Metabolic Profiling of Echinacea Genotypes and a Test of Alternative Taxonomic Treatments

Lankun Wu
Iowa State University

Philip M. Dixon
Iowa State University, pdixon@iastate.edu

Basil J. Nikolau
Iowa State University, dimmas@iastate.edu

George A. Kraus
Iowa State University, gakraus@iastate.edu

Mark P. Widrlechner
United States Department of Agriculture, isumw@iastate.edu

Follow this and additional works at: http://lib.dr.iastate.edu/ncrpis_pubs

Part of the Agricultural Science Commons, Agriculture Commons, Agronomy and Crop Science Commons, Biochemistry Commons, Biostatistics Commons, Genetics Commons, Horticulture Commons, Molecular Biology Commons, Organic Chemistry Commons, and the Plant Breeding and Genetics Commons

The complete bibliographic information for this item can be found at http://lib.dr.iastate.edu/ncrpis_pubs/3. For information on how to cite this item, please visit http://lib.dr.iastate.edu/howtocite.html.
Authors
Lankun Wu, Philip M. Dixon, Basil J. Nikolau, George A. Kraus, Mark P. Widrlechner, and Eve S. Wurtele
Metabolic Profiling of *Echinacea* Genotypes and a Test of Alternative Taxonomic Treatments

**Author**

Lankun Wu¹, Philip M. Dixon², Basil J. Nikolau³, George A. Kraus⁴, Mark P. Widrlechner⁵, Eve Syrkin Wurtele¹

**Affiliation**

¹ Department of Genetics, Development, and Cell Biology, Iowa State University, Ames, Iowa, USA
² Department of Statistics, Iowa State University, Ames, Iowa, USA
³ Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, Iowa, USA
⁴ Department of Chemistry, Iowa State University, Ames, Iowa, USA
⁵ USDA-ARS North Central Regional Plant Introduction Station, Ames, Iowa, USA

**Key words**

- *Echinacea* (Asteraceae)
- alkamides
- alkylamides
- HPLC
- GC-MS
- PCA

**Abstract**

The genus *Echinacea* is used as an herbal medicine to treat a variety of ailments. To better understand its potential chemical variation, 40 *Echinacea* accessions encompassing broad geographical and morphological diversity were evaluated under controlled conditions. Metabolites of roots from these accessions were analyzed by HPLC-photo diode array (HPLC-PDA), GC-MS, and multivariate statistical methods. In total, 43 lipophilic metabolites, including 24 unknown compounds, were detected. Weighted principal component analysis (WPCA) and clustering analysis of the levels of these metabolites across *Echinacea* accessions, based on Canberra distances, allowed us to test two alternative taxonomic treatments of the genus, with the further goal of facilitating accession identification. A widely used system developed by McGregor based primarily on morphological features was more congruent with the dendrogram generated from the lipophilic metabolite data than the system more recently developed by Binns et al. Our data support the hypothesis that *Echinacea pallida* is a diverse allopolyploid, incorporating the genomes of *Echinacea simulata* and another taxon, possibly *Echinacea sanguinea*. Finally, most recognized taxa of *Echinacea* can be identified by their distinct lipophilic metabolite fingerprints.

**Supporting information** available online at http://www.thieme-connect.de/ejournals/toc/plantamedica

**Introduction**

*Echinacea* extracts, particularly from roots, have historically been used by Native Americans and, more recently, by Western cultures as herbal remedies to treat ailments ranging from snake bites to pain, burns, cough, sore throats, and toothache [1], [2], [3]. *Echinacea* products are promoted for immune system enhancement and are among the best-selling herbal preparations in the United States [4]. Many unusual phytochemicals have been found in *Echinacea*, some with reports of bioactivity in animal cells or animals, including alkamides and ketones, caffeic acid derivatives, glycoproteins, and unusual polysaccharides (as reviewed by Bauer [5]). Despite its popularity as an herbal dietary supplement, and many pertinent pharmacological and clinical studies, little is known regarding the specific compounds primarily responsible for observed bioactivity or whether they are consistently efficacious in humans [6].

Although morphologically distinct *Echinacea* groups exhibit differences in phytochemical composition [5] and bioactivity [7], most studies have focused on three major medicinal species: *Echinacea angustifolia* DC., *E. pallida* (Nutt.) Nutt., and *E. purpurea* (L) Moench. The most comprehensive study to date was conducted by Binns et al. [8], who reported on phytochemical variation in *Echinacea* from roots and capitula of wild and cultivated populations representing all nine *Echinacea* species recognized by McGregor [9]. However, in that report, most alkamides and ketones were identified by UV spectra and relative retention times compared with one major pair of alkamide standards. Baum et al. [10], in a recent review of the status of *Echinacea* systematics and phytochemistry, indicated that *Echinacea* taxa are readily distinguishable on the basis of HPLC profiles and that HPLC profiles for lipophilic compounds contain more information than those based on caffeic acid derivatives. This is noteworthy because of the existence of two al-
alternative taxonomic treatments for Echinacea. The older, developed by McGregor [9], was based on field observations, common garden studies, and cytological and anatomical analyses. McGregor's classification has been widely used by botanists and herbalists [11] and serves as the basis for the recent Flora North America treatment [12]. Binns et al. [13] proposed a revision, based on morphometric data and phytochemical data from greenhouse-grown and wild plants, using canonical discriminant and cladistic analyses. This revision recognizes all but one of McGregor's taxa, the most significant changes being a reduction in the number of species, an increase in the number of varieties, and, in particular, the incorporation of 5 morphologically diverse clades characterized as species by McGregor [9] into a single species. This Binns et al. [13] revision is controversial in the botanical community [11].

The current DNA-based molecular marker evidence is not yet refined sufficiently to generate accession-level systematics [14, 15]. Even an extensive DNA-sequencing study using these identical accessions and based on multiple loci has thus far been unable to completely evaluate systematic relationships among these accessions [16].

Here we have taken a targeted, metabolite-profiling approach to investigate the accumulation of putatively bioactive alkamides and ketones in 40 geographically and morphologically diverse Echinacea populations, which already had been well characterized morphologically and as to origin. We used as standards authentic alkamides and ketones that were purchased or synthesized by our group [17], [18], [19], as well as structural information obtained by a combination of HPLC-PDA and GC-MS, for more comprehensive compound identification. Because we lacked reference standards for many of the metabolites reported in this study (43 in all), we used relative instead of absolute metabolite concentrations to compare overall lipophilic-metabolite profiles across the 40 accessions. This approach, coupled with weighted principal component analysis (WPCA) and clustering analysis based on Canberra distances [20], provides an opportunity to test these two taxonomic classifications. Furthermore, these metabolic profiles may help standardize Echinacea products, characterize plant material of unknown provenance, and identify genetic sources to select for increased production of desired compounds.

Materials and Methods

Plant materials
We selected 40 well-characterized Echinacea accessions (Table 1) representing a broad geographic and morphological sampling of the germplasm conserved by the U.S. National Plant Germplasm System, USDA-ARS North Central Regional Plant Introduction Station, Ames, Iowa [21]. Initially we looked at roots of two ages of plants: 6-month-old and 3-year-old. We found that the relative levels of metabolites vary, but the same identified alkamides, ketones, and unknown metabolites are present at both ages (Fig. 15, Supporting Information); therefore, we focused on 6-month-old plants because we are able to grow them under well-controlled conditions. Characterization data for a broad range of (> 40) morphological traits are available at Germplasm Resources Information Network database (http://www.ars-grin.gov/cgi-bin/npgs/html/desc_form.pl?221). Accessions were keyed to species (or subspecies) during initial regeneration on the basis of McGregor [9], and we converted McGregor identifications to the treatment of Binns et al. [13] via Table 2 in [10]. Growth conditions and sampling methods are available in the Supporting Information.

Extraction, HPLC, and GC-MS analysis
Plant extraction, HPLC, and GC-MS analyses were performed as in [17], [18], and [22].

Compound identification and relative abundance
In addition to 19 known alkamides and ketones, another 24 unknown lipophilic metabolites were detected and grouped according to their retention times and UV spectra (Table 15, Supporting Information). Methods for determining the relative abundance of metabolites are provided in Supporting Information.

Statistical analysis
Multivariate analyses by WPCA and hierarchical clustering analysis were performed in R software, version 2.2.1 (http://www.r-project.org/). Detailed information is provided in the Supporting Information.

Supporting information
Detailed methods and additional data are available as Supporting Information.

Results and Discussion

Most Echinacea extractions use dry materials and rigorous methods that last 1–24 hours (e.g., soxhlet extraction [23], ultrasonic extraction [24], and shaking [8]). To minimize possible degradation of unstable metabolites during extraction, we used a quick extraction method by powderizing a small amount of fresh tissue with liquid N2 and extracting at low temperature [17], [18]. By using authentic synthesized standards, combined with GC-MS and HPLC-PDA, we evaluated 40 accessions (Fig. 11) and detected 43 UV-absorbing lipophilic metabolites. Of these, 19 metabolites were identified, including 16 alkamides and 2 ketones among those reported so far by the pioneering studies of Bauer and colleagues [5] and another recently reported alkamide, herein referred to as “Chen alkamide” [25] (for structures, see Fig. 25, Supporting Information). In addition, we detected 24 unknown lipophilic metabolites, some of which (e.g., unknown B5 and unknown 9) are relatively abundant. Five unknowns (unknowns A1-A5) have UV spectra similar to the 2,4-diene alkamides (1, 2, 3, 4, 7, 10, and 11). Six unknowns (B1-B6) have UV spectra similar to the monoene alkamides (12, 13, 14, 16, and 17) (Table 15, Supporting Information). Interestingly, 10 unknown metabolites have atypical UV spectra. Most of the unknowns (3–10) are highly lipophilic and thus elute at later times. Identification of these unknowns is currently being conducted by HPLC-tandem mass spectrometry (LC-MS/MS), semi-preparative HPLC, and NMR.

Our observation of 43 lipophilic metabolites (Table 15, Supporting Information) can be contrasted with the findings of Binns et al. [8], who distinguished 15 unique alkamides, 2 pairs of alkamides (alkamide 8/9 and alkamide 5/15), and 3 ketones. We detected the presence of all but 4 of these, 3 of the alkamides and 1 of the ketones. Our ability to distinguish more than twice the compounds from these samples is likely attributable to two factors: rapid extraction under low temperature, minimizing possible degradation, and a more sensitive HPLC separation method with extended retention times.

To elucidate how Echinacea accessions are related in terms of their overall metabolite profiles, two multivariate statistical approaches were used: WPCA and clustering analysis. Traditional
PCA assumes that all observations of a particular metabolite have the same variance, although variances may differ between metabolites. However, for our dataset, standard deviations among biological replicates increase with metabolite abundance, i.e., the abundant compounds are more variable (Fig. 2). We used WPCA to account for this pattern. Each element of the data is given a corresponding “weight”, proportional to the inverse of the variance. Thus, smaller peaks with smaller errors are given a larger weight, placing more emphasis on these less abundant compounds.

WPCA reveals large quantitative and qualitative differences in lipophilic metabolites among Echinacea populations (Fig. 3). Each composite metabolite profile indicates metabolites that tend to be present together or absent together (Fig. 3A). The accession profiles indicate the relative abundance of corresponding metabolite profiles for each accession (Fig. 3B). Profile 1 (WPC1) focuses on the most abundant compounds, primarily amides 8 and 9, with smaller amounts of amides 3, 5, 11, and 12. Profile 1 is most abundant in accessions identified on the basis of McGregor’s (9) two described varieties E. angustifolia and E. purpurea (Nut.) Nutt., E. pavonia, and E. sanguinea. Profile 2 (WPC2) contains predominantly amides 2 and 3, the Chen alkamide, ketones 22, and 24, and unknown 8. Profile 2 is most abundant in E. purpurea as well as in many other species, except E. angustifolia, E. atrorubens, and E. tennesseensis (Beadle) Small. Profile 3 (WPC3) is almost exclusively composed of amides 12, 13, and 14 and is most abundant in E. angustifolia and E. tennesseensis. The three-profile solution explains 94.8% of the variance of the whole dataset. The hierarchical dendrogram constructed from Canberra distances (20) for all pairs of 40 accessions (3 plants per accession),

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>E. angustifolia</td>
<td>E. pallida var. angustifolia</td>
<td>PI631 267</td>
<td>Murray County, OK</td>
</tr>
<tr>
<td>E. angustifolia var. angustifolia</td>
<td>E. pallida var. angustifolia</td>
<td>PI631 272</td>
<td>Comanche County, OK</td>
</tr>
<tr>
<td>E. angustifolia var. angustifolia</td>
<td>E. pallida var. angustifolia</td>
<td>PI631 285</td>
<td>Lyon County, IA</td>
</tr>
<tr>
<td>E. angustifolia var. angustifolia</td>
<td>E. pallida var. angustifolia</td>
<td>PI631 318</td>
<td>Rooks County, KS</td>
</tr>
<tr>
<td>E. angustifolia var. strigosa</td>
<td>E. pallida var. angustifolia</td>
<td>PI631 266</td>
<td>Murray County, OK</td>
</tr>
<tr>
<td>E. atrorubens</td>
<td>E. atrorubens var. atrorubens</td>
<td>PI631 320</td>
<td>Pontotoc County, OK</td>
</tr>
<tr>
<td>E. atrorubens</td>
<td>E. atrorubens var. atrorubens</td>
<td>PI631 255</td>
<td>Douglas County, KS</td>
</tr>
<tr>
<td>E. atrorubens</td>
<td>E. atrorubens var. atrorubens</td>
<td>PI631 266</td>
<td>Murray County, OK</td>
</tr>
<tr>
<td>E. atrorubens</td>
<td>E. atrorubens var. atrorubens</td>
<td>PI631 299</td>
<td>Osage County, KS</td>
</tr>
<tr>
<td>E. laevigata</td>
<td>E. laevigata</td>
<td>PI631 310</td>
<td>Oconee County, SC</td>
</tr>
<tr>
<td>E. laevigata</td>
<td>E. laevigata</td>
<td>PI631 312</td>
<td>Oconee County, SC</td>
</tr>
<tr>
<td>E. laevigata</td>
<td>E. laevigata</td>
<td>PI631 314</td>
<td>Granville County, NC</td>
</tr>
<tr>
<td>E. pallida</td>
<td>E. pallida var. pallida</td>
<td>PI631 275</td>
<td>Osage County, OK</td>
</tr>
<tr>
<td>E. pallida</td>
<td>E. pallida var. pallida</td>
<td>PI631 290</td>
<td>Sac County, IA</td>
</tr>
<tr>
<td>E. pallida</td>
<td>E. pallida var. pallida</td>
<td>PI631 293</td>
<td>Stone County, AR</td>
</tr>
<tr>
<td>E. pallida</td>
<td>E. pallida var. pallida</td>
<td>PI631 296</td>
<td>Taney County, MO</td>
</tr>
<tr>
<td>E. pallida var. neglecta</td>
<td>E. atrorubens var. neglecta</td>
<td>PI631 263</td>
<td>Murray County, OK</td>
</tr>
<tr>
<td>E. paradoxa var. neglecta</td>
<td>E. atrorubens var. neglecta</td>
<td>PI631 264</td>
<td>Murray County, OK</td>
</tr>
<tr>
<td>E. paradoxa var. neglecta</td>
<td>E. atrorubens var. neglecta</td>
<td>PI631 265</td>
<td>Murray County, OK</td>
</tr>
<tr>
<td>E. paradoxa var. neglecta</td>
<td>E. atrorubens var. neglecta</td>
<td>PI631 292</td>
<td>Stone County, AR</td>
</tr>
<tr>
<td>E. paradoxa var. neglecta</td>
<td>E. atrorubens var. neglecta</td>
<td>PI631 301</td>
<td>Camden County, MO</td>
</tr>
<tr>
<td>E. paradoxa var. neglecta</td>
<td>E. atrorubens var. neglecta</td>
<td>PI631 321</td>
<td>Camden County, MO</td>
</tr>
<tr>
<td>E. purpurea</td>
<td>E. purpurea</td>
<td>PI631 307</td>
<td>Franklin County, MO</td>
</tr>
<tr>
<td>E. purpurea</td>
<td>E. purpurea</td>
<td>PI631 313</td>
<td>Madison County, NC</td>
</tr>
<tr>
<td>E. purpurea</td>
<td>E. purpurea</td>
<td>PI633 669</td>
<td>Caldwell Parish, LA</td>
</tr>
<tr>
<td>E. purpurea</td>
<td>E. purpurea</td>
<td>PI631 257</td>
<td>Vernon Parish, LA</td>
</tr>
<tr>
<td>E. purpurea</td>
<td>E. purpurea</td>
<td>PI631 258</td>
<td>Vernon Parish, LA</td>
</tr>
<tr>
<td>E. purpurea</td>
<td>E. purpurea</td>
<td>PI631 672</td>
<td>Bienville Parish, LA</td>
</tr>
<tr>
<td>E. simulata</td>
<td>E. simulata var. simulata</td>
<td>PI631 249</td>
<td>Grayson County, KY</td>
</tr>
<tr>
<td>E. simulata</td>
<td>E. simulata var. simulata</td>
<td>PI631 304</td>
<td>Franklin County, MO</td>
</tr>
<tr>
<td>E. simulata</td>
<td>E. simulata var. simulata</td>
<td>PI631 308</td>
<td>Rutherford County, TN</td>
</tr>
<tr>
<td>E. tennesseensis</td>
<td>E. tennesseensis var. tennesseensis</td>
<td>PI631 250</td>
<td>Wilson County, TN</td>
</tr>
<tr>
<td>E. tennesseensis</td>
<td>E. tennesseensis var. tennesseensis</td>
<td>PI631 324</td>
<td>Wilson County, TN</td>
</tr>
<tr>
<td>E. tennesseensis</td>
<td>E. tennesseensis var. tennesseensis</td>
<td>PI631 325</td>
<td>Wilson County, TN</td>
</tr>
<tr>
<td>E. tennesseensis</td>
<td>E. tennesseensis var. tennesseensis</td>
<td>PI631 326</td>
<td>Wilson County, TN</td>
</tr>
</tbody>
</table>

* This accession was intermediate in characteristics between McGregor’s (9) two described varieties.

*b/ab = Could not be determined through the application of Baum et al. (13).
based on average linkage (Figs. 4A and 4B). The primary area of agreement between the two taxonomic schemes is the recognition of a distinct species pair, *E. purpurea* and *E. laevigata* (C. L. Boynton & Beadle) S. F. Blake, which share a distinct stem anatomy, leaf shape, and phylary structure, clustered in adjacent groups in Figs. 4C and 4D.

The major difference between the two treatments centers on the remaining taxa, which McGregor [9] treated as seven species with two additional varieties and Binns et al. [13] treated as only two species with six additional varieties. The dendrogram presented in Fig. 4 lends little support for the circumscription of the two, diverse species as proposed by Binns et al. [13], for two primary reasons. First, the observed degree of differentiation that distinguishes *E. laevigata* from *E. purpurea* in Fig. 4, recognized by Binns et al. [13] as a clear distinction at the subgeneric level, would support the recognition of three additional subgenera, which we do not feel is warranted based on all other relevant data. In addition, the two inclusive species recognized by Binns et al. [13] as *E. pallida* and *E. atrorubens* cluster in an intercalated fashion within the dendrogram, above and below the cluster containing *E. laevigata* and *E. purpurea*. Thus, lipophilic metabolic profiles do not support the broad species combinations proposed by Binns et al. [13].

In contrast, there is a better correspondence to the species and at least one of the varieties recognized by McGregor [9] (Fig. 4), Of the nine species recognized by McGregor [9], accessions from...
six cluster together into a single branch in the tree: *E. laevigata*, *E. purpurea*, *E. tennesseensis*, *E. sanguinea*, *E. atrorubens*, and *E. angustifolia*. The three that do not are *E. pallida*, *E. simulata*, and *E. paradoxa* (Norton) Britton.

For *E. pallida*, three of five accessions cluster together (Fig. 4), but accessions PI 631315 and PI 631275 do not. Interestingly, the three *E. pallida* accessions that are clustered together are adjacent to *E. simulata*, which McGregor [9] considered to be very close to *E. pallida* and a likely progenitor of this tetraploid species. He hypothesized that the other species involved in the parentage of *E. pallida* was *E. sanguinea*; the two “atypical” accessions are located on our dendrogram closer to *E. sanguinea* than to *E. simulata*. Support for a close biochemical relationship between *E. pallida* and *E. sanguinea* was reported recently by Senchina et al. [26], who conducted a phenetic analysis of the immunomodulatory characteristics of seven *Echinacea* taxa. Binns et al. [13] treated both *E. simulata* and *E. sanguinea* as varieties of *E. pallida*.

For *E. paradoxa*, the two varieties recognized by McGregor [9] form clean clusters adjacent to each other (Fig. 4). Taken as a single group, the *E. paradoxa* “cluster” also includes two putative hybrid accessions, both of which were collected from populations where *E. paradoxa* was sympatric with other taxa (Table 1), and *E. angustifolia* var. *strigosa* McGregor PI 631320. Five of six accessions of *E. angustifolia* cluster together, the outlier being *E. angustifolia* var. *strigosa* PI 631320. We speculate that *E. angustifolia* var. *strigosa* occupies some “hybrid middle ground” between *E. paradoxa* var. *neglecta* McGregor and *E. angustifolia* var. *angustifolia* (supported geographically) and/or that *E. angustifolia* var. *strigosa* is not well differentiated based on lipophilic compounds. Variety *strigosa* has been recognized as problematic by other researchers as well. McGregor [9] considered it to be of hybrid origin, as did Binns et al. [13], and the conversion table presented by Baum et al. [10] does not recognize it as a distinct taxon nor does *Flora North America* [12].

In general, the dendrogram generated on the basis of Canberra distances for lipophilic metabolite profiles among our 40 accessions supports the taxonomic treatment presented by McGregor [9], with the possible exception of *E. angustifolia* var. *strigosa*. The metabolic profiles also indicate that there are diverse chemotypes of *E. pallida*, consistent with its proposed alloploid origin. Although we sampled a broad geographic distribution of accessions representing each taxon, these accessions generally clustered consistently with taxa as identified by morphological, anatomical, and cytological characteristics used by McGregor [9] rather than by geographic or environmental variables gleaned from their provenance data. Thus, our analyses imply that the distribution and types of alkaloid and ketone metabolites in *Echinacea* do not evolve in a convergent manner in response to particular geo/environmental conditions.

Finally, the relative concentrations of the 43 lipophilic compounds appear to be distinctive enough by taxon to allow us to develop “typical” profiles for *Echinacea* fingerprinting, which could be validated by evaluating additional populations. Thus,
this research expands the basis for the evaluation, standardization, and identification of plant material of unknown provenance for cultivated Echinacea and for commercial Echinacea products. In addition, these data contribute to the identification of genetic resources for the production of specific alkamides and ketones. In the course of this study, we found more than 20 unidentified metabolites, some of which may be alkamides or ketones. Bioactivity-guided fractionation together with compound identification will elucidate these as yet unidentified metabolites.

Acknowledgements

This publication was made possible by grant number 9 P50 AT004155-06 from the National Center for Complementary and Alternative Medicine (NCCAM) and was supported by the Hatch Act and State of Iowa funds to the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa, Project No. 1018. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NCCAM or NIH. We thank Dr. Ann Perera and Dr. Jonathan Wendel for suggestions.

References

3. Jackson JR. Toothache medicine: A customary use of pale purple coneflower (Echinacea pallida (Nutt.) Nutt.) among the Yuchi in eastern Oklahoma, USA. Econ Bot 2006; 60: 386–8
11. Blumenthal M, Urbatsch LE. Echinacea taxonomy — is the re-classification of the genus warranted? Herbal Gram 2006; 72: 30–1, 80

Fig. 4 Hierarchical dendrogram constructed from profiles generated from 43 lipophilic metabolites, using Canberra distance, illustrating the distances between 40 Echinacea accessions (3 plants per accession). Panel A is marked only with abbreviations of the accessions following McGregor [9] and the final three digits of the accession number. Panel B also includes blocking corresponding to the four species recognized by Binns et al. [13], with A = E. atrorubens, L = E. laevigata, Pa = E. pallida, and Pu = E. purpurea.