Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Biochemical Systematics and Ecology 32 (2004) 875–891
biochemical systematics and ecology
www.elsevier.com/locate/biochemsyseco

A chemotaxonomic analysis of terpenoid variation in *Cannabis*

Karl W. Hillig *

Department of Biology, Indiana University, Bloomington, IN 47405, USA

Received 19 September 2003; accepted 25 April 2004

Abstract

To determine whether the terpenoid composition of the essential oil of *Cannabis* is useful for chemotaxonomic discrimination, extracts of pistillate inflorescences of 162 greenhouse-grown plants of diverse origin were analyzed by gas chromatography. Peak area ratios of 48 compounds were subjected to multivariate analysis and the results interpreted with respect to geographic origin and taxonomic affiliation. A canonical analysis in which the plants were pre-assigned to *C. sativa* or *C. indica* based on previous genetic, morphological, and chemotaxonomic studies resulted in 91% correct assignment of the plants to their pre-assigned species. A scatterplot on the first two principal component axes shows that plants of accessions from Afghanistan assigned to the wide-leaflet drug biotype (an infraspecific taxon of unspecified rank) of *C. indica* group apart from the other putative taxa. The essential oil of these plants usually had relatively high ratios of guaiol, isomers of eudesmol, and other unidentified compounds. Plants assigned to the narrow-leaflet drug biotype of *C. indica* tended to have relatively high ratios of *trans*-β-farnesene. Cultivars of the two drug biotypes may exhibit distinctive medicinal properties due to significant differences in terpenoid composition.

Keywords: *Cannabis*; Chemotaxonomy; Essential oil; Eudesmol; Guaiol; Hemp; Taxonomy; Terpene

1. Introduction

The essential oil of *Cannabis* is a complex mixture composed primarily of monoterpenoids, sesquiterpenoids, and cannabinoids, as well as alkanes and numerous

* Corresponding address. 1010 Saratoga Road, Ballston Lake, NY 12019, USA.
E-mail address: khillig@bio.indiana.edu (K.W. Hillig).

0305-1978/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bse.2004.04.004

oxygen-containing compounds (Hendriks et al., 1978; Turner et al., 1980; Ross and ElSohly, 1996). The cannabinoids comprise a family of over 60 terpenophenolic compounds unique to *Cannabis* (Turner et al., 1980). Small (1979) regarded cannabinoids to be "extremely important" chemotaxonomic markers. In contrast, terpenoids are common in many plant families and their utility in differentiating *Cannabis* taxa is not well established. The primary purpose of this investigation was to characterize the terpenoid composition of the essential oil of *Cannabis* plants of diverse geographic origin to determine whether differences exist that are useful for taxonomic discrimination.

Some taxonomists treat *Cannabis* (Cannabaceae) as a single highly variable species, *C. sativa* L., while others consider the genus to be polytypic. Lamarck (1785) differentiated *C. indica* Lam. from *C. sativa* based on differences in morphology and geographic range, as well as its stronger aroma and greater inebriant ability. Other species of *Cannabis* have been proposed (reviewed in Schultes et al., 1974; Small and Cronquist, 1976); of these, only *C. ruderalis* Janisch. is widely accepted. Vavilov and Bukinich (1929) studied putative wild populations of *Cannabis* in central Asia and Afghanistan and recognized two species, *C. sativa* and *C. indica*. Wild populations of *C. indica* were assigned the name *C. indica* var. *kafiristanica* Vav. Vavilov (1926) considered *C. ruderalis* to be synonymous with *C. sativa* var. *spontanea* Vav.

Small and coworkers conducted a systematic investigation of *Cannabis* including morphological and chemotaxonomic studies (Small and Beckstead, 1973a,b; Small et al., 1976; Small, 1979). Small and Beckstead (1973b) determined that individual *Cannabis* plants can be assigned to one of three chemical phenotypes (chemotypes) having different ratios of Δ^9-tetrahydrocannabinol (THC, the primary psychoactive component of *Cannabis* resin) relative to cannabidiol (CBD, which is not considered psychoactive). Small and Beckstead (1973a) characterized *Cannabis* populations with respect to their mean THC/CBD ratios but this method blurs the qualitative distinction between the chemotypes of individual plants. The THC/CBD chemotype of a plant is determined by its genotype at a single locus (de Meijer et al., 2003; Mandolino et al., 2003). Plants with putative genotypes B_D/B_D, B_D/B_T, or B_T/B_T produce low (<0.20), intermediate (0.25–1.50), or high (>50) ratios of THC/CBD (Hillig and Mahlberg, 2004).

Small and Cronquist (1976) published a taxonomic treatment of *Cannabis* similar to that of Vavilov and Bukinich (1929), but they reduced *C. indica* to the rank of subspecies. *Cannabis sativa* subsp. *sativa* and *C. sativa* subsp. *indica* (Lam.) Small and Cronq. were each bifurcated into a cultivated and a wild variety. The two subspecies were delimited on the basis of their "intoxicant ability" determined by quantitative levels of THC in their leaf tissues. Schultes and Hofmann (1980) contended that an objective interpretation of Small and Becksteads' (1973a) cannabinoid data tends to support a polytypic concept of *Cannabis*.

Emboden (1974) advanced a polytypic treatment of *Cannabis* based on morphological differences. He objected to Small and Cronquists' (1976) treatment, arguing that the quantitative level of THC in a *Cannabis* plant is unstable and the

presumed ''psychotomimetic'' effect of a plant on a human being is not a legitimate character for taxonomic delimitation (Emboden, 1981).

Schultes et al. (1974) and Anderson (1980) recognized three species of *Cannabis* based on morphological criteria. Their similar circumscriptions of *C. indica* included short, densely branched plants from Afghanistan. Anderson assigned plants with narrow, lanceolate or linear-lanceolate leaflets to *C. sativa* and plants with wide, oblanceolate leaflets to *C. indica*. This treatment differed from Lamarck's (1785) original concept that described *C. indica* as having narrower leaflets than *C. sativa*. Wide-leaflet ''indica'' drug strains are traditionally cultivated for the production of hashish and narrow-leaflet ''sativa'' drug strains are usually cultivated for the production of marijuana (de Meijer, 1999). In recent years, hybrids between these two taxa have been widely cultivated for illicit drug production and are used for either purpose. McPartland et al. (2000) treated the wide-leaflet drug (WLD) ''biotype'' as a separate species from *C. sativa* and *C. indica*, but Hillig (2004) determined that the narrow-leaflet drug (NLD) and WLD biotypes were both derived from the *C. indica* gene pool.

Biochemical evidence of restricted gene flow between the *indica* and *sativa* taxa (regardless of rank) was provided by an analysis of flavonoid variation among 53 *Cannabis* plants from eight countries (Clark and Bohm, 1979). Luteolin-C-glycuronide was detected in 30 of 31 specimens assigned to *C. sativa* subsp. *sativa* but in only one of 22 specimens assigned to *C. sativa* subsp. *indica*.

Taxonomic inferences based on DNA studies conducted for forensic or plant breeding purposes are difficult to make because of limited sampling of putative taxa and the absence of voucher specimens (Faeti et al., 1996; Jagadish et al., 1996; Shirota et al., 1998; Siniscalco Gigliano, 2001; Gilmore et al., 2003). However, a random amplified polymorphic DNA (RAPD) analysis of 13 *Cannabis* accessions shows a Korean accession (here assigned to *C. indica*) divergent from the European hemp accessions (Faeti et al., 1996).

To reconcile unresolved taxonomic issues, a systematic investigation of a diverse *Cannabis* germplasm collection was conducted that included analyses of genetic, morphological, and biochemical variation within a set of 157 accessions of known geographic origin (Hillig, 2004; Hillig, in press; Hillig and Mahlberg, 2004). Hillig (2004) surveyed sample populations of each accession for allozyme variation at 17 gene loci. Principal components (PC) analysis of the allozyme allele frequencies revealed a bifurcation of the *Cannabis* collection into two major gene pools that correspond (more or less) to previous circumscriptions of *C. sativa* and *C. indica*. A third putative gene pool consisted of ruderal accessions from central Asia that correspond to *C. ruderalis*. The *sativa* gene pool includes hemp (cultivated for fiber or seed) and feral accessions from Europe, Asia Minor, and central Asia. The *indica* gene pool includes hemp accessions from southern and eastern Asia, drug accessions from the Indian subcontinent, Africa, and other drug producing regions, and wild or feral accessions from India and Nepal. Hillig rejected a single-species hypothesis and recognized *C. sativa* and *C. indica* as separate species. *Cannabis ruderalis* was given tentative recognition for further hypothesis testing.

Hillig (in press) studied phenotypic variation in the same set of accessions that was utilized in his genetic study. The results were interpreted with respect to an a priori assignment of accessions to species and to putative infraspecific taxa of unassigned rank (“biotypes”) based on genetic variation, geographic origin, and presumed purpose of cultivation (Hillig, 2004). Canonical analyses resulted in 100% correct assignment of the accessions to their pre-assigned species and infraspecific taxa. Because few morphological differences were found between plants assigned to *C. ruderalis* and to the feral biotype of *C. sativa*, a two-species concept was favored.

Hillig and Mahlberg (2004) determined qualitative and quantitative cannabinoid levels in the pistillate inflorescences of 253 individual plants and the frequency of the B_T allele in sample populations of the same set of accessions that Hillig (in press) analyzed for genetic and morphological variation. The data were analyzed with respect to an a priori assignment of accessions to putative taxa. The mean frequency of B_T was significantly higher in the hemp, drug, and feral biotypes of *C. indica* than in the hemp and feral biotypes of *C. sativa* and putative *C. ruderalis*. Plants with enhanced levels of cannabinoids having propyl side chains were common only in *C. indica*. Hillig and Mahlberg interpreted these results to support recognition of *C. sativa* and *C. indica* as separate species. Accessions assigned to *C. ruderalis* and the feral biotype of *C. sativa* did not differ significantly in cannabinoid content so recognition of *C. ruderalis* as a separate species was not supported.

Terpenoid variation may provide another means of taxonomic discrimination in *Cannabis*. Forensic studies of the feasibility of using terpenoid profiles to determine the geographic origin of confiscated *Cannabis* material gave mixed results (Hood et al., 1973; Hood and Barry, 1978; Brenneisen and ElSohly, 1988). Hood and Barry (1978) reported a low correspondence between headspace volatiles of marijuana and its geographic origin. However, Brenneisen and Elsohly (1988) analyzed a small number of marijuana samples and determined that enhanced levels of particular terpenes may be useful for determining the country of origin. A study of terpenoid variation primarily among European hemp cultivars revealed differences in terpenoid composition and aroma but these differences were not interpreted with respect to geographic origin or taxonomic affiliation (Mediavilla and Steinemann, 1997). Levels of terpenoid components in the essential oil of two Austrian populations of “*C. sativa* subsp. *spontanea*” (=*C. ruderalis*) were within the ranges reported by Mediavilla and Steinemann for European hemp cultivars (Novak and Franz, 2003). None of these studies of terpenoid variation included a broad sampling of the *Cannabis* gene pool.

Monoterpenoids and other volatile compounds are primarily responsible for differences in fragrance among *Cannabis* strains (Ross and ElSohly, 1996; Mediavilla and Steinemann, 1997; Clarke, 1998). Sesquiterpenoids are less volatile than monoterpenoids and a pronounced odor is generally only observed in derivatives with simple substitutions such as alcohols or ketones (Herout, 1970). Ross and ElSohly (1996) determined that drying fresh female inflorescences results in significantly greater loss of monoterpenoids than sesquiterpenoids but none of the major components of the oil completely disappear. Mediavilla and Steinemann (1997) repor-

ted that terpenoid yield and floral aroma vary with regard to the degree of maturity of pistillate inflorescences and whether or not the inflorescences are pollinated. They were unable to determine a relationship between terpenoid composition and "scent quality" of the essential oil. Some terpenoids present in the essential oil of *Cannabis* are pharmacologically active and may synergize the effects of the cannabinoids (McPartland and Russo, 2001; D. Watson, pers. comm., 2004).

This study characterized variation in the terpenoid composition of the essential oil among *Cannabis* plants grown from the same set of accessions that were characterized in previous taxonomic studies, to determine whether differences exist that are useful for taxonomic discrimination (Hillig, 2004; Hillig, in press; Hillig and Mahlberg, 2004). Variation in levels of pharmacologically active terpenoids and the utility of using terpenoid profiles in forensic investigations were also of interest.

2. Materials and methods

2.1. Plant material

A diverse germplasm collection consisting of 157 accessions of known geographic origin was obtained from *Cannabis* breeders, researchers, germplasm repositories, and law enforcement agencies. Each accession consisted of a small packet of viable achenes ("seeds"). Passport data, accession codes, and the assignment of accessions to putative taxa were previously published (Hillig, 2004). Voucher specimens are deposited in the herbarium at Indiana University (IND).

Plants were grown individually in 13 cm clay pots in soil consisting of three parts black peat, two parts vermiculite, and one part each of sand and topsoil. Each plant received 100 ml of nutrient solution at weekly intervals consisting of 25 ml of Dyna-Gro 7-9-5 (Dyna-Gro Corporation, San Pablo, California, USA) and 18 ml of 1 M KOH (to neutralize acidity) per 3.8 l of water. Supplemental lighting during the vegetative stage of growth was supplied by three 1000-watt metal-halide lamps in each greenhouse to extend the daylight length to 16 h.

Staminate and pistillate plants were segregated into separate greenhouses when their sex was determined. Inflorescences of unpollinated pistillate plants were individually sampled when resin production of each plant was visually assessed to have reached its peak. Samples were air dried at room temperature. A period of up to 2 months preceded analysis, during which time the samples were stored in open containers at room temperature. The samples were oven-dried overnight at 30 °C prior to extraction. Primary and larger secondary leaves were removed and only visibly resinous floral bracts and subtending leaves were extracted. Fragrance was evaluated by rubbing the fresh stems and smelling the fingers, and by smelling the fresh and dried inflorescences.

2.2. Samples evaluated

Floral samples of 162 plants representing 82 accessions were extracted and analyzed by gas chromatography (GC). One to five plants of each accession were

evaluated. Country codes and number of accessions analyzed were: Af = Afghanistan (9); Bg = Bulgaria (2); Ch = China (3); Cl = Colombia (1); Gm = Gambia (1); Gr = Germany (1); Hn = Hungary (7); In = India (4); It = Italy (3); Jm = Jamaica (1); Jp = Japan (1); Ls = Lesotho (1); Mx = Mexico (4); Np = Nepal (2); Ng = Nigeria (1); Rm = Romania (2); Rs = Russia (10); SA = South Africa (3); SK = South Korea (7); Sp = Spain (2); Sw = Swaziland (1); Th = Thailand (6); Tk = Turkey (5); Ug = Uganda (1); Uk = Ukraine (3); Yg = former Yugoslavia (1).

2.3. Sample extraction

Sample material (50 mg) was placed in a test tube with 1 ml of chloroform. The sample was crushed with a glass rod and briefly sonicated to dislodge and/or rupture the resin heads of the glandular trichomes. The sample remained in the solvent at room temperature for at least 1 h and was sonicated again for a few seconds. A 150 μl volume of extract was transferred to a clean test tube and the solvent was evaporated with a gentle stream of nitrogen. The residue was redissolved in 50 μl of acetone containing 0.25 mg/ml of *n*-eicosane, the internal standard (i.s.).

2.4. GC conditions

Chromatograms were generated with a Hewlett-Packard (HP) 5710A gas chromatograph fitted with a 30 m × 0.53 mm DB-5ms column (J & W Scientific, Rancho Cordova, California, USA), with a film thickness of 1.5 μm. Injector and detector port temperatures were 250 and 300 °C, respectively. Carrier and make-up gas flow rates were 5.8 ml/min He and 24.2 ml/min N_2. The signal was detected with a flame-ionization detector (FID). The oven temperature was 90 °C for 8 min, increased at 4 °C/min to 300 °C, and held for 8 min. A 0.5 μl volume of extract was injected into the chromatograph using splitless injection.

The identities of certain peaks were determined by comparing relative retention times (RRT's) on the DB-5ms column with those of terpenoid standards of compounds reported to be present in the essential oil of *Cannabis*. Purified terpenoids were obtained from Sigma Chemical Company (St. Louis, Missouri, USA) and Fluka Chemical Corp. (Ronkonkoma, New York, USA). These included α-pinene, β-pinene, β-myrcene, limonene, γ-terpinene, terpinolene, linalool, α-terpineol, β-caryophyllene, *trans*-β-farnesene, α-humulene, caryophyllene oxide, and guaiol. GC/mass spectrometry (GC/MS) of four plant extracts was performed in a separate laboratory to help identify the terpenoid peaks. These analyses utilized an HP 5890 Series II GC/5970B MSD fitted with a 60m × 0.25 mm Supelco (Sigma-Aldrich, St. Louis, Missouri, USA) SPB-5 column with a film thickness of 0.25 μm. The GC signal was detected with a FID and the mode of injection was splitless. The GC/MS data were compared with the other chromatograms and with mass spectra and retention indices published by Adams (1989) and elsewhere.

Table 1
Working hypothesis of a taxonomic circumscription of the *Cannabis* germplasm collection. Three species and seven putative taxa are recognized

Putative taxon	Description
C. indica hemp biotype	Hemp landraces of southern and eastern Asia
C. indica feral biotype	Wild or feral populations of India and Nepal
C. indica NLD biotype	Narrow-leaflet drug (NLD) strains of the Indian subcontinent, Africa, and other drug producing regions
C. indica WLD biotype	Wide-leaflet drug (WLD) strains of Afghanistan and Pakistan
C. sativa hemp biotype	Hemp landraces of Europe, Asia Minor, and central Asia
C. sativa feral biotype	Naturalized populations of eastern Europe
C. ruderalis	Ruderal populations of central Asia

2.5. Statistical analysis

Forty-eight peaks with retention times less than that of *n*-eicosane were selected for analysis. Individual peak areas relative to the total area under all 48 peaks were tabulated. This method was used because the relative proportions of the various compounds were presumed to be more stable than their absolute levels (relative to the i.s.), which vary with the maturity of the sample material and the length of time between harvest and analysis (Ross and ElSohly, 1996; Meier and Mediavilla, 1998). Statistical analysis was performed with JMP version 5.0 (SAS Institute, 2002). The data matrix was subjected to PC analysis and canonical analysis (i.e., equally weighted multiple discriminant analysis). The PC axes were extracted from the correlation matrix. The data were analyzed and interpreted with respect to the taxonomic hypothesis outlined in Table 1.

3. Results

Fig. 1 shows a gas chromatogram of the essential oil of a plant of Afghani origin (accession Af-3). Peaks 36 (guaiol), 38 (γ-eudesmol), and 40 (β-eudesmol) are bicyclic sesquiterpene alcohols (Fig. 2) that were often prominent on chromatograms of plants of Afghani origin. Although peak 9 (terpinolene) is prominent in Fig. 1, this was not a general feature of plants assigned to the WLD biotype. Peak 19 (β-caryophyllene) was prominent on most chromatograms.

Table 2 shows RRT's, identities, and mass spectral data for the 48 peaks that were statistically analyzed. Not all of the numbered peaks were detected in all samples. Numbered peaks with RRT's less than that of β-caryophyllene (peak 19) had molecular weights ≤ 154 and are known or presumed to be monoterpenoids. Numbered peaks with RRT's greater than that of β-caryophyllene had molecular weights ≥ 204 and are known or presumed to be sesquiterpenoids.

Fig. 3 shows a scatterplot of plants on PC axis 1 and PC axis 2, which accounted for 15.4% and 11.6% of the total variance, respectively. The plants form a dense cluster toward the left side of the plot with outliers toward the top and the lower right quadrant. Different plants of a given accession tend to occupy the same

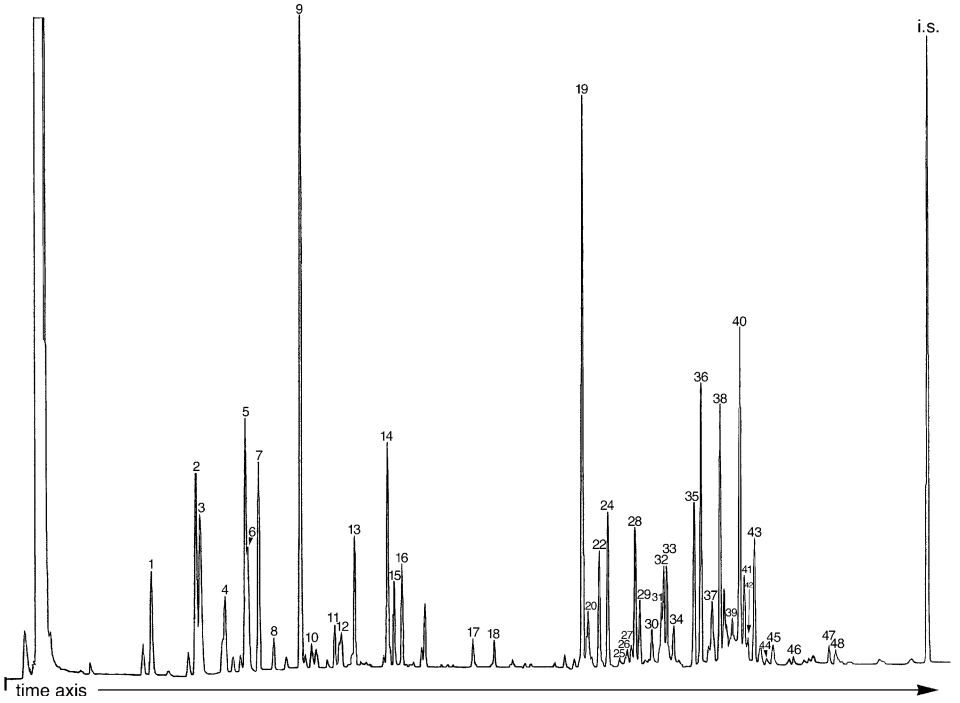

Fig. 1. Gas chromatogram of the essential oil of a plant of accession from Afghanistan assigned to the wide-leaflet drug (WLD) biotype of *C. indica*. i.s. = internal standard. Column = DB-5ms.

region of the plot. The dense cluster is comprised of plants of accessions from several countries, assigned to all putative taxa except the WLD biotype of *C. indica*. Plants of accessions of Afghani origin were assigned to the WLD biotype and are outliers, mostly located in the lower right quadrant of the plot. Outliers toward the top of the plot include several plants of accessions assigned to the NLD biotype of *C. indica*. Plants of accessions assigned to other putative taxa are also located in this region. Table 3 lists the peaks with eigenvector loadings having the largest absolute values for PC axis 1 and PC axis 2. The six peaks with the largest loadings on PC axis 1 are sesquiterpenoids and the five peaks with the largest loadings

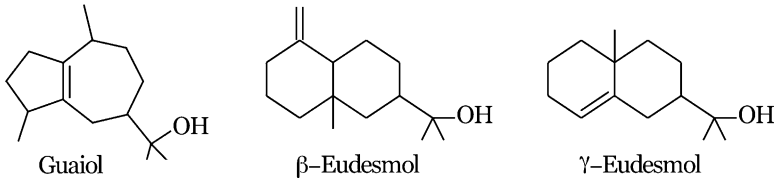

Fig. 2. Molecular structure of three sesquiterpene alcohols typically found at relatively high ratios in the essential oil of *C. indica* plants of Afghani origin.

Table 2
Relative retention times (RRT's), known identities, and mass spectral data for 48 terpenoid compounds in the essential oil of *Cannabis*

Peak	RRT[a]	Identity, M^+ (m/z)
1	0.159	α-Pinene
2	0.206	β-Pinene
3	0.211	β-Myrcene
4	0.238	Δ-3-Carene
5	0.260	Limonene
6	0.263	
7	0.274	*Cis*-ocimene
8	0.290	γ-Terpinene
9	0.320	Terpinolene
10	0.332	Linalool
11	0.357	β-Fenchol
12	0.364	M^+ 136 (m/z: 43 71 81 99 93 55)
13	0.379	
14	0.415	
15	0.422	
16	0.430	α-Terpineol
17	0.507	
18	0.530	
19	0.626	β-Caryophyllene
20	0.632	(Z)-α-*trans*-bergamotene
21	0.635	α-Guaiene
22	0.644	*Trans*-β-farnesene
23	0.647	M^+ 204 (m/z: 41 69 93 105 133)
24	0.653	α-Humulene
25	0.666	M^+ 204 (m/z: 41 189 91 105 119)
26	0.674	M^+ 204 (m/z: 189 41 133 91 105)
27	0.679	M^+ 204 (m/z: 41 79 93 105 67 55)
28	0.683	M^+ 204 (m/z: 189 41 93 204 107)
29	0.688	M^+ 204 (m/z: 93 107 41 119 79)
30	0.701	M^+ 204 (m/z: 41 69 93 161 79 55)
31	0.711	M^+ 204 (m/z: 93 41 79 119 67 105)
32	0.713	α-Gurjunene (?)
33	0.717	M^+ 204 (m/z: 161 204 122 107 91)
34	0.724	M^+ 204 (m/z: 41 69 93 55 107 79)
35	0.747	Caryophyllene oxide
36	0.754	Guaiol
37	0.766	
38	0.775	γ-Eudesmol
39	0.788	M^+ 236 (m/z: 59 43 105 161 93 81)
40	0.797	β-Eudesmol
41	0.802	M^+ 222 (m/z: 59 107 43 135 161)
42	0.805	
43	0.813	
44	0.825	
45	0.833	
46	0.855	
47	0.894	
48	0.901	
i.s.	1.000	*n*-Eicosane

[a] On DB-5ms column.

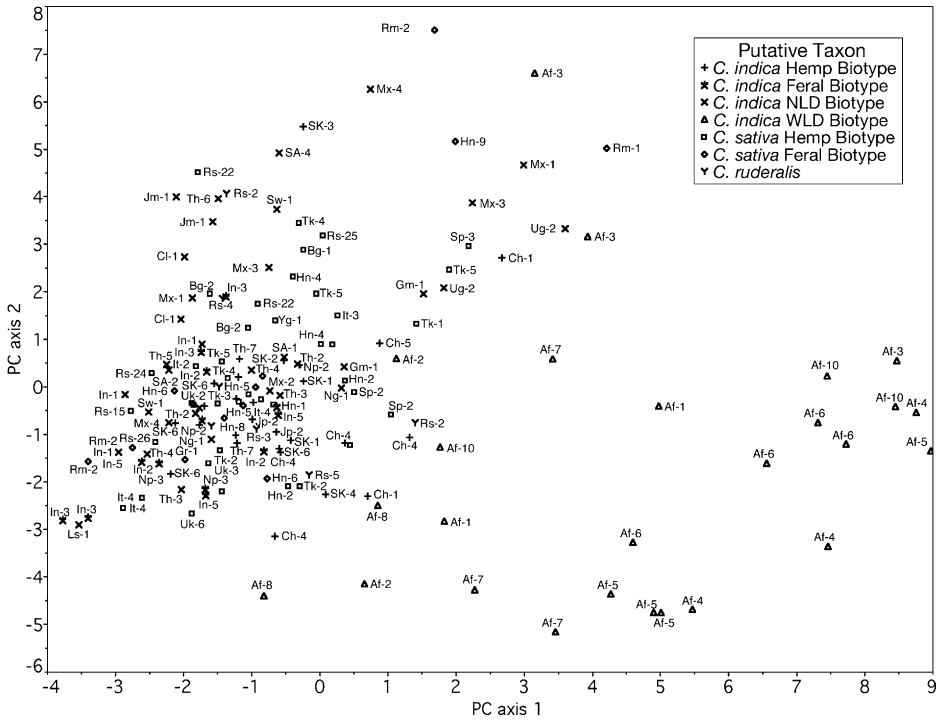

Fig. 3. Scatterplot of 162 *Cannabis* plants on the 1st and 2nd principal component (PC) axes. The plants were pre-assigned to seven putative taxa. Country codes are given in the text. Accession codes were previously published (Hillig, 2004). NLD, narrow-leaflet drug; WLD, wide-leaflet drug.

on PC axis 2 are monoterpenoids. A positive eigenvector loading indicates a trend toward increasing values of the corresponding trait when proceeding in the positive

Table 3
Numbered peaks with largest eigenvector loadings (absolute values) for PC axis 1 and PC axis 2

PC axis 1		PC axis 2	
Peak	Eigenvector loading	Peak	Eigenvector loading
40	0.33	14	0.32
36	0.30	13	0.32
38	0.29	8	0.27
47	0.28	15	0.27
48	0.27	6	0.26
24	−0.21	32	−0.23
12	0.21	33	−0.22
39	0.21	18	0.22
11	0.21	31	−0.19
19	−0.20	19	−0.18

direction along a given PC axis. A negative loading indicates the same trend in the opposite direction.

Canonical analysis resulted in 56% correct assignment of the plants to their preassigned species. When the seven accessions preassigned to *C. ruderalis* were reassigned to *C. sativa*, canonical analysis resulted in 91% correct assignment of the plants to *C. sativa* or *C. indica*. A canonical analysis in which the accessions were preassigned to the seven putative taxa in Table 1 resulted in 31% correct assignment of the plants to their preassigned taxa. A scatterplot of plants on the first two canonical axes (Fig. 4) shows that the plants preassigned to the NLD and WLD biotypes of *C. indica* were well discriminated from the other putative taxa. With the ruderal accessions from central Asia reassigned to the feral biotype of *C. sativa*, canonical analysis resulted in 88% correct assignment of the plants to the six biotypes of *C. sativa* and *C. indica*.

Means and standard deviations of the ratios of the peak areas relative to the area under all 48 peaks for the seven putative taxa are given in Appendix A. Student's *t*-tests show which means are significantly different ($P \leq 0.05$). The mean ratios for peaks 11, 12, 16, 32, 33, 36, 38, 39, 40, 41, 47, and 48 were significantly greater for the WLD accessions than for all other taxa and differed by factors of 2–8. Relatively high ratios of peaks 36, 38, and 40 also occurred in a few plants of accessions from China, South Korea, Nepal, South Africa, and Uganda assigned to *C. indica* and an accession from Turkey assigned to *C. sativa*. The mean peak area ratios for peaks 19 and 24 were significantly lower for the WLD biotype than

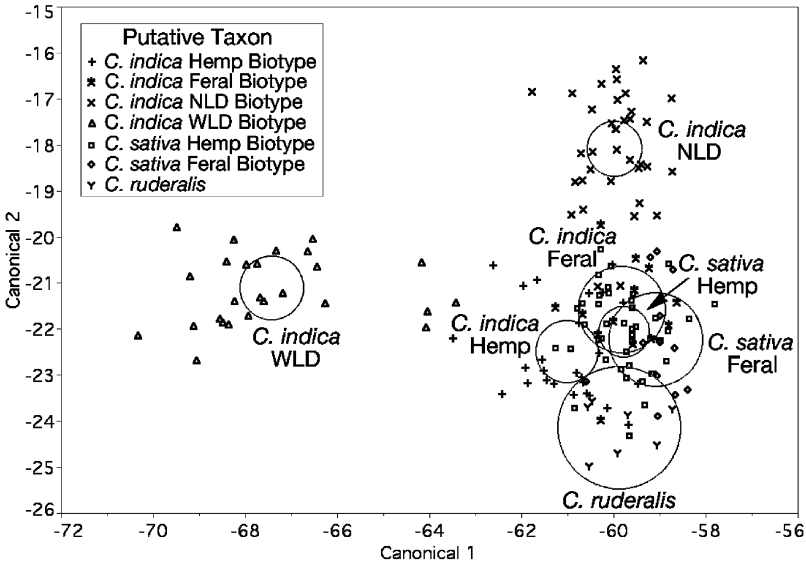

Fig. 4. Scatterplot of 162 *Cannabis* plants on the 1st and 2nd canonical axes. Circles show 95% confidence limits for the multivariate means of the putative taxa. NLD, narrow-leaflet drug; WLD, wide-leaflet drug.

for all other putative taxa except *C. ruderalis*. The mean ratios for peaks 29 and 30 were significantly higher for the feral biotype of *C. indica* than for all other taxa, and the mean ratios for peaks 22 and 31 were significantly higher for the NLD and hemp biotypes of *C. indica*, respectively, than for all other taxa. The mean peak area ratios (not shown) were also computed for *C. sativa* (including ruderal central Asian accessions) and *C. indica*. Significant differences ($P \leq 0.05$) between species were found for peaks 3, 8, 12, 16, 20, 22, 23, 26, 28, 29, 30, 32, 33, 34, 36, 37, 38, 40, 41, 43, 45, and 47. Only peaks 3, 12, 23, 33, 36, 38, 40, 41, and 47 differed by more than a factor of two between the two species.

Individual histograms for each numbered peak (not shown) show that relatively high ratios of peak 4 were characteristic of plants of hemp accessions from South Korea. These plants were notable for the candy-like lemon/pine aroma of their stems. High ratios of peak 22 were characteristic of plants of NLD accessions from Thailand. The stems of these plants had a pleasant green-tea-like fragrance.

4. Discussion

Numerous significant differences in mean terpenoid ratios were found among the *Cannabis* taxa. However, these differences were of limited use for chemotaxonomic discrimination, with sesquiterpenoids generally more useful than monoterpenoids. Most plants of Afghani origin had components in their essential oil with peak area ratios that distinguished them from the other putative taxa, but terpenoid composition varied between plants and was not always a reliable indicator of an Afghani plant's affiliation to the WLD biotype. The hemp, NLD, and feral biotypes of *C. indica* each had one or two components in their essential oil with ratios significantly higher than those of the other putative taxa, but the means differed by less than a factor of two and do not provide a firm basis for chemotaxonomic discrimination. Several ratios were significantly different between *C. sativa* (including ruderal accessions from central Asia) and *C. indica* but these differences were mostly for peaks that distinguished the WLD biotype of *C. indica* from the other putative taxa. Only two other peaks (3 and 23) had mean ratios that differed by more than a factor of two between the two species. These differences do not appear to be of practical significance for chemotaxonomic discrimination at the species level.

Human selection may have contributed to differences in terpenoid composition of the essential oil among drug strains as a result of recurrent selection and inbreeding. Terpenoids contribute to the distinctive smoking qualities and possibly to the character of the "high" associated with marijuana and hashish produced from different drug strains (Clarke, 1998; D. Watson, pers. comm., 2004). Because the essential oil of *Cannabis* is a complex mixture and individual sesquiterpenoids (including guaiol, which has a faint phenolic aroma) are difficult to detect except by laboratory analysis, it seems likely that differences in terpenoid composition that characterize the WLD biotype are naturally occurring in its wild progenitor and are not primarily a result of human selection.

The terpenoid composition of the essential oil may be useful in forensic investigations for inferring the geographic origin of confiscated marijuana and hashish. However, a large number of samples of known provenance must be statistically analyzed to ascertain whether the source of a sample of unknown origin can be inferred with confidence. In the present study, the extracts of several plants of accessions from Thailand were characterized by elevated ratios of peak 22 (*trans*-β-farnesene), but Brenneisen and ElSohly (1988) did not identify this peak as indicative of confiscated marijuana of Thai origin. A chromatogram of a hashish sample published by Brenneisen and ElSohly appears to show elevated peaks corresponding to guaiol and isomers of eudesmol, which suggests that the sample material may have been from Afghanistan. However, the essential oil of *Cannabis* plants from other hashish producing regions should be analyzed to determine whether elevated ratios of these compounds are unique to plants of Afghani origin. Ross and ElSohly (1996) reported levels of guaiol and isomers of eudesmol ranging from about 0.5% to 1.8% in the volatile oil of a "high potency hybrid". It is reasonable to assume that the recent worldwide dissemination of high yielding hybrid drug strains of diverse origin is obscuring regional differences, making forensic determination of the country of origin of confiscated marijuana based on physical characteristics less certain.

The pharmacology of the essential oils of NLD and WLD cultivars warrants further study due to differences in terpenoid composition. The terpenoid fraction of the essential oil may modify or enhance the physiological effects of the cannabinoids resulting in greater medicinal benefit than the pure cannabinoid compounds (McPartland and Russo, 2001). Sesquiterpene alcohols that distinguish the WLD biotype occur in the essential oil of other medicinal plants including *Michelia champaca* L. and *Bulnesia sarmienti* Lorentz ex. Griseb. (=G*uajacum officinale* L.) (Merck Index, 1989).

Further studies are needed to positively identify the terpenoid compounds that differentiated the accessions from Afghanistan. The geographic range of wild and cultivated plants with elevated ratios of these compounds should be determined, and quantitative levels of these compounds should be ascertained to use as guidelines for chemotaxonomic discrimination of the WLD biotype of *C. indica*.

Acknowledgements

I thank Prof. Paul G. Mahlberg for facilitating this study. I also thank all those who contributed seeds for this investigation. This study was supported by a grant from HortaPharm B.V., The Netherlands. Drug Enforcement Administration number PI0043113 (PGM).

Appendix A. Arithmetic means (±SD) of the ratios of the peak areas (multiplied by 100) relative to the total area under all 48 peaks for 162 *Cannabis* plants assigned to seven putative taxa

Peak	*C. indica* Hemp Biotype $N = 27$[a]	*C. indica* Feral Biotype $N = 14$	*C. indica* NLD[b] Biotype $N = 35$	*C. indica* WLD[c] Biotype $N = 26$	*C. sativa* Hemp Biotype $N = 41$	*C. sativa* Feral Biotype $N = 12$	*C. ruderalis* $N = 7$
1	13.6a ± 7.8	6.3c ± 3.3	6.5ab ± 5.6	8.6bc ± 6.6	10.7 ab ± 5.7	10.2abc ± 6.6	12.0ab ± 6.8
2	3.8a ± 3.4	1.1b ± 1.6	2.4ab ± 3.3	2.0b ± 3.4	2.5ab ± 2.7	2.0ab ± 2.0	1.9ab ± 2.5
3	7.6d ± 9.4	14.9bc ± 10.9	5.8d ± 5.9	9.0cd ± 6.5	19.6b ± 13.2	19.6ab ± 16.2	28.3a ± 15.1
4	2.3a ± 2.6	0.2b ± 0.3	0.4b ± 0.5	0.1b ± 0.4	0.9b ± 3.1	0.1b ± 0.2	2.7a ± 4.0
5	1.8bc ± 1.9	3.0ab ± 3.1	1.3bc ± 1.2	4.0a ± 4.3	1.5bc ± 1.7	0.9c ± 1.2	3.1abc ± 6.7
6	0.3c ± 0.4	0.4abc ± 0.5	0.7a ± 0.8	0.4bc ± 0.7	0.6ab ± 0.6	0.8ab ± 0.7	0.6abc ± 0.7
7	4.3ab ± 4.9	5.5ab ± 4.4	3.0bc ± 2.3	1.9c ± 3.0	4.6ab ± 4.0	3.6abc ± 4.6	6.7a ± 7.0
8	0.1bc ± 0.2	0.1bc ± 0.2	0.2ab ± 0.2	0.1c ± 0.2	0.2ab ± 0.2	0.3a ± 0.5	0.2abc ± 0.2
9	1.1b ± 2.2	3.7ab ± 5.4	4.4a ± 8.0	1.0b ± 2.9	2.9ab ± 3.4	1.2b ± 2.4	1.2ab ± 1.9
10	0.7a ± 1.0	1.3a ± 0.9	1.0a ± 1.0	1.2a ± 1.1	1.0a ± 1.2	0.8a ± 0.9	0.5a ± 0.3
11	0.3b ± 0.4	0.2b ± 0.2	0.2b ± 0.2	0.8a ± 0.9	0.2b ± 0.3	0.2b ± 0.2	0.3b ± 0.6
12	0.2b ± 0.3	0.1b ± 0.1	0.2b ± 0.2	0.6a ± 0.6	0.1b ± 0.2	0.1b ± 0.2	0.1b ± 0.4
13	0.3ab ± 0.6	0.1b ± 0.2	0.6a ± 0.8	0.3ab ± 0.9	0.5ab ± 0.6	0.6ab ± 1.1	0.5ab ± 0.9
14	1.1ab ± 1.9	0.3b ± 0.6	1.9a ± 1.9	0.6b ± 1.5	1.1ab ± 1.3	1.8a ± 3.2	1.0ab ± 1.4
15	0.4bc ± 0.5	0.2c ± 0.3	1.0a ± 1.3	0.5bc ± 0.9	0.5bc ± 0.6	0.9ab ± 1.5	0.5abc ± 0.7
16	0.3b ± 0.3	0.4b ± 0.4	0.5b ± 0.3	1.0a ± 0.9	0.4b ± 0.3	0.3b ± 0.4	0.3b ± 0.4
17	0.2b ± 0.4	0.2ab ± 0.5	0.6ab ± 0.9	0.9a ± 2.4	0.3b ± 0.5	0.6ab ± 0.9	0.3ab ± 0.5
18	0.2ab ± 0.5	0.1ab ± 0.1	0.2a ± 0.3	0.1b ± 0.2	0.1ab ± 0.1	0.2ab ± 0.3	0.1ab ± 0.2
19	18.7ab ± 7.7	21.9a ± 7.3	15.7bc ± 7.2	9.7d ± 6.2	17.6ab ± 10.1	16.7abc ± 10.9	9.9cd ± 5.6
20	1.4c ± 1.1	2.8abc ± 1.2	4.0a ± 2.6	3.7ab ± 5.5	1.5c ± 1.5	2.0bc ± 1.4	1.4bc ± 0.8
21	0.5b ± 1.1	0.5ab ± 0.9	1.0a ± 1.3	0.4b ± 0.7	0.5b ± 1.0	< 0.1b ± 0.1	< 0.1b ± 0.1
22	2.0c ± 2.0	4.3b ± 2.5	7.6a ± 4.4	4.1b ± 3.3	3.1bc ± 2.7	4.0bc ± 2.9	2.7bc ± 1.5
23	0.7a ± 1.3	0.4ab ± 0.5	0.7a ± 0.8	0.7a ± 1.3	0.2b ± 0.5	0.2ab ± 0.3	0.1ab ± 0.1
24	5.6a ± 2.2	6.8a ± 2.4	5.8a ± 2.9	2.9c ± 2.0	5.5a ± 3.0	5.2ab ± 2.9	3.0bc ± 1.7
25	0.4b ± 0.4	1.3a ± 1.2	0.7b ± 0.4	0.6b ± 0.5	0.4b ± 0.5	0.7b ± 0.5	1.4a ± 1.4

26	1.2a ± 0.6	0.6bc ± 0.5	0.7bc ± 0.4	0.5c ± 0.4	0.9ab ± 0.6	1.2a ± 0.6	1.3a ± 0.7
27	3.3bc ± 1.4	3.0bc ± 2.4	5.6a ± 3.6	2.3c ± 1.7	2.8c ± 2.0	4.5ab ± 2.9	3.5bc ± 2.4
28	2.9b ± 1.8	2.8bc ± 2.4	4.4a ± 2.0	1.8c ± 1.1	2.2bc ± 1.7	3.4ab ± 2.5	2.1bc ± 1.3
29	1.1b ± 0.7	2.4a ± 1.7	1.6b ± 1.6	1.2b ± 0.9	1.1b ± 1.1	0.9b ± 0.4	0.9b ± 0.6
30	0.7c ± 0.4	2.4a ± 2.7	1.4b ± 0.9	0.8c ± 0.5	0.7c ± 0.6	0.9bc ± 0.6	0.5bc ± 0.4
31	3.0a ± 1.6	2.0b ± 1.5	0.6c ± 1.2	1.9 b ± 1.5	1.3b ± 1.5	1.3bc ± 2.3	1.1bc ± 0.9
32	0.8b ± 1.1	0.2bc ± 0.2	0.2c ± 0.2	2.4a ± 2.1	0.4bc ± 0.7	0.4bc ± 0.3	0.7bc ± 0.8
33	1.1b ± 1.2	0.2c ± 0.2	0.4c ± 0.4	2.9a ± 2.5	0.5c ± 0.8	0.4bc ± 0.3	0.9bc ± 0.9
34	0.9bc ± 0.5	2.1a ± 1.0	1.7a ± 1.1	0.9bc ± 0.7	1.2b ± 0.8	0.6c ± 0.5	0.7bc ± 0.4
35	7.0ab ± 4.8	2.5c ± 3.1	8.9a ± 7.9	4.2bc ± 4.2	5.4bc ± 6.1	5.2abc ± 5.4	3.6bc ± 3.7
36	0.5b ± 0.5	0.2b ± 0.5	0.2b ± 0.4	3.5a ± 1.8	0.3b ± 0.3	0.1b ± 0.2	0.1b ± 0.2
37	2.7ab ± 2.0	1.5b ± 1.4	3.2a ± 2.7	2.2ab ± 1.9	1.8b ± 1.8	1.9ab ± 1.8	1.4b ± 1.1
38	1.5b ± 1.2	0.6c ± 0.3	0.6c ± 0.6	4.8a ± 2.1	0.9c ± 0.7	0.8bc ± 0.7	0.7bc ± 0.7
39	0.1c ± 0.2	0.1bc ± 0.2	< 0.1c ± 0.1	0.9a ± 0.7	0.3b ± 0.6	0.3bc ± 0.5	0.2bc ± 0.3
40	1.0b ± 0.8	0.5b ± 0.4	0.8b ± 0.6	7.4a ± 4.0	0.7b ± 0.4	0.5b ± 0.5	0.8b ± 0.4
41	0.3b ± 0.5	0.1b ± 0.2	0.1b ± 0.3	1.4a ± 1.4	0.1b ± 0.4	0.1b ± 0.3	0.2b ± 0.2
42	0.9ab ± 0.8	0.4b ± 0.3	1.1a ± 0.8	1.1a ± 1.0	0.9ab ± 0.7	1.4a ± 1.2	0.7ab ± 0.4
43	2.2a ± 1.6	2.1a ± 1.4	0.8b ± 1.1	2.1a ± 2.0	0.9b ± 0.9	1.8ab ± 2.8	1.0ab ± 1.2
44	0.3a ± 0.2	0.1bc ± 0.2	0.1c ± 0.2	0.3a ± 0.3	0.2ab ± 0.2	0.3ab ± 0.4	0.2abc ± 0.2
45	0.2bc ± 0.5	< 0.1c ± < 0.1	0.5ab ± 1.0	0.7a ± 0.5	0.2c ± 0.5	0.2bc ± 0.6	0.1bc ± 0.1
46	0.2bc ± 0.4	0.1c ± 0.2	0.5a ± 0.8	0.5ab ± 0.4	0.2bc ± 0.5	0.4abc ± 0.6	0.1abc ± 0.1
47	< 0.1b ± 0.1	< 0.1b ± < 0.1	< 0.1b ± < 0.1	0.8a ± 0.8	< 0.1b ± < 0.1	0.1b ± 0.3	< 0.1b ± 0.1
48	0.2bc ± 0.2	0.1c ± 0.1	0.3b ± 0.2	0.7a ± 0.5	0.2bc ± 0.2	0.3b ± 0.3	0.2bc ± 0.1

[a] *N*, number of plants.
[b] NLD, narrow-leaflet drug.
[c] WLD, wide-leaflet drug.

References

Adams, R.P., 1989. Identification of Essential Oils by Ion Trap Mass Spectroscopy. Academic Press, New York.

Anderson, L.C., 1980. Leaf variation among *Cannabis* species from a controlled garden. Bot. Mus. Leafl. Harvard Univ. 28, 61–69.

Brenneisen, R., ElSohly, M.A., 1988. Chromatographic and spectroscopic profiles of *Cannabis* of different origins: part I. J. Forensic Sci. 33, 1385–1404.

Clark, M.N., Bohm, B.A., 1979. Flavonoid variation in *Cannabis* L. J. Linn. Soc. 79, 249–257.

Clarke, R.C., 1998. Hashish! Red Eye Press, Los Angeles, CA.

de Meijer, E.P.M., 1999. Cannabis germplasm resources. In: Ranalli, P. (Ed.), Advances in Hemp Research. Haworth Press, Binghamton, New York, pp. 133–151.

de Meijer, E.P.M., Bagatta, M., Carboni, A., Crucitti, P., Cristiana Moliterni, V.M., Ranalli, P., Mandolino, G., 2003. The inheritance of chemical phenotype in *Cannabis sativa* L. Genetics 163, 335–346.

Emboden, W.A., 1974. *Cannabis*—a polytypic genus. Econ. Bot. 28, 304–310.

Emboden, W.A., 1981. The genus *Cannabis* and the correct use of taxonomic categories. J. Psychoactive Drugs 13, 15–21.

Faeti, V., Mandolino, G., Ranalli, P., 1996. Genetic diversity of *Cannabis sativa* germplasm based on RAPD markers. Plant Breed. 115, 367–370.

Gilmore, S., Peakall, R., Robertson, J., 2003. Short tandem repeat (STR) DNA markers are hypervariable and informative in *Cannabis sativa*: implications for forensic investigations. Forensic Sci. Int. 131, 65–74.

Hendriks, H., Malingré, Th.M., Batterman, S., Bos, R., 1978. The essential oil of *Cannabis sativa* L. Pharm. Weekblad 113, 413–424.

Herout, V., 1970. Biochemistry of sesquiterpenoids. In: Goodwin, T.W. (Ed.), Aspects of Terpenoid Chemistry and Biochemistry. Academic Press, New York, pp. 53–94.

Hillig, K.W., 2004. Genetic evidence for speciation in *Cannabis* (Cannabaceae). Genet. Res. Crop. Evol. 51, xxx–xxx (in press).

Hillig, K.W., in press. A multivariate analysis of phenotypic variation in *Cannabis*. Syst. Bot.

Hillig, K.W., Mahlberg, P.G., 2004. A chemotaxonomic analysis of cannabinoid variation in *Cannabis* (Cannabaceae). Am. J. Bot. 91, 966–975.

Hood, L.V.S., Barry, G.T., 1978. Headspace volatiles of marihuana and hashish: gas chromatographic analysis of samples of different geographic origin. J. Chrom. 166, 499–506.

Hood, L.V.S., Dames, M.E., Barry, G.T., 1973. Headspace volatiles of marijuana. Nature 242, 402–403.

Jagadish, V., Robertson, J., Gibbs, A., 1996. RAPD analysis distinguishes *Cannabis sativa* samples from different sources. Forensic. Sci. Int. 79, 113–121.

Lamarck, J.B. de., 1785. Encyclopédie méthodique. Botanique. 1 part 2, Panckoucke, Paris, pp. 694–695.

Mandolino, G., Bagatta, M., Carboni, A., Ranalli, P., de Meijer, E., 2003. Qualitative and quantitative aspects of the inheritance of chemical phenotype in *Cannabis*. J. Indust. Hemp 8 (2), 51–72.

McPartland, J.M., Clarke, R.C., Watson, D.P., 2000. Hemp Diseases and Pests. CABI Publishing, Wallingford, UK.

McPartland, J.M., Russo, E.B., 2001. *Cannabis* and *Cannabis* extracts: greater than the sum of their parts? J. Cannabis Therap. 1, 103–132.

Mediavilla, V., Steinemann, S., 1997. Essential oil of *Cannabis sativa* L. strains. J. Int. Hemp Assoc. 4, 82–84.

Meier, C., Mediavilla, V., 1998. Factors influencing the yield and the quality of hemp (*Cannabis sativa* L.) essential oil. J. Int. Hemp Assoc. 5, 16–20.

Merck Index, 1989. Guaiol (entry 4466). Merck Index. Eleventh ed. Merck and Company, Rahway, NJ, pp. 717.

Novak, J., Franz, C., 2003. Composition of the essential oils and extracts of two populations of *Cannabis sativa* L. ssp. *spontanea* from Austria. J. Essent. Oil Res. 15, 158–160.

Ross, S.A., ElSohly, M.A., 1996. The volatile oil composition of fresh and air-dried buds of *Cannabis sativa*. J. Nat. Prod. 59, 49–51.

SAS Institute, 2002. JMP Statistics and Graphics Guide. SAS Institute, Cary, North Carolina.

Schultes, R.E., Hofmann, A., 1980. The Botany and Chemistry of Hallucinogens. Charles C. Thomas, Springfield, IL.

Schultes, R.E., Klein, W.M., Plowman, T., Lockwood, T.E., 1974. *Cannabis*: an example of taxonomic neglect. Bot. Mus. Leafl. Harvard Univ. 23, 337–367.

Shirota, O., Watanabe, A., Yamazaki, M., Saito, K., Shibano, K., Sekita, S., Satake, M., 1998. Random amplified polymorphic DNA and restriction fragment length polymorphism analyses of *Cannabis sativa*. Nat. Med. 52, 160–166.

Siniscalco Gigliano, G., 2001. *Cannabis sativa* L—botanical problems and molecular approaches in forensic investigations. Forensic Sci. Rev. 13, 2–17.

Small, E., 1979. The Species Problem in *Cannabis* 1. Science, Corpus Information Services, Toronto.

Small, E., Beckstead, H.D., 1973a. Common cannabinoid phenotypes in 350 stocks of *Cannabis*. Lloydia 36, 144–165.

Small, E., Beckstead, H.D., 1973b. Cannabinoid phenotypes in *Cannabis sativa*. Nature 245, 147–148.

Small, E., Cronquist, A., 1976. A practical and natural taxonomy for *Cannabis*. Taxon 25, 405–435.

Small, E., Jui, P.Y., Lefkovitch, L.P., 1976. A numerical taxonomic analysis of *Cannabis* with special reference to species delimitation. Syst. Bot. 1, 67–84.

Turner, C.E., ElSohly, M.A., Boeren, E.G., 1980. Constituents of *Cannabis sativa* L. XVII. A review of the natural constituents. J. Nat. Prod. 43, 169–234.

Vavilov, N.I., 1926. The origin of the cultivation of ''primary'' crops, in particular cultivated hemp. In: Studies on the Origin of Cultivated Crops. Institute of Applied Botany and Plant Breeding, Leningrad, pp. 221–233.

Vavilov, N.I., Bukinich, D.D., 1929. Zemledel'cheskii Afghanistan. Trudy po Prikl. Bot. Gen. Sel. Suppl. 33, 380–382, ([Reissued 1959, Izdatel'stuo Akademii Nauk SSSR, Moskva-Leningrad]).