: 489-496.

Tsuji, S. A., K. Itoh, S. Sasazaki, H. Mannen, K. Oyama, F. Shojo, and M. Mukai. 2004. An association study using AFLP markers and application to a beef cattle breeding population. J. Anim. Genet. 35: 40-43

Tyler, V. E. 1993. Phytomedicines in Western Europe. Potential impact on herbal medicine in the United States. In A. D. Klinghorn and M. F. Balandrin, (eds). Human medicinal agents from Plants. ACS series No. 534. pp. 25-37.

U.S. Fish and Wildlife service Threatened and Endangered Species System (TESS) http://www.fws.gov. (Web page accessed Jan. 9, 2004).

U.S. Fish and Wildlife service U. S. Listed Flowering Plant species Index by Lead Region and Status (Web page accessed Jan. 9, 2004). http://www.fws.gov

Urbatsch, L. E., and R. K. Jansen. 1995. Phylogenetic affinities among and within the cone flower genera (Asteraceae, Heliantheae): a chloroplast DNA analysis. Syst. Bot. 20(1): 28-39.

Urbatsch, L. E, B. G. Baldwin and M. J. Donoghue. 2000. Phylogeny of the coneflowers and relatives (Helaitheae: Asteraceae) based on nuclear rDNA internal transcribed spacers (ITS) sequences and chloroplast DNA restriction site data. Systematic Botany. 25. 539-565.

Virk P. S., H. J. Newbury, M. T. Jackson, and B. V. Ford Lloyd. 2001. Are mapped markers more useful for assessing genetic diversity? Theor. Appl. Genet. 100: 607-613.

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

Watrous, L. E. and Q. D.Wheeler. 1981. The outgroup comparison method of character analysis. Syst. Zool. 30: 1-11.

Waugh, R., N. Bonar, E. Baird, B. Thomas, A. Garner, P. Hayes, and W. Powell. 1997. Homology of AFLP products in three mapping populations of barley. Mol. Gen. Genet. 255: 311-321.

Weir, B. S. 1996. Genetic data analysis II. Methods for discrete population genetic data. Sinauer Associates, Sunderland, Mass. USA.

Wiens, J. J. 2001. Character analysis in morphological phylogenetics: problems and solutions. Syst. Biol. 50: 689-699.

Wilkinson, T., A. Wetten, C. Prychid, and M. F. Fay. 2003. Suitability of cryopreservation for the long-term storage of rare and endangered plant species: a case history for *Cosmos atrosanguineus*. Annals of Bot. 91: 68-74.

Williams, J. G. K., A. R. Kubelik, K. J. Levak, J. A. Rafalski, and S. V. Tingey. 1990. DNA polymorphism amplification by arbitrary primers is useful as markers. Nucleic Acids Res. 18: 6531-6535.

Winfield, M. O. 1998. A study of genetic diversity in *Populus nigra* in the upper Severn area of the UK using the AFLP markers. Mol. Ecol. 7(1): 3-10.

Wong, A., M. R. Forbes, and M. L. Smith. 2001. Characterization of AFLP markers in damselflies: prevalence of co-dominant markers and implications for population genetic applications. Genome. 11: 677-684.

Yeh, F. C., and T. Boyle. 1997. Population genetic analysis of co-dominant and dominant markers and quantitative traits. Belgian J. Bot. 129: 157-163.

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

Zerega, N. J. C., S. Mori, C. Lindqvist, Q. Zheng, and T. J. Motley. 2002. Using Amplified Fragment Length Polymorphisms (AFLP) to Identify Black Cohosh (*Actaea racemosa*). Economic Botany. 56: 154–164.

APPENDIX. 1.

Raw AFLP data. (see CD attached)

APPENDIX. 2.

Original Phylogram (see CD attached)