Genetic Resources and Crop Evolution **52:** 161–180, 2005.

Genetic evidence for speciation in *Cannabis* (Cannabaceae)

Karl W. Hillig
Department of Biology, Indiana University, Bloomington, IN, USA; Current address: 1010 Saratoga Road, Ballston Lake, NY 12019, USA (e-mail: khillig@bio.indiana.edu)

Received 7 January 2003; accepted in revised form 28 June 2003

Key words: Allozyme, *Cannabis*, Evolution, Genetics, Origin, Taxonomy

Abstract

Sample populations of 157 *Cannabis* accessions of diverse geographic origin were surveyed for allozyme variation at 17 gene loci. The frequencies of 52 alleles were subjected to principal components analysis. A scatter plot revealed two major groups of accessions. The *sativa* gene pool includes fiber/seed landraces from Europe, Asia Minor, and Central Asia, and ruderal populations from Eastern Europe. The *indica* gene pool includes fiber/seed landraces from eastern Asia, narrow-leafleted drug strains from southern Asia, Africa, and Latin America, wide-leafleted drug strains from Afghanistan and Pakistan, and feral populations from India and Nepal. A third putative gene pool includes ruderal populations from Central Asia. None of the previous taxonomic concepts that were tested adequately circumscribe the *sativa* and *indica* gene pools. A polytypic concept of *Cannabis* is proposed, which recognizes three species, *C. sativa*, *C. indica* and *C. ruderalis*, and seven putative taxa.

Abbreviations: PCA – principal components analysis

Introduction

Cannabis is believed to be one of humanity's oldest cultivated crops, providing a source of fiber, food, oil, medicine, and inebriant since Neolithic times (Chopra and Chopra 1957; Schultes 1973; Li 1974; Fleming and Clarke 1998). *Cannabis* is normally a dioecious, wind-pollinated, annual herb, although plants may live for more than a year in subtropical regions (Cherniak 1982), and monoecious plants occur in some populations (Migal 1991). The indigenous range of *Cannabis* is believed to be in Central Asia, the northwest Himalayas, and possibly extending into China (de Candolle 1885; Vavilov 1926; Zhukovsky 1964; Li 1974). The genus may have two centers of diversity, Hindustani and European–Siberian (Zeven and Zhukovsky 1975). *Cannabis* retains the ability to escape from cultivation and return to a weedy growth habit, and is considered to be only semi-domesticated (Vavilov 1926; Bredemann et al. 1956). Methods of *Cannabis* cultivation are described in the ancient literature of China, where it has been utilized continuously for at least six thousand years (Li 1974). The genus may have been introduced into Europe ca. 1500 B.C. by nomadic tribes from Central Asia (Schultes 1970). Arab traders may have introduced *Cannabis* into Africa, perhaps one to two thousand years ago (Du Toit 1980). The genus is now distributed worldwide from the equator to about 60 °N latitude, and throughout much of the southern hemisphere.

Cannabis cultivated for fiber and/or achenes (i.e., 'seeds') is herein referred to as 'hemp.' *Cannabis* breeders distinguish eastern Asian hemp from the common hemp of Europe (Bócsa and Karus 1998; de Meijer 1999). Russian botanists recognize four 'eco-geographical' groups of hemp: Northern,

Middle-Russian, Southern, and Far Eastern (Serebriakova and Sizov 1940; Davidyan 1972). The Northern hemp landraces are smaller in stature and earlier maturing than the landraces from more southerly latitudes, with a series of overlapping gradations in phenotypic traits between the Northern, Middle-Russian, and Southern types. The Far-east Asian hemp landraces are most similar to the Southern eco-geographical group (Dewey 1914). Two basic types of drug plant are commonly distinguished, in accord with the taxonomic concepts of Schultes et al. (1974) and Anderson (1980): the narrow-leafleted drug strains and the wide-leafleted drug strains (Cherniak 1982; Anonymous 1989; de Meijer 1999).

The taxonomic treatment of *Cannabis* is problematic. Linnaeus considered the genus to consist of a single undivided species, *Cannabis sativa* L. Lamarck (1785) determined that *Cannabis* strains from India are distinct from the common hemp of Europe, and named the new species *C. indica* Lam. Distinguishing characteristics include more branching, a thinner cortex, narrower leaflets, and the general ability of *C. indica* to induce a state of inebriation. Opinions differ whether Lamarck adequately differentiated *C. indica* from *C. sativa*, but they are both validly published species. Other species of *Cannabis* have been proposed (reviewed in Schultes et al. 1974; and Small and Cronquist 1976), including *C. chinensis* Delile, and *C. ruderalis* Janisch. Vavilov (1926) considered *C. ruderalis* to be synonymous with his own concept of *C. sativa* L. var. *spontanea* Vav. He later recognized wild *Cannabis* populations in Afghanistan to be distinct from *C. sativa* var. *spontanea*, and named the new taxon *C. indica* Lam. var. *kafiristanica* Vav. (Vavilov and Bukinich 1929).

Small and Cronquist (1976) proposed a monotypic treatment of *Cannabis*, which is a modification of the concepts of Lamarck and Vavilov. They reduced *C. indica* in rank to *C. sativa* L. subsp. *indica* (Lam.) Small and Cronq. and differentiated it from *C. sativa* L. subsp. *sativa*, primarily on the basis of 'intoxicant ability' and purpose of cultivation. Small and Cronquist bifurcated both subspecies into 'wild' (sensu lato) and domesticated varieties on the basis of achene size, and other achene characteristics. This concept was challenged by other botanists, who used morphological traits to delimit three species: *C. indica*, *C. sativa*, and *C. ruderalis* (Anderson 1974, 1980; Emboden 1974; Schultes et al. 1974). Schultes et al. and Anderson narrowly circumscribed *C. indica* to include relatively short, densely branched, wide-leafleted strains from Afghanistan. The differences of opinion between taxonomists supporting monotypic and polytypic concepts of *Cannabis* have not been resolved (Emboden 1981).

Few studies of genetic variation in *Cannabis* have been reported. Lawi-Berger et al. (1982) studied seed protein variation in five fiber strains and five drug strains of *Cannabis*, and found no basis for discriminating these predetermined groups. de Meijer and Keizer (1996) conducted a more extensive investigation of protein variation in bulked seed lots of 147 *Cannabis* accessions, and on the basis of five variable proteins concluded that fiber cultivars, fiber landraces, drug strains, and wild or naturalized populations could not be discriminated. A method that shows greater promise for taxonomic investigation of *Cannabis* is random amplified polymorphic DNA (RAPD) analysis. Using this technique, *Cannabis* strains from different geographic regions can be distinguished (Faeti et al. 1996; Jagadish et al. 1996; Siniscalco Gigliano 2001; Mandolino and Ranalli 2002), but the number and diversity of accessions that have been analyzed in these investigations are too small to provide a firm basis for drawing taxonomic inferences.

Allozyme analysis has proven useful in resolving difficult taxonomic issues in domesticated plants (Doebley 1989). Allozymes are enzyme variants that have arisen through the process of DNA mutation. The genetic markers (allozymes) that are commonly assayed are part of a plant's primary metabolic pathways, and presumed neutral to the effects of human selection. Through allozyme analysis, it is possible to discern underlying patterns of variation that have been outwardly obscured by the process of domestication. Because these genetic markers are cryptic, it is necessary to associate allozyme frequencies with morphological differences in order to synthesize the genetic data into a formal taxonomic treatment (Pickersgill 1988). Other types of biosystematic data may be included in the synthesis as well.

The purpose of this research is (1) to elucidate underlying genetic relationships among *Cannabis* accessions of known geographic origin, and (2) to assess previous taxonomic concepts in light of the

genetic evidence. The research reported herein is part of a broader systematic investigation of morphological, chemotaxonomic, and genetic variation in *Cannabis*, which will be reported separately.

Materials and methods

The Cannabis germplasm collection

A diverse collection of 157 *Cannabis* accessions of known geographic origin was obtained from breeders, researchers, genebanks, and law enforcement agencies (Table 1). Each accession consisted of an unspecified number of viable achenes. Many of the landraces that were studied are no longer cultivated, and exist only in germplasm repositories. Sixty-nine accessions were from hemp landraces conserved at the N.I. Vavilov Institute of Plant Industry (VIR) in Russia (Lemeshev et al. 1994). Ten accessions were from Small's taxonomic investigation of *Cannabis* (Small and Beckstead 1973; Small et al. 1976). Thirty-three accessions were from de Meijer's study of agronomic diversity in *Cannabis* (de Meijer and van Soest 1992; de Meijer 1994, 1995; de Meijer and Keizer 1996). The accessions from Afghanistan were obtained from *Cannabis* breeders in Holland, and at least three of these strains (Af-4, Af-5, Af-9) are inbred (Anonymous 1989). Six Asian accessions were collected from extant populations, including a drug landrace from Pakistan (Pk-1), three feral populations from India (In-2, In-3, In-5), and fiber landraces from India (In-4) and China (Ch-4). Accession Ch-4 was collected in Shandong Province from seed propagated on the island of Hunan (Clarke 1995). Five accessions from Central Asia were collected from roadsides and gardens in the Altai region of Russia, and identified by the provider as *C. ruderalis*. Several weedy accessions from Europe were identified as *C. ruderalis*, 'ssp. *ruderalis*,' or 'var. *spontanea*.'

A priori grouping of accessions

The accessions were assigned to drug or hemp plant-use groups, or ruderal (wild or naturalized) populations as shown in Table 1. They were also assigned to putative taxa according to the concepts of Lamarck (1785), Delile (1849), Schultes et al. (1974) and Anderson (1980), and Small and Cronquist (1976), based on morphological differences, geographic origin, and presumed reason for cultivation. Not all of the accessions could be unambiguously assigned to a taxon for each concept. To depict the various groups of interest, bivariate density ellipses were drawn on the PC scatter plot. A probability value of 0.75 was chosen because at this value the ellipses encompass the majority of accessions in a given group, but not the outliers.

Allozyme analysis

An initial survey was conducted to identify enzymes that produce variable banding patterns in *Cannabis* that can be visualized and interpreted reliably (Wendel and Weeden 1989). Eleven enzymes encoded at 17 putative loci were selected for a genetic survey of the entire *Cannabis* germplasm collection. Previously published methods of starch gel electrophoresis and staining were employed (Shields et al. 1983; Soltis et al. 1983; Morden et al. 1987; Wendel and Weeden 1989; Kephart 1990).

Gel/electrode buffer systems

Three gel/electrode buffer systems were utilized. A Tris–citrate buffer system (modified from Wendel and Weeden 1989) was used to resolve aconitase (ACN), leucine aminopeptidase (LAP), malic enzyme (ME), 6-phosphogluconate dehydrogenase (6PGD), phosphoglucoisomerase (PGI), phosphoglucomutase (PGM), and shikimate dehydrogenase (SKDH). A lithium–borate buffer system (modified from Soltis et al. 1983) was used to resolve hexokinase (HK) and triosephosphate isomerase (TPI). A morpholine–citrate buffer system (modified from Wendel and Weeden 1989) was used to resolve LAP, malate dehydrogenase (MDH), ME, PGI, PGM, and an unknown enzyme (UNK) that appeared on gels stained for isocitrate dehydrogenase (IDH). IDH could not be interpreted reliably, and was not used in the analysis. A phosphate buffer (modified from Soltis et al. 1983) was used for enzyme extraction.

Electrophoresis and staining

For both the Tris–citrate and morpholine–citrate buffer systems, 5-mm thick gels were held at 30 mA,

Table 1. Passport data for the 157 *Cannabis* accessions examined.

Origin	ID	*n*	Region/name	Use	Parallel ID	Source	Taxon
Afghanistan	Af-1	10		Drug	891383[b]	CPRO	*C. ind.*[j]; *ind. ind.*[k]
Afghanistan	Af-2	12	Ghazni	Drug	91-100[c]	AMSRS	*C. ind.*[j]; *ind. ind.*[k]
Afghanistan	Af-3	15	'Afghani No. 1'	Drug		AMSRS	*C. ind.*[j]; *ind. ind.*[k]
Afghanistan	Af-4	10	'G13'	Drug		SB	*C. ind.*[j]; *ind. ind.*[k]
Afghanistan	Af-5	10	'Hash Plant'	Drug	921199[b]	SB	*C. ind.*[j]; *ind. ind.*[k]
Afghanistan	Af-6	9	'Heavily High'	Drug	M 40	SSSC	*C. ind.*[j]; *ind. ind.*[k]
Afghanistan	Af-7	10	Mazar i Sharif	Drug	921200[b]	SB	*C. ind.*[j]; *ind. ind.*[k]
Afghanistan	Af-8	10		Drug		BPDIN	*C. ind.*[j]; *ind. ind.*[k]
Afghanistan	Af-9	10	'N. Lights 1'	Drug		SB	*C. ind.*[j]; *ind. ind.*[k]
Afghanistan	Af-10	10	Afghan mix	Drug		SB	*C. ind.*[j]; *ind. ind.*[k]
Armenia	Ar-1	8		Hemp	VIR 472[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Armenia	Ar-2	9		Hemp	VIR 482[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Belorus	Br-1	10		Hemp	VIR 296[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Bulgaria	Bg-1	10	'Lovrin 110'	Hemp	883173[b]	CPRO	*C. sat.*[i,j]; *sat. sat.*[k]
Bulgaria	Bg-2	10	Silistrenski	Hemp	901107[b]	CPRO	*C. sat.*[i,j]; *sat. sat.*[k]
Bulgaria	Bg-3	9		Hemp	VIR 73[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Bulgaria	Bg-4	7		Hemp	VIR 335[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Bulgaria	Bg-5	4		Hemp	VIR 369[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Bulgaria	Bg-6	4		Hemp	VIR 370[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Cambodia	Cm-1	10		Drug	No. 154[a]	SMALL	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
China	Ch-1	10		Hemp	901078[b]	CPRO	*C. chi.*[h]; *C. sat.*[j]; *sat. sat.*[k]
China	Ch-2	12		Rud.	No. 338[a], 921201[b]	NJBG	*C. chi.*[h]; *C. sat.*[j]; *sat. spo.*[k]
China	Ch-3	10		Hemp		NJBG	*C. chi.*[h]; *C. sat.*[j]; *sat. sat.*[k]
China	Ch-4	10	Shandong	Hemp	921198[b]	AMSRS	*C. chi.*[h]; *C. sat.*[j]; *sat. sat.*[k]
China	Ch-5	10	'Shun-Da'	Hemp	921051[b], VIR 175[d]	CPRO	*C. chi.*[h]; *C. sat.*[j]; *sat. sat.*[k]
China	Ch-6	12	'Tin-Yan'	Hemp	883249[b], VIR 184[d]	CPRO	*C. chi.*[h]; *C. sat.*[j]; *sat. sat.*[k]
China	Ch-7	15	'Shan-Va'	Hemp	921218[b], VIR 185[d]	VIR	*C. chi.*[h]; *C. sat.*[j]; *sat. sat.*[k]
Colombia	Cl-1	10		Drug		BPDIN	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
Colombia	Cl-2	10		Drug		BPDIN	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
Gambia	Gm-1	10		Drug		AMSRS	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
Germany	Gr-1	10	var. *spontanea*	Rud.	883141[b]	CPRO	*C. sat.*[i,j]; *sat. spo.*[k]
Hungary	Hn-1	10	'Szegedi-9'	Hemp	883044[b]	CPRO	*C. sat.*[i,j]; *sat. sat.*[k]
Hungary	Hn-2	10	Nyiregyházái	Hemp	883050[b]	CPRO	*C. sat.*[i,j]; *sat. sat.*[k]
Hungary	Hn-3	10	Leveleki	Hemp	883051[b]	CPRO	*C. sat.*[i,j]; *sat. sat.*[k]
Hungary	Hn-4	10	Kisszekeresi	Hemp	883058[b]	CPRO	*C. sat.*[i,j]; *sat. sat.*[k]
Hungary	Hn-5	10	var. *spontanea*	Rud.	883113[b]	CPRO	*C. sat.*[i,j]; *sat. spo.*[k]
Hungary	Hn-6	10	var. *spontanea*	Rud.	883114[b]	CPRO	*C. sat.*[i,j]; *sat. spo.*[k]
Hungary	Hn-7	12	*C. ruderalis*	Rud.	No. 316[f]	HBIPM	*C. sat.*[i,j]; *sat. spo.*[k]
Hungary	Hn-8	8		Rud.	No. 317[f]	HBIPM	*C. sat.*[i,j]; *sat. spo.*[k]
Hungary	Hn-9	10	*C. ruderalis*	Rud.	No. 1247[f]	HBIPM	*C. sat.*[i,j]; *sat. spo.*[k]
India	In-1	12	Munar, Kerala	Drug	91-194[c]	AMSRS	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
India	In-2	12	Almora	Rud.		NBPGR	*C. ind.*[i]; *C. sat.*[j]; *ind. kaf.*[k]
India	In-3	12	Delhi	Rud.		NBPGR	*C. ind.*[i]; *C. sat.*[j]; *ind. kaf.*[k]
India	In-4	12	Pauri, Garhwal	Hemp	921207[b]	INDBS	*C. chi.*[h]; *C. sat.*[j]; *sat. sat.*[k]
India	In-5	12	Saharanpur	Rud.		NBPGR	*C. ind.*[i]; *C. sat.*[j]; *ind. kaf.*[k]
Italy	It-1	10	'Kompolti'	Hemp	883048[b]	CPRO	*C. sat.*[i,j]; *sat. sat.*[k]
Italy	It-2	10		Hemp		MDCC	*C. sat.*[i,j]; *sat. sat.*[k]
Italy	It-3	12		Hemp	VIR 106[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Italy	It-4	10		Hemp	921050[b], VIR 112[d]	CPRO	*C. sat.*[i,j]; *sat. sat.*[k]
Italy	It-5	8	Turin	Hemp	VIR 195[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Italy	It-6	7	Napoletana	Hemp	VIR 278[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Italy	It-7	4	Distr. di Fatza	Hemp	VIR 280[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Italy	It-8	9	Carmagnola	Hemp	VIR 282[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Italy	It-9	4		Hemp	VIR 462[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Jamaica	Jm-1	10		Drug	No. 66[a], 921209[b]	SMALL	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
Japan	Jp-1	14			No. 152[a], 921208[b]	SMALL	*C. chi.*[h]; *C. sat.*[j]; *sat. sat.*[k]
Japan	Jp-2	18	Kozuhara zairai	Hemp	883213[b]	CPRO	*C. chi.*[h]; *C. sat.*[j]; *sat. sat.*[k]
Kazakhstan	Kz-1	9		Hemp	VIR 468[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Kazakhstan	Kz-2	9		Hemp	VIR 469[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]

Table 1. Continued.

Origin	ID	*n*	Region/name	Use	Parallel ID	Source	Taxon
Kazakhstan	Kz-3	8		Hemp	VIR 470[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Kazakhstan	Kz-4	6	Alma Ata	Hemp	VIR 484[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Lesotho	Ls-1	10		Drug		SAP	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
Mexico	Mx-1	12		Drug	No. 24[a], 921231[b]	SMALL	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
Mexico	Mx-2	8		Drug	No. 41[a]	SMALL	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
Mexico	Mx-3	12		Drug	No. 289[a], 921232[b]	SMALL	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
Mexico	Mx-4	10		Drug	921230[b]	SHOY	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
Moldavia	Ml-1	5		Hemp	VIR 116[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Nepal	Np-1	10	Kalopani	Rud.	891192[b]	CPRO	*C. ind.*[i]; *C. sat.*[j]; *ind. kaf.*[k]
Nepal	Np-2	10	Dana	Hemp	891193[b]	CPRO	*C. chi.*[h]; *C. sat.*[j]; *sat. sat.*[k]
Nepal	Np-3	10		Rud.	921233[b]	SB	*C. ind.*[i]; *C. sat.*[j]; *ind. kaf.*[k]
Nigeria	Ng-1	10		Drug		AMSRS	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
Pakistan	Pk-1	30	NW Frontier	Drug		PAKI	*C. ind.*[j]; *ind. ind.*[k]
Poland	Pl-1	7	*C.s.* 'gigantea'	Hemp	VIR 443[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Poland	Pl-2	10		Hemp	VIR 474[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Poland	Pl-3	10		Hemp	VIR 475[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Poland	Pl-4	8		Hemp	VIR 476[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Romania	Rm-1	10	ssp. *ruderalis*	Rud.	883154[b]	CPRO	*C. sativa*[i,j]; *sat. spo.*[k]
Romania	Rm-2	10	ssp. *ruderalis*	Rud.	901047[b]	CPRO	*C. sativa*[i,j]; *sat. spo.*[k]
Romania	Rm-3	10		Hemp	VIR 374[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Russia	Rs-1	6	Khakass	Rud.	N 38[g]	CSBG	*C. sat.*[i]; *C. rud.*[j]; *sat. spo.*[k]
Russia	Rs-2	5	Novosibirsk	Rud.	N 77[g]	CSBG	*C. sat.*[i]; *C. rud.*[j]; *sat. spo.*[k]
Russia	Rs-3	10	Altai	Rud.	N 79[g]	CSBG	*C. sat.*[i]; *C. rud.*[j]; *sat. spo.*[k]
Russia	Rs-4	10	Gorno-Altay	Rud.	N 82[g]	CSBG	*C. sat.*[i]; *C. rud.*[j]; *sat. spo.*[k]
Russia	Rs-5	4	Khakass	Rud.	N 102[g]	CSBG	*C. sat.*[i]; *C. rud.*[j]; *sat. spo.*[k]
Russia	Rs-6	10	Dalnevostochnaya	Hemp	921214[b], VIR 58[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Russia	Rs-7	7	Altaiskaya	Hemp	VIR 90[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Russia	Rs-8	10	Altaiskaya	Hemp	883248[b], VIR 100[d]	CPRO	*C. sat.*[i,j]; *sat. sat.*[k]
Russia	Rs-9	10	Altaiskaya	Hemp	VIR 107[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Russia	Rs-10	7	Altaiskaya	Hemp	VIR 141[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Russia	Rs-11	12	Novosibirskaya	Hemp	921217[b], VIR 142[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Russia	Rs-12	8	Ermakovskaya	Hemp	VIR 310[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Russia	Rs-13	10	Dalnevostochnaya	Hemp	VIR 387[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Russia	Rs-14	6	Trubchevskaya	Hemp	VIR 41[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Russia	Rs-15	12	Orlovskaya	Hemp	883247[b], VIR 48[d]	CPRO	*C. sat.*[i,j]; *sat. sat.*[k]
Russia	Rs-16	8	Toguchinskaya	Hemp	VIR 77[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Russia	Rs-17	7	Tyumenskaya	Hemp	VIR 85[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Russia	Rs-18	4	Smolenskaya	Hemp	VIR 110[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Russia	Rs-19	8	Permskaya	Hemp	VIR 140[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Russia	Rs-20	7	Maryiskaya	Hemp	VIR 151[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Russia	Rs-21	7	Tatarskaya	Hemp	VIR 156[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Russia	Rs-22	12	Kirovskaya	Hemp	VIR 313[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Russia	Rs-23	10	Kirovskaya	Hemp	883289[b], VIR 315[d]	CPRO	*C. sat.*[i,j]; *sat. sat.*[k]
Russia	Rs-24	10	Maryiskaya	Hemp	891327[b], VIR 349[d]	CPRO	*C. sat.*[i,j]; *sat. sat.*[k]
Russia	Rs-25	14	Chuvashskaya	Hemp	921223[b], VIR 354[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Russia	Rs-26	14	Maryiskaya	Hemp	921224[b], VIR 356[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Russia	Rs-27	10	Arkhonskaya	Hemp	921226[b], VIR 405[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Russia	Rs-28	8	Tyumenskaya	Hemp	VIR 528[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Sierra Leone	SL-1	10		Drug	No. 63[a], 921236[b]	SMALL	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
Spain	Sp-1	10		Hemp	880973[b]	CPRO	*C. sat.*[i,j]; *sat. sat.*[k]
Spain	Sp-2	10		Hemp	891240[b]	CPRO	*C. sat.*[i,j]; *sat. sat.*[k]
Spain	Sp-3	10		Hemp	921213[b], VIR 57[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Spain	Sp-4	6		Hemp	VIR 163[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
South Africa	SA-1	12	Pietersburg	Drug		SAP	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
South Africa	SA-2	10	Transkei	Drug		SAP	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
South Africa	SA-3	4	Transkei	Drug		AMSRS	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
South Africa	SA-4	10		Drug	921235[b]	DNHSA	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
South Korea	SK-1	12	Andong	Hemp	901161[b]	CPRO	*C. chi.*[h]; *C. sat.*[j]; *sat. sat.*[k]

Continued on next page

Table 1. Continued.

Origin	ID	*n*	Region/name	Use	Parallel ID	Source	Taxon
South Korea	SK-2	10	Bonghwa	Hemp	901162[b]	CPRO	*C. chi.*[h]; *C. sat.*[j]; *sat. sat.*[k]
South Korea	SK-3	10	Milyang	Hemp	901163[b]	CPRO	*C. chi.*[h]; *C. sat.*[j]; *sat. sat.*[k]
South Korea	SK-4	12	Chonnamjong	Hemp		RDASK	*C. chi.*[h]; *C. sat.*[j]; *sat. sat.*[k]
South Korea	SK-5	10	Kangwansong	Hemp	IT.180388[e]	RDASK	*C. chi.*[h]; *C. sat.*[j]; *sat. sat.*[k]
South Korea	SK-6	12	Sunchangsong	Hemp	IT.180384[e]	RDASK	*C. chi.*[h]; *C. sat.*[j]; *sat. sat.*[k]
South Korea	SK-7	12	Sungjusong	Hemp	IT.180386[e]	RDASK	*C. chi.*[h]; *C. sat.*[j]; *sat. sat.*[k]
Swaziland	Sw-1	12		Drug		SAP	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
Syria	Sy-1	10		Hemp	VIR 397[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Thailand	Th-1	12		Drug	No. 10[a]	SMALL	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
Thailand	Th-2	10	Sakon Nokhon	Drug	91-170[c]	AMSRS	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
Thailand	Th-3	12		Drug	91-171[c]	AMSRS	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
Thailand	Th-4	8		Drug	91-172.8[c]	AMSRS	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
Thailand	Th-5	10		Drug	92-176[c]	AMSRS	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
Thailand	Th-6	10		Drug		AMSRS	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
Thailand	Th-7	10	Meao, THCVA	Hemp	921237[b]	SHOY	*C. chi.*[h]; *C. sat.*[j]; *sat. sat.*[k]
Turkey	Tk-1	10	Tokumu	Hemp	883272[b]	CPRO	*C. sat.*[i,j]; *sat. sat.*[k]
Turkey	Tk-2	12		Hemp	891088[b]	CPRO	*C. sat.*[i,j]; *sat. sat.*[k]
Turkey	Tk-3	10		Hemp	891090[b]	CPRO	*C. sat.*[i,j]; *sat. sat.*[k]
Turkey	Tk-4	10		Hemp	891093[b]	CPRO	*C. sat.*[i,j]; *sat. sat.*[k]
Turkey	Tk-5	10	Kurdistan	Hemp		RBREN	*C. sat.*[i,j]; *sat. sat.*[k]
Turkey	Tk-6	7		Hemp	VIR 52[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Turkey	Tk-7	10		Hemp	VIR 54[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Turkey	Tk-8	7		Hemp	VIR 464[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Turkey	Tk-9	9		Hemp	VIR 465[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Uganda	Ug-1	10		Drug	No. 76[a]	SMALL	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
Uganda	Ug-2	10	Mbale district	Drug	921239[b]	KWNDA	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]
Ukraine	Uk-1	9	Novgorod-Severskaya	Hemp	VIR 37[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Ukraine	Uk-2	12	Transcarpathian	Hemp	921215[b], VIR 125[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Ukraine	Uk-3	12	Transcarpathian	Hemp	921216[b], VIR 126[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Ukraine	Uk-4	4	Transcarpathian	Hemp	VIR 128[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Ukraine	Uk-5	7	Transcarpathian	Hemp	VIR 130[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Ukraine	Uk-6	12		Hemp	921219[b], VIR 205[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Uzbekistan	Uz-1	5	Kokand	Rud.		AMSRS	*C. sat.*[i]; *C. rud.*[j]; *sat. spo.*[k]
Yugoslavia	Yg-1	12	Domaca local	Hemp	921210[b], VIR 11[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Yugoslavia	Yg-2	5	Nisca	Hemp	VIR 19[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Yugoslavia	Yg-3	10		Hemp	921211[b], VIR 22[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Yugoslavia	Yg-4	10		Hemp	921212[b], VIR 29[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Yugoslavia	Yg-5	7	Leskovacha	Hemp	VIR 377[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Yugoslavia	Yg-6	10	Novosadska	Hemp	VIR 442[d]	VIR	*C. sat.*[i,j]; *sat. sat.*[k]
Zimbabwe	Zm-1	10		Drug	No. 235[a], 921234[b]	SMALL	*C. ind.*[i]; *C. sat.*[j]; *ind. ind.*[k]

Origin – country of origin; ID – accession code; *n* – approximate number of plants sampled for genetic analysis (varies with enzyme); Region/Name – region where achenes were originally collected (if known)/name (if a commercial cultivar); Use – *a priori* assignment to plant-use group: Drug, Hemp, or Rud. = Ruderal (wild or naturalized); Parallel ID – parallel accession codes: [a]SMALL; [b]CPRO; [c]AMSRS; [d]VIR; [e]RDASK; [f]HBIPM; [g]CSBG.

Source: AMSRS – HortaPharm B.V., Amsterdam, Holland; BPDIN – Bloomington Police Department, Bloomington, IN, USA; CPRO – Centre for Plant Breeding and Reproduction Research, Wageningen, Holland; CSBG – Central Siberian Botanical Garden, Novosibirsk, Russia; DNHSA – Department of National Health, Pretoria, Republic of South Africa; HBIPM – Hortus Botanicus, Institui Plantarum Medicinalium, Budakalasz, Hungary; HBP – Hortus Botanicus Pekinensis, Instituti Botanici Academiae Sinicae, Beijing, China; INDBS – Botanical Survey of India, Dehra Dun, India; KWNDA – Kawanda Research Station, Kampala, Uganda; MDCC – Museo Della Civilta Contadina, Bologna, Italy; NBPGR – National Bureau of Plant Genetic Resources, New Delhi, India; NJBG – Nanjing Botanical Garden, Mem. Sun Yat-Sen, Jiangsu, China; PAKI – Pakistan Narcotics Control, Islamabad, Pakistan; RBREN – Dr. Rudolph Brenneisen, Institute of Pharmacy, Berne, Switzerland; RDASK – Rural Development Administration, Suwon, South Korea; SAP – Forensic Science Laboratory, Pretoria, Republic of South Africa; SB – The Seed Bank, Ooy, Holland (commercial seed company); SHOY – Dr Y. Shoyama, Faculty of Pharmaceutical Sciences, Kyushu University, Japan; SMALL – Dr E. Small, Biosystematics Research Institute, Ottawa, Canada; SSSC – Super Sativa Seed Club, Amsterdam, Holland (commercial seed company); VIR – N.I. Vavilov All-Union Institute of Plant Industry, St. Petersburg, Russia.

Taxon: *a priori* assignment of accessions to taxonomic concepts of [h]Delile; [i]Lamarck; [j]Schultes et al. and Anderson; [k]Small and Cronquist. Taxon abbreviations: *C. chi.* –*C. chinensis*; *C. ind.* –*C. indica*; *C. sat.* –*C. sativa*; *C. rud.* –*C. ruderalis*; *sat. sat.* – *C. sativa* subsp. *sativa* var. *sativa*; *sat. spo.* –*C. sativa* subsp. *sativa* var. *spontanea*; *ind. ind.* –*C. sativa* subsp. *indica* var. *indica*; *ind. kaf.* –*C. sativa* subsp. *indica* var. *kafiristanica*.

and 10-mm thick gels at 45 mA throughout electrophoresis. For the lithium–borate buffer system, only 5-mm thick gels were used. These were held at 50 mA for the first 10 min (after which the wicks were removed), and at 200 V subsequently. Current was applied for about 6 h to obtain good band separation. Staining recipes for all enzymes except HK were modified from Soltis et al. (1983). The HK recipe was modified from Morden et al. (1987).

Tissue sample collection

Sample populations of each accession were grown in two secure greenhouses at Indiana University, Bloomington, Indiana. Voucher specimens are deposited in the Deam Herbarium (IND) at Indiana University. About 10 plants of each accession were surveyed, except for accessions obtained late in the investigation. Thirty *Cannabis* plants were sampled for each gel. To make the gels easier to interpret, two lanes were left blank or loaded with a plant other than *Cannabis*. Tissue samples were collected the afternoon before extraction and electrophoresis, and stored overnight on moist filter paper in small Petri dishes, under refrigeration. Shoot tips generally produced the darkest bands, although mature leaf tissue was better for visualizing PGM.

Multivariate analysis

Putative genotypes were inferred from the allozyme banding patterns, and allele frequencies were calculated for small populations of each accession (Wendel and Weeden 1989). Allele frequencies were analyzed using JMP version 5.0 (SAS Institute 2002). Principal components analysis (PCA), commonly employed in numerical taxonomic investigations, was used to visualize the underlying pattern of genetic variation. The principal components were extracted from the correlation matrix of allele frequencies. Each PC axis is defined by a linear combination of the allele frequencies. PC axis 1 accounts for the largest amount of variance that can be attributed to a single multivariate axis, and each succeeding axis accounts for a progressively smaller proportion of the remaining variance. PC analysis simplifies the original n-dimensional data set (n = the number of alleles) by enabling the data to be plotted on a reduced number of orthogonal axes while minimizing the loss of information. The degree of similarity among the accessions can be inferred from their proximity in PC space (Wiley 1981; Hillig and Iezzoni 1988).

The average number of alleles per locus (A), number of alleles per polymorphic locus (Ap), and percent polymorphic loci (P) were calculated for each accession, and the expected heterozygosity (H_e) averaged over all loci was calculated using the mean allele frequencies of each sample population, for the 11 enzymes that were assayed (Nei 1987; Doebley 1989).

Several industrial hemp strains developed in European breeding programs were genetically characterized, but excluded from the statistical analysis because of their possible hybrid origin (de Meijer and van Soest 1992; de Meijer 1995). For the purpose of this investigation, an accession was considered hybrid if the parental strains came from more than one country. Nine Chinese accessions from the VIR collection were excluded because of suspected hybridization during seed regeneration. Only accessions analyzed in this investigation are shown in Table 1.

Results

Gel interpretation

The allozyme banding patterns were interpreted as shown in Figure 1. Only diploid banding patterns were observed. When more than one set of bands appeared on a gel, the loci were numbered sequentially starting with the fastest migrating (most anodal) locus. Alleles at a given locus were lettered sequentially, starting with the fastest migrating band. Monomeric enzymes (ACN, HK, LAP, PGM, SKDH, UNK) showed a single band for homozygous individuals, and two bands for heterozygous individuals. Dimeric enzymes (6PGD, MDH, PGI, TPI) typically showed one band for homozygotes, and three bands for heterozygotes. Malic enzyme (ME) is tetrameric (Weeden and Wendel 1989), and heterozygous individuals produced a five-banded pattern. Curiously, a pair of bands appeared at the bottom of gels stained for LAP due to cannabidiolic acid (CBDA) and tetrahydrocannabinolic acid (THCA) migrating into

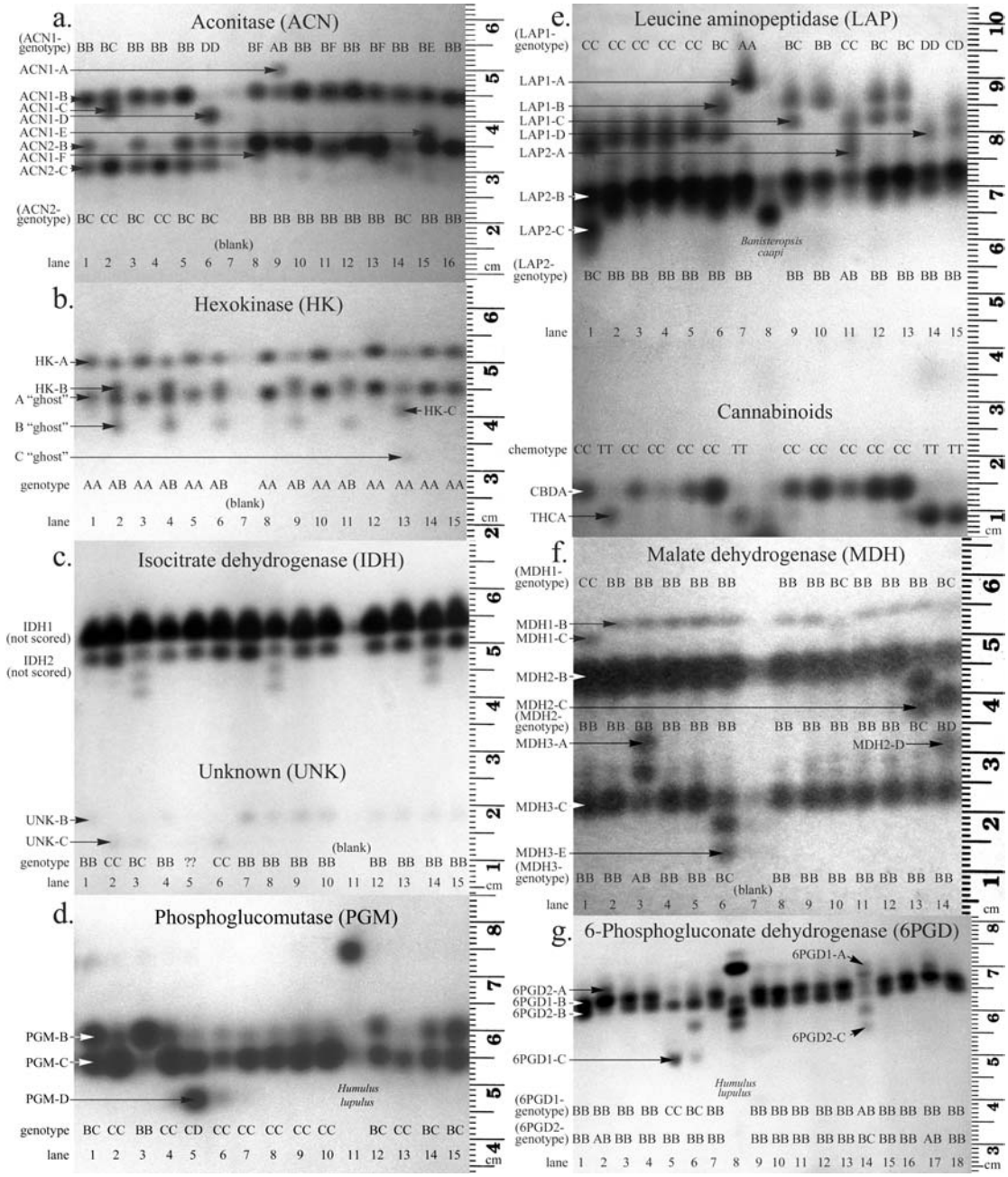

Figure 1. Starch gels stained for enzyme activity. The scale (cm) shows the distance of migration from the origin. (a) ACN; (b) HK; so-called 'ghost' bands are artifacts and can be ignored. (c) IDH (not used in analysis) and UNK; (d) PGM; (e) LAP; cannabinoids CBDA and THCA appear toward the bottom of the gel. (f) MDH; (g) 6PGD; (h) ME; (i) SKDH; (j) TPI; (k) PGI; (l) PGI; the two-banded pattern in lane 3 is attributed to the expression of a 'silent' allele (As).

the gels (Figure 1e). Cannabinoid data were not included in the statistical analysis.

A total of 65 alleles were detected for the 11 enzymes that were assayed. Thirteen of these were excluded from the analysis because they appeared in just a single accession. Although they are not useful in this study for taxonomic discrimination, these alleles may indicate regions of high genetic

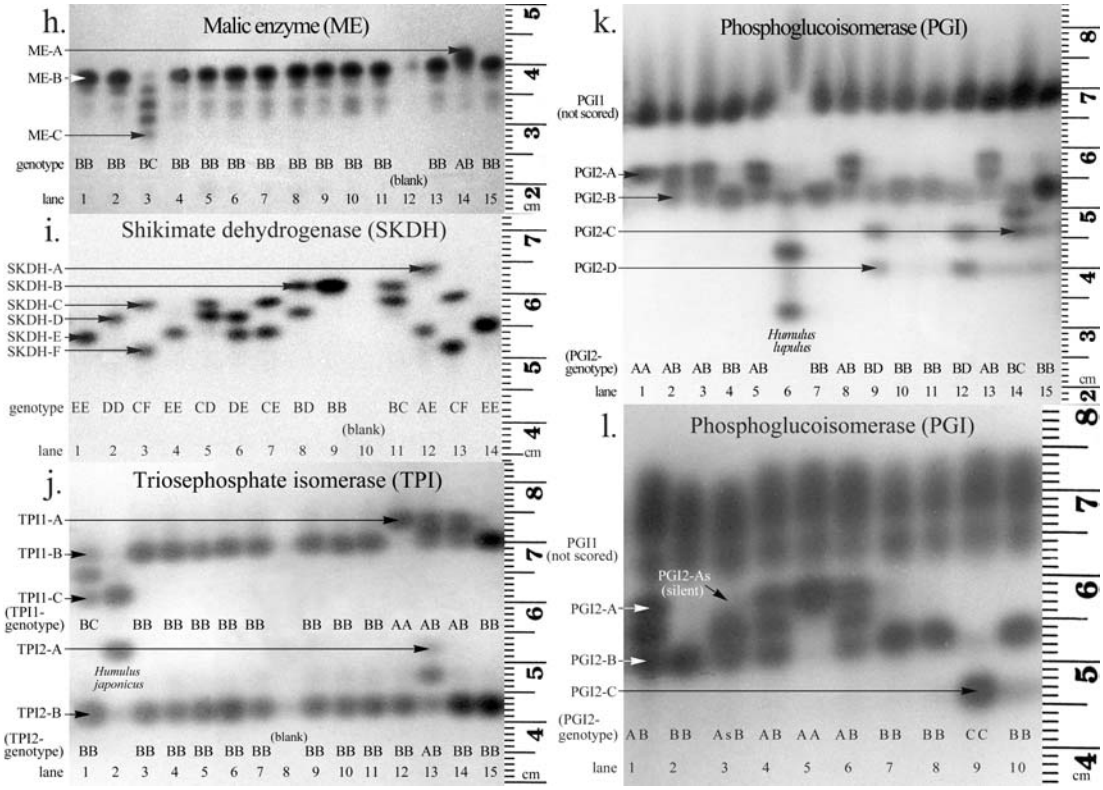

Figure 1. Continued.

diversity. Ten of the 13 rare alleles were detected in accessions from southern and eastern Asia (India, Japan, Pakistan, South Korea), and just two were detected in accessions from Europe. The 52 alleles that were detected in more than one accession were included in the statistical analysis.

Principal components analysis

The *Cannabis* accessions were plotted on PC axis 1 (PC1) and PC axis 2 (PC2), which account for 12.3 and 7.3% of the total variance, respectively (Figure 2). Two large clusters of accessions, as well as several outliers, are evident on a density contour overlay of the PC scatter plot (Figure 3). A line separating the two major groups is arbitrarily drawn at PC1 = −1. The geographic distribution of the accessions was visualized by drawing bivariate density ellipses ($P = 0.75$) on the PC plot for the 19 countries of origin represented by three or more accessions (Figure 4). It can be seen in Figure 4 that the ellipses cluster into the two major groups visualized in Figure 3. Accessions with values of PC1 > −1 are mostly from Asian and African countries, including Afghanistan, Cambodia, China, India, Japan, Nepal, Pakistan, South Korea, Thailand, and Uzbekistan, as well as Gambia, Lesotho, Nigeria, Sierra Leone, South Africa, Swaziland, Uganda, and Zimbabwe. Accessions from Colombia, Jamaica, and Mexico are also associated with this group. The other major group, with values of PC1 > −1, is comprised of accessions from Europe, Asia Minor, and Asiatic regions of the former Soviet Union, including Armenia, Belorus, Bulgaria, Germany, Hungary, Italy, Kazakhstan, Moldavia, Poland, Romania, Russia, Spain, Syria, Turkey, Ukraine, and former Yugoslavia. Although the ellipses for Russia and former Yugoslavia extend into the neighboring cluster, none of the Yugoslavian accessions, and only two of the Russian accessions (Rs-1, Rs-3) had values of PC1 < −1. The ellipse for Russia is relatively large because of several outliers, including a group of five accessions (Rs-7, Rs-9,

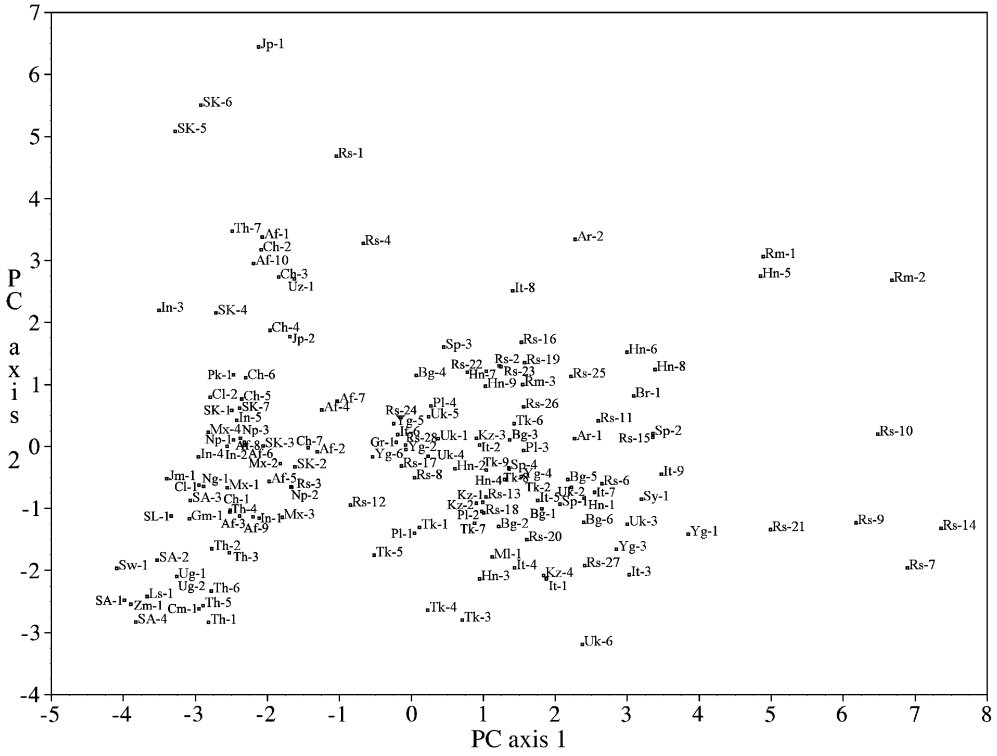

Figure 2. Scatter plot of 156 *Cannabis* accessions on PC axis 1 and PC axis 2. Accession codes are given in Table 1. Rs-5, a distant outlier, is not shown.

Rs-10, Rs-14, Rs-21), three of which are from the Altai region of Central Asia. Three ruderal accessions from the same region (Rs-1, Rs-4, Rs-5) are also outliers, but situated apart from the previous group. Two ruderal Romanian accessions (Rm-1, Rm-2) are outliers, resulting in an elongated ellipse that extends beyond the main cluster, and envelops five ruderal Hungarian accessions (Hn-5, Hn-6, Hn-7, Hn-8, Hn-9) as well.

For further analysis, accessions with values of PC1 < -1 were assigned to the *indica* gene pool, and those with values of PC1 > -1 were assigned to the *sativa* gene pool. The gene pools are so-named because they correspond (more or less) to the *indica*/*sativa* dichotomy perceived by Lamarck and others. A map showing the countries of origin of accessions from Eurasia and Africa is shaded to indicate the approximate geographic range of the *indica* and *sativa* gene pools on these continents (Figure 5). A third *ruderalis* gene pool was hypothesized, to accommodate the six Central Asian ruderal accessions (Rs-1 through Rs-5, Uz-1) situated on the PC plot between the *indica* and *sativa* gene pools. The *ruderalis* accessions correspond to Janischevsky's (1924) description of C. *ruderalis*. The indigenous range of the putative *ruderalis* gene pool is believed to be in Central Asia. A more detailed analysis of spontaneous *Cannabis* populations along the migratory routes of ancient nomadic people, ranging from Central Asia to the Carpathian Basin, may reveal further details regarding the *ruderalis* gene pool.

The frequencies (f) of 29 out of 52 alleles differed significantly ($P \leq 0.05$) between accessions assigned to the *indica* and *sativa* gene pools (Table 2). The most common allele at each locus is the same for both gene pools, but their frequencies differed significantly for 10 of the 17 loci surveyed. The absolute values of the eigenvectors (Table 2) indicate the relative contribution of each allele to a given PC axis. Several alleles that account for much of the differentiation between the two major gene pools on PC1 (ACN1-F, LAP1-B, 6PGD2-A, PGM-B, SKDH-D, UNK-C) are

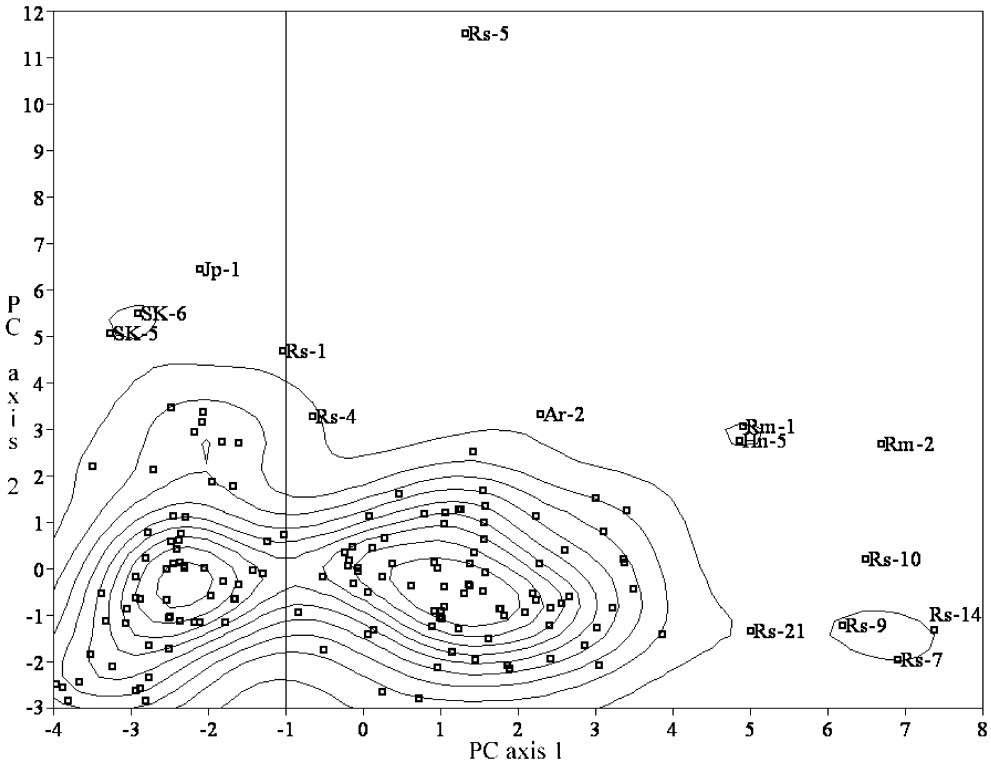

Figure 3. Density contour overlay of the PC scatter plot. The two large clusters of accessions are separated by a line drawn at PC1 = −1. Several outlying accessions are evident, including Rs-5, not shown in Figure 2. Density contours are in 10% increments, with 0.7 kernel sizes for both axes.

relatively common ($f \geq 0.10$) in the *sativa* gene pool, and uncommon ($f \leq 0.05$) in the *indica* gene pool. Four of these alleles (ACN1-F, 6PGD2-A, PGM-B, SKDH-D) are also common in the *ruderalis* gene pool. Several other alleles that largely contribute to the differentiation of accessions on PC2 (ACN1-A, LAP1-C, ME-C, UNK-A) are significantly more common in the *ruderalis* gene pool than in the *indica* or *sativa* gene pools. Only two alleles (ACN2-C, LAP1-D) were found that are common ($f \geq 0.10$) in accessions assigned to the *indica* gene pool, and uncommon in accessions assigned to the *sativa* gene pool. However, several less-common ($0.05 \leq f < 0.10$) alleles in the *indica* gene pool were uncommon or rare ($f \leq 0.03$) in the *sativa* gene pool (PGI2-C, SKDH-A, SKDH-B, SKDH-F).

The ruderal accessions from Europe and Central Asia tend to group apart. Although Rs-5 is a distant outlier, plants of this accession appeared morphologically similar to others from the same region. The outlying position of Rs-5 may be partially due to sampling error, since only four viable achenes were obtained. Allele LAP2-A is common among the ruderal accessions from Europe and Central Asia, but relatively uncommon among the other accessions in the collection, particularly those assigned to the *indica* gene pool.

The germplasm collection included two very early maturing Russian hemp accessions typical of the Northern eco-geographical group (Rs-22, Rs-23). These are situated on the PC plot with early maturing accessions from nearby regions (Rs-25, Rs-26), and with three ruderal accessions (Hn-7, Hn-9, Rs-2). However, accessions from more southerly latitudes in Europe also cluster nearby (Bg-4, Rm-3, Sp-3). No formal distinction was made in this investigation between the Middle-Russian and Southern eco-geographic groups of hemp, or between fiber and seed accessions. There appears to be little basis for differentiating these groups on the PC scatter plot. The large ellipse for Russia (Figure 4) envelops accessions assigned to both the *sativa* and *ruderalis* gene pools. Allele

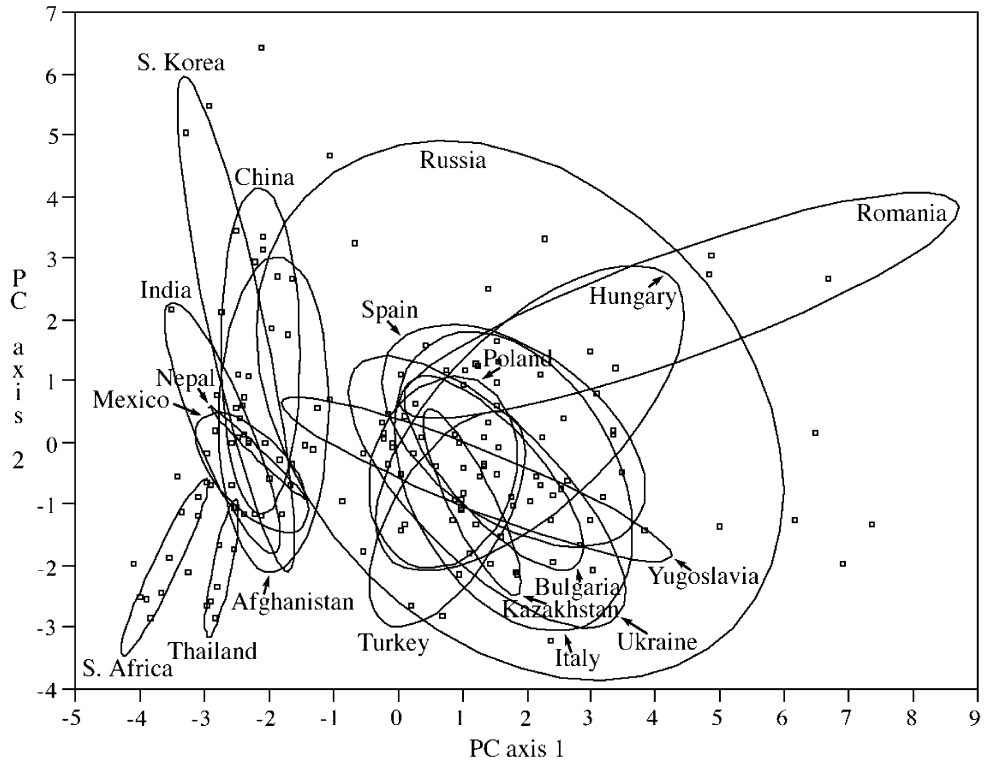

Figure 4. Density ellipses ($P = 0.75$) are drawn on the PC scatter plot for the countries of origin of the various accessions. Ellipses were only generated for countries represented by a minimum of three accessions.

MDH2-C was detected in four of the five Russian outliers situated toward the right side of the PC scatter plot (Rs-7, Rs-9, Rs-14, Rs-21). This allele was not found in any of the other accessions. The taxonomic significance of this group, if any, is unknown.

The fiber/seed accessions assigned to the *indica* gene pool are genetically diverse. All but six of the 57 alleles detected in the *indica* gene pool were present in this group, including seven rare alleles that were detected in just a single accession. The outliers in the upper left corner of the PC scatter plot are mostly hemp landraces from eastern Asia that had allele frequencies outside the normal range, which sets them apart from the other *indica* accessions.

The narrow-leafleted drug accessions are relatively devoid of genetic variation, compared to the other conceptual groups recognized in this study. Even so, geographic patterns of genetic variation are apparent within this group. The 12 African accessions are from three regions: western Africa (Nigeria, Gambia, Sierra Leone), east-central Africa (Uganda) and southern Africa (South Africa, Swaziland, Lesotho, Zimbabwe). Sample populations of the two Ugandan accessions (Ug-1, Ug-2) consisted entirely of monoecious plants devoid of detectable allozyme variation. The position of these two accessions on the PC scatter plot represents a region of low genetic variation, with drug accessions from southern Africa and Southeast Asia situated nearby. A rare allele (SKDH-A) was found in all seven southern African accessions, but in only two other accessions, from Nigeria and Colombia. For the African accessions, an allele (SKDH-C) that was commonly found in most other accessions was not detected.

The wide-leafleted drug accessions from Afghanistan and Pakistan (Af-1 thru Af-10, Pk-1) cluster with the other accessions assigned to the *indica* gene pool. Allele HK-B was found in nine of the 11 wide-leafleted drug accessions, and in a few hemp accessions from China and South Korea, but not in any of the narrow-leafleted drug accessions

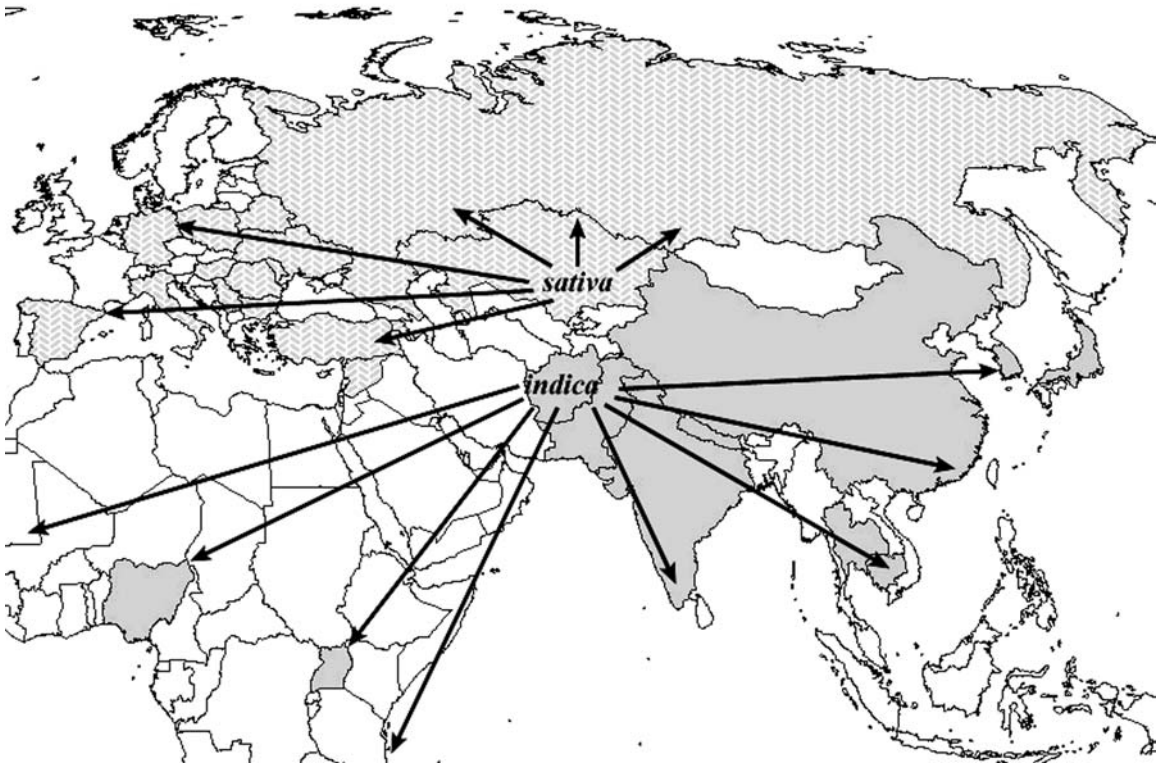

Figure 5. Map showing the countries of origin of accessions assigned to the *indica* and *sativa* gene pools. The arrows suggest human-vectored dispersal from the presumed origin of *Cannabis* in Central Asia.

or feral *indica* accessions. HK-B is common in the *sativa* gene pool, being found in 60 of the 89 accessions assigned to that group. However, several other alleles that are common in the *sativa* gene pool (ACN1-F, LAP1-B, 6PGD2-A, PGM-B, TPI1-A, UNK-C) were rare or undetected in the wide-leafleted drug accessions.

Taxonomic interpretation

One objective of this study is to assess previous taxonomic concepts in light of the genetic evidence. *Cannabis* is commonly divided into drug and hemp plant-use groups, and a third group of ruderal (wild or naturalized) populations. The density ellipse for the drug accessions (Figure 6a) overlies the *indica* gene pool, while the ellipse for the hemp accessions overlies both major gene pools, as does the ellipse for the ruderal accessions.

Delile's (1849) concept of *C. chinensis* is given consideration, because hemp accessions from southern and eastern Asia group separately from those assigned to the *sativa* gene pool, and Delile was the first taxonomist to describe a separate taxon of eastern Asian hemp. The density ellipse for accessions assigned to *C. chinensis* (Figure 6b) shows that they comprise a subset of the *indica* gene pool.

Lamarck's (1785) taxonomic concept differentiates the narrow-leafleted *C. indica* drug accessions from *C. sativa*, but it is ambiguous how he would have classified the wide-leafleted drug accessions, or the eastern Asian hemp accessions. Figure 6c shows good separation of the two species proposed by Lamarck, but his concept of *C. indica* does not circumscribe all of the accessions assigned to the *indica* gene pool.

Schultes et al. (1974) and Anderson (1980) narrowly circumscribed *C. indica* to include wide-leafleted strains from Afghanistan. The narrow-leafleted drug strains, together with hemp strains from all locations are circumscribed under

Table 2. Mean allele frequencies for accessions assigned to the *indica*, *sativa* and *ruderalis* gene pools. For a given allele, means (in rows) not connected by the same letter are significantly different using Student's *t*-test ($P = 0.05$). The most common allele at each locus is shown in bold. n = number of accessions assigned to each group. Also shown are the Eigenvectors for the first two principal component axes (PC1 and PC2).

Allele	*indica* $n = 62$ Mean	*sativa* $n = 89$ Mean	*ruderalis* $n = 6$ Mean	Eigenvector	
				PC1	PC2
ACN1-A	0.02 b	0.01 b	0.11 a	−0.039	0.280
ACN1-B	0.95 a	0.89 b	0.79 b	−0.082	−0.183
ACN1-D	0.02 a	0.00 a	0.02 a	−0.023	0.039
ACN1-E	0.01 a	0.00 a	0.00 a	−0.045	0.067
ACN1-F	0.00 b	0.10 a	0.09 a	0.161	0.025
ACN2-B	0.90 b	0.99 a	0.80 b	0.105	−0.342
ACN2-C	0.10 a	0.01 b	0.20 a	−0.104	0.341
HK-A	0.92 a	0.85 b	0.82 ab	−0.080	−0.189
HK-B	0.08 b	0.15 a	0.18 ab	0.080	0.187
LAP1-A	0.00 a	0.01 a	0.00 a	0.095	0.082
LAP1-B	0.03 b	0.33 a	0.00 b	0.231	−0.154
LAP1-C	0.68 b	0.64 b	0.93 a	−0.037	0.288
LAP1-D	0.30 a	0.03 b	0.07 b	−0.190	−0.189
LAP2-A	0.01 b	0.07 a	0.20 a	0.126	0.178
LAP2-B	0.99 a	0.92 b	0.81 b	−0.154	−0.175
LAP2-C	0.00 a	0.02 a	0.00 a	0.140	0.030
MDH1-A	0.01 a	0.00 a	0.00 a	−0.017	0.017
MDH1-B	0.99 a	0.94 b	0.93 ab	−0.218	−0.132
MDH1-C	0.00 b	0.06 a	0.07 ab	0.237	0.133
MDH2-B	1.00 a	0.99 a	1.00 a	−0.154	0.059
MDH2-C	0.00 a	0.01 a	0.00 a	0.156	−0.060
MDH3-A	0.00 a	0.00 a	0.00 a	0.030	0.067
MDH3-C	0.99 a	0.98 a	0.97 a	−0.045	−0.092
MDH3-E	0.00 b	0.02 a	0.03 a	0.077	0.041
ME-B	0.99 a	0.99 a	0.93 b	0.011	−0.160
ME-C	0.01 b	0.01 b	0.07 a	−0.004	0.168
6PGD1-A	0.00 a	0.00 a	0.00 a	−0.047	0.040
6PGD1-B	0.99 a	1.00 a	1.00 a	0.038	−0.143
6PGD2-A	0.02 b	0.17 a	0.15 a	0.252	0.045
6PGD2-B	0.98 a	0.82 b	0.85 b	−0.249	−0.044
6PGD2-C	0.00 a	0.00 a	0.00 a	0.022	−0.011
PGI2-A	0.08 b	0.21 a	0.00 b	0.143	−0.066
PGI2-As	0.01 a	0.00 a	0.00 a	−0.036	0.111
PGI2-B	0.86 a	0.79 b	0.98 a	−0.095	0.033
PGI2-C	0.05 a	0.00 b	0.02 ab	−0.083	0.025
PGM-B	0.01 c	0.34 a	0.20 b	0.294	−0.011
PGM-C	0.98 a	0.66 c	0.80 b	−0.291	0.009
PGM-D	0.01 a	0.00 b	0.00 ab	−0.035	0.040
SKDH-A	0.05 a	0.00 b	0.00 ab	−0.124	−0.123
SKDH-B	0.09 a	0.02 b	0.00 ab	−0.104	−0.058
SKDH-C	0.31 a	0.37 a	0.04 b	0.083	−0.132
SKDH-D	0.05 b	0.14 a	0.20 a	0.137	0.105
SKDH-E	0.42 b	0.43 ab	0.63 a	−0.036	0.068
SKDH-F	0.08 a	0.03 b	0.13 a	−0.098	0.239
TPI1-A	0.05 b	0.10 a	0.11 ab	0.098	0.097
TPI1-B	0.95 a	0.90 b	0.89 ab	−0.096	−0.097
TPI2-A	0.01 a	0.01 a	0.00 a	−0.023	0.034
TPI2-B	0.99 a	0.99 a	1.00 a	0.019	−0.013
TPI2-C	0.00 a	0.00 a	0.00 a	0.005	−0.049
UNK-A	0.00 b	0.00 b	0.03 a	0.029	0.219
UNK-B	0.99 a	0.60 b	0.97 a	−0.305	0.114
UNK-C	0.01 b	0.39 a	0.00 b	0.304	−0.126

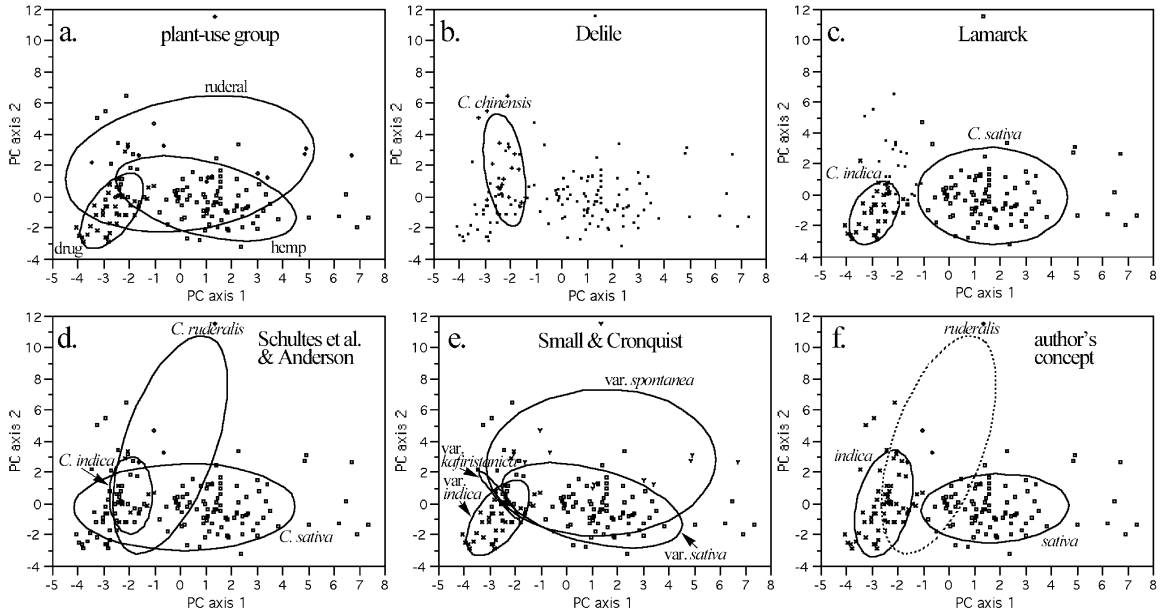

Figure 6. The PC scatter plot, with density ellipses ($P = 0.75$) showing how well various conceptual groups coincide with the genetic data. The accessions were sorted according to the following concepts: (a) plant-use group; (b) Delile; (c) Lamarck; (d) Schultes et al. and Anderson; (e) Small and Cronquist; (f) author's concept.

C. sativa. The density ellipse for *C. indica* shows that the accessions assigned to this concept comprise a subset of the *indica* gene pool (Figure 6d), while the ellipse for *C. sativa* includes accessions assigned to both the *indica* and *sativa* gene pools. Schultes et al. and Anderson also recognized *C. ruderalis*, and emphasized that it only exists in regions where *Cannabis* is indigenous. The ellipse for the six Central Asian accessions assigned to *C. ruderalis* lies between and overlaps both the *indica* and *sativa* gene pools.

Small and Cronquist (1976) proposed two subspecies and four varieties of *C. sativa*. Their circumscription of *C. sativa* L. subsp. *sativa* var. *sativa* includes hemp strains from all regions, and the resulting ellipse overlaps the *indica* and *sativa* gene pools (Figure 6e). *C. sativa* L. subsp. *sativa* var. *spontanea* (Vav.) Small and Cronq. includes ruderal accessions from both Europe and Central Asia. The resulting ellipse encompasses most of the *sativa* gene pool and a portion of the *indica* gene pool, although only two accessions assigned to var. *spontanea* (Rs-1, Rs-3) had values of PC1 < -1. The density ellipses for *C. sativa* L. subsp. *indica* Lam. var. *indica* (Lam.) Wehmer, and for *C. sativa* L. subsp. *indica* Lam. var. *kafiristanica* (Vav.) Small and Cronq. encompass different subsets of the *indica* gene pool.

The author's concept is illustrated by density ellipses for the *indica*, *sativa*, and *ruderalis* gene pools (Figure 6f). The ellipses for accessions assigned to the *indica* and *sativa* gene pools overlay the two major clusters of accessions, while the ellipse for the *ruderalis* accessions is intermediate, and overlaps the other two. Since the existence of a separate *ruderalis* gene pool is less certain, it is indicated with a dotted line.

Genetic diversity statistics

Genetic diversity statistics for gene pools and putative taxa of *Cannabis* are given in Table 3. The taxa listed in Table 3 circumscribe different subsets of the *indica* and *sativa* gene pools. *C. ruderalis* is also included here. The circumscriptions of *C. sativa* subsp. *sativa* var. *sativa* and *C. sativa* subsp. *sativa* var. *spontanea* exclude accessions assigned to *C. chinensis* and *C. ruderalis*, respectively, while *C. indica* sensu Lamarck excludes accessions assigned to *C. sativa* subsp. *indica* var. *kafiristanica*. In general, the *sativa* accessions exhibited greater genetic diversity than the *indica* accessions

Table 3. Means for the number of alleles per locus (A), number of alleles per polymorphic locus (Ap), percentage of polymorphic loci (P) and average expected heterozygosity (He) for gene pools and putative taxa of *Cannabis*. Means (in columns) not connected by the same letter are significantly different using Student's *t*-test ($P = 0.05$). The gene pools and putative taxa were tested separately. n = number of accessions.

	n	A	Ap	P	He
Gene pool					
sativa	89	1.60 a	2.20 b	48.3 a	0.17 a
indica	62	1.35 b	2.39 a	22.2 c	0.08 c
ruderalis	6	1.39 b	2.13 b	34.0 b	0.13 b
Putative taxon					
C. sativa subsp. *sativa* var. *sativa*[a] Small and Cronq.	81	1.60 a	2.20 bc	48.4 a	0.17 a
C. sativa subsp. *sativa* var. *spontanea*[b] Small and Cronq.	8	1.59 ab	2.19 bc	47.0 a	0.17 a
C. sativa subsp. *indica* var. *kafiristanica* Small and Cronq.	5	1.44 bc	2.38 ab	22.4 cde	0.09 cd
C. indica Lam.[c]	27	1.19 d	2.43 a	12.8 e	0.05 e
C. indica sensu Schultes et al. and Anderson	11	1.29 c	2.21 bc	22.1 d	0.07 d
C. chinensis Delile	19	1.59 a	2.44 a	35.6 b	0.12 bc
C. ruderalis Janisch.	6	1.39 c	2.13 c	34.0 bc	0.13 b

[a]Excluding accessions assigned to *C. chinensis*.
[b]Excluding accessions assigned to *C. ruderalis*.
[c]Excluding accessions assigned to *C. sativa* subsp. *indica* var. *kafiristanica*.

(including *C. sativa* subsp. *indica* var. *kafiristanica* and *C. chinensis*), and the *ruderalis* accessions were intermediate. Within the *indica* gene pool, the accessions assigned to *C. chinensis* exhibited the greatest genetic diversity, and the narrow-leafleted drug accessions (*C. indica* sensu Lamarck) exhibited the least. Within the *sativa* gene pool, the cultivated (var. *sativa*) and weedy (var. *spontanea*) accessions exhibited virtually identical levels of genetic diversity.

Discussion

The allozyme data show that the *Cannabis* accessions studied in this investigation were derived from two major gene pools, ruling out the hypothesis of a single undivided species. The genetic divergence of the cultivated accessions approximates the *indica*/*sativa* split perceived by previous investigators. However, none of the earlier taxonomic treatments of *Cannabis* adequately represent the underlying relationships discovered in the present study.

The allozyme data, in conjunction with the different geographic ranges of the *indica* and *sativa* gene pools and previous investigations that demonstrate significant morphological and chemotaxonomic differences between these two taxa (Small and Beckstead 1973; Small et al. 1976), support the formal recognition of *C. sativa*, *C. indica*, and possibly *C. ruderalis* as separate species. This opinion represents a synthesis of the species concepts of Lamarck, Delile, Janischevsky, Vavilov, Schultes et al. and Anderson. It rejects the single-species concepts of Linnaeus, and Small and Cronquist, because the genetic data demonstrate a fundamental split within the *Cannabis* gene pool. It is more 'practical and natural' to assign the *indica* and *sativa* gene pools to separate species, and to leave the ranks of subspecies and variety available for further classification of the putative taxa recognized herein.

The *C. sativa* gene pool includes hemp landraces from Europe, Asia Minor and Central Asia, as well as weedy populations from Eastern Europe. The *C. indica* gene pool is more diverse than Lamarck originally conceived. Besides the narrow-leafleted drug strains, the *C. indica* gene pool includes wide-leafleted drug strains from Afghanistan and Pakistan, hemp landraces from southern and eastern Asia, and feral populations from India and Nepal. *C. ruderalis*, assumed to be indigenous to Central Asia, is delimited to exclude naturalized *C. sativa* populations occurring in regions where *Cannabis* is not native. The existence of a separate *C. ruderalis* gene pool is less certain, since only six accessions of this type were available for study.

The first two PC axes account for a relatively small proportion of the total variance (19.6%), compared with a typical PC analysis of

morphological data. Morphological data sets often have a high degree of 'concomitant character variation,' such as the size correlation between different plant parts (Small 1979). As a result, the first few PC axes often account for a relatively large proportion of the variance. This type of 'biological correlation' was absent from the data set of allele frequencies. Although the less common alleles are of taxonomic importance, the common alleles largely determined the outcome of the PC analysis. When only the most frequent allele at each locus was entered into the analysis, the first two PC axes accounted for 25.8% of the total variance, and the *C. indica* and *C. sativa* gene pools were nearly as well discriminated.

The role of human selection in the divergence of the *C. indica* and *C. sativa* gene pools is uncertain. Small (1979) presumed the dichotomy to be largely a result of selection for drug production in the case of the *indica* taxon, and selection for fiber/seed production in the case of *sativa*. The genetic evidence challenges this assumption, since the fiber/seed accessions from India, China, Japan, South Korea, Nepal, and Thailand all cluster with the *C. indica* gene pool. An alternate hypothesis is that the *C. indica* and *C. sativa* hemp landraces were derived from different primordial gene pools and independently domesticated, and that the drug strains were derived from the same primordial gene pool as the *C. indica* hemp landraces. It is assumed that, in general, when humans introduced *Cannabis* into a region where it did not previously exist, the gene pool of the original introduction largely determined the genetic make-up of the *Cannabis* populations inhabiting the region thereafter. It remains to be determined whether the *C. indica* and *C. sativa* gene pools diverged before, or after the beginning of human intervention in the evolution of *Cannabis*.

The amount of genetic variation in *Cannabis* is similar to levels reported for other crop plants (Doebley 1989). Hamrick (1989) compiled data from different sources that show relatively high levels of genetic variation within out-crossed and wind-pollinated populations, and low levels of variation within weedy populations. Differentiation between populations is relatively low for dioecious and out-crossed populations, and high for annuals and plants (such as *Cannabis*) with gravity-dispersed seeds. Hamrick reported the within-population means of 74 dicot taxa. The number of alleles per locus (1.46), percentage of polymorphic loci (31.2%) and mean heterozygosity (0.113) are within the ranges estimated for the putative taxa of *Cannabis*. The extensive overlap of the density ellipses for the countries of origin of accessions assigned to the *C. sativa* gene pool (Figure 4) suggests that this group is relatively homogeneous throughout its range. In comparison, the ellipses for the *C. indica* gene pool do not all overlap, suggesting that regional differences within this gene pool are more distinct.

Divergence in allele frequencies between populations (gene pools) can occur in two principle ways (Witter, cited in Crawford 1989). Initially, a founder population can diverge partly or wholly by genetic drift. The second process, which presumably takes much longer, involves the accumulation of new mutations in the two populations. Both of these processes may help to explain the patterns of genetic variation present in *Cannabis*, albeit on a larger scale. The alleles that differentiate *C. indica* from *C. sativa* on PC1 are common in the *C. sativa* gene pool and uncommon in the *C. indica* gene pool, which suggests that a founder event may have narrowed the genetic base of *C. indica*. However, a considerable number of mutations appear to have subsequently accumulated in both gene pools, indicating that the *indica*/*sativa* split may be quite ancient.

The assumption that the alleles that were surveyed in this study are selectively neutral does not imply that humans have not affected allele frequencies in *Cannabis*. It only means that these genetic markers are 'cryptic' and not subject to deliberate manipulation. Humans have undoubtedly been instrumental in both the divergence and mixing of the *Cannabis* gene pools. For example, the commercial hemp strain 'Kompolti Hybrid TC' takes advantage of heterosis (hybrid vigor) in a cross between a European hemp strain corresponding to *C. sativa*, and a Chinese 'unisexual' hemp strain corresponding to *C. indica* (Bócsa 1999). Evidence of gene flow from eastern Asian hemp to cultivated *C. sativa* is provided by certain alleles (e.g., LAP1-D, PGI2-C, SKDH-B, SKDH-F) that occur in low frequency in the *C. sativa* gene pool, and are significantly more common among the hemp accessions assigned to *C. indica*. There is also limited evidence of gene flow in the reverse direction; allele

PGM-B, which is common in accessions assigned to *C. sativa*, was detected at low frequency in a few of the hemp accessions assigned to *C. indica*.

Some of the accessions in the collection encompass little genetic variation, which may be a result of inbreeding, genetic drift, or sampling error (e.g., the achenes may have been collected from a single plant). In general, the accessions cultivated for drug use, particularly the narrow-leafleted drug accessions, show more signs of inbreeding than those cultivated for fiber or seed. The absence of allele PGM-B in the gene pool of narrow-leafleted drug accessions indicates a lack of gene flow from *C. sativa*. Although it is possible that the entire gene pool of narrow-leafleted drug strains passed through a 'genetic bottleneck,' the low genetic diversity of this group may also be a result of the way these plants are often cultivated. It is not unusual for growers to select seeds from the few best plants in the current year's crop to sow the following year, thereby reducing the genetic diversity of the initial population. Since staminate plants are often culled before flowering, the number of pollinators may also be extremely limited.

The gene pool of a cultivated taxon is expected to contain a subset of the alleles present in the ancestral gene pool (Doebley 1989). In the case of *Cannabis*, the available evidence is insufficient to make an accurate determination of progenitor–derivative relationships. Aboriginal populations may have migrated from Central Asia into Europe as 'camp followers,' along with the cultivated landraces (Vavilov 1926). If so, then the weedy populations of Europe may represent the aboriginal gene pool into which individuals that have escaped from cultivation have merged. Although fewer alleles were detected in the ruderal accessions from Central Asia and Europe than in the cultivated *C. sativa* gene pool, this result is preliminary given the relatively small number of ruderal accessions available for study. Similarly, the feral *C. indica* accessions from India and Nepal do not encompass as much genetic variation as the cultivated accessions of *C. indica*, but again this result is based on insufficient data to draw firm conclusions. Even so, both results suggest that ruderal (feral) populations are secondary to the domesticated ones. From the evidence at hand, it appears that the feral *C. indica* accessions could represent the ancestral source of the narrow-leafleted drug accessions, but perhaps not of the wide-leafleted drug accessions, since allele HK-B was found in nine of the 11 wide-leafleted drug accessions, but not in any of the ruderal *C. indica*, or narrow-leafleted drug accessions. Vavilov and Bukinich (1929) reported finding wild *Cannabis* populations in eastern Afghanistan (*C. indica* Lam. f. *afghanica* Vav.), which could represent the progenitor of the wide-leafleted drug strains. Unfortunately, wild populations from Afghanistan were not represented in the present study.

Conclusion

This investigation substantiates the existence of a fundamental split within the *Cannabis* gene pool. A synthesis of previous taxonomic concepts best describes the underlying patterns of variation. The progenitor–derivative relationships within *Cannabis* are not well understood, and will require more extensive sampling and additional genetic analyses to further resolve. A revised circumscription of the infraspecific taxonomic groups is warranted, in conjunction with analyses of morphological and chemotaxonomic variation within the germplasm collection under study.

Acknowledgements

I am grateful to Professor Paul G. Mahlberg for facilitating this investigation. Thanks also to Professor Gerald Gastony and Valerie Savage for technical assistance, and to Dr Etienne de Meijer, David Watson and the others who donated germplasm for this study. I appreciate the help of Drs Beth and William Hillig, Dr John McPartland, Dr Paul Mahlberg, and two anonymous referees in reviewing this manuscript. This research was supported by a grant from HortaPharm B.V., The Netherlands.

References

Anderson L.C. 1974. A study of systematic wood anatomy in *Cannabis*. Harvard Univ. Bot. Mus. Leafl. 24: 29–36.

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

Anonymous 1989. The Seed Bank Catalogue. Ooy, The Netherlands. [authorship attributed to N. Schoenmakers.]
Bócsa I. 1999. Genetic improvement: conventional approaches. In: Ranalli P. (ed.), Advances in Hemp Research, Haworth Press, Binghamton, NY, pp. 153–184.
Bócsa I. and Karus M. 1998. The Cultivation of Hemp. Hemptech, Sebastopol, CA.
Bredemann G., Schwanitz Fr. and von Sengbusch R. 1956. Problems of modern hemp breeding, with particular reference to the breeding of varieties with little or no hashish. Bull. Narc. 8: 31–35.
de Candolle A. 1885. Hemp – *Cannabis sativa* L. In: Origin of Cultivated Plants, D. Appleton, New York, pp. 148–149.
Cherniak L. 1982. The Great Books of *Cannabis*, vol. I, Book II. Cherniak/Damele Publishing, Oakland, CA.
Chopra I.C. and Chopra R.N. 1957. The use of *Cannabis* drugs in India. Bull. Narc. 9: 4–29.
Clarke R.C. 1995. Hemp (*Cannabis sativa* L.) cultivation in the Tai'an district of Shandong Province, Peoples Republic of China. J. Int. Hemp Assoc. 2: 57, 60–65.
Crawford D.J. 1989. Enzyme electrophoresis and plant systematics. In: Soltis D.E. and Soltis P.S. (eds), Isozymes in Plant Biology, Dioscorides Press, Portland, OR, pp. 146–164.
Davidyan G.G. 1972. Hemp: biology and initial material for breeding. Tr. Prikl. Bot. Genet. Sel. 48: 1–160 (in Russian).
Delile R.-A. 1849. Index seminum horti botanici Monspeliensis. Ann. Sci. Nat. Bot. Biol. Veg. 12: 365–366.
Dewey L.H. 1914. Hemp. In: USDA Yearbook 1913. United States Department of Agriculture, Washington, DC, pp. 283–347.
Doebley J. 1989. Isozymic evidence and the evolution of crop plants. In: Soltis D.E. and Soltis P.S. (eds), Isozymes in Plant Biology, Dioscorides Press, Portland, OR, pp. 165–191.
Du Toit B.M. 1980. *Cannabis* in Africa. A.A. Balkema, Rotterdam, The Netherlands.
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. and Ranalli P. 1996. Genetic diversity of *Cannabis sativa* germplasm based on RAPD markers. Plant Breed. 115: 367–370.
Fleming M.P. and Clarke R.C. 1998. Physical evidence for the antiquity of *Cannabis sativa* L. J. Int. Hemp Assoc. 5: 80–93.
Hamrick J.L. 1989. Isozymes and the analysis of genetic structure in plant populations. In: Soltis D.E. and Soltis P.S. (eds), Isozymes in Plant Biology, Dioscorides Press, Portland, OR, pp. 87–105.
Hillig K.W. and Iezzoni A.F. 1988. Multivariate analysis of a sour cherry germplasm collection. J. Am. Soc. Hort. Sci. 113: 928–934.
Jagadish V., Robertson J. and Gibbs A. 1996. RAPD analysis distinguishes *Cannabis sativa* samples from different sources. Forensic Sci. Int. 79: 113–121.
Janischevsky D.E. 1924. Forma konopli na sornykh mestakh v Yugo-vostochnoi Rossii. In: Chiuevsky I.A. (ed.), Uchen. Zap. Gosud. Saratovsk. Chernyshevskogo Univ., 2(2): 3–17, Saratov University Press, Saratov, USSR.
Kephart S.R. 1990. Starch gel electrophoresis of plant isozymes: a comparative analysis of techniques. Am. J. Bot. 77: 316–368.
de Lamarck J.B. 1785. Encyclopédique Méthodique de Botanique, vol. 1, Pt. 2. Paris, France, pp. 694–695.
Lawi-Berger C., Miège M.N., Kapétanidis I. and Miège J. 1982. Contribution a l'etude chimiotaxonomique de *Cannabis sativa* L. CR Acad. Sci. Paris 295: 397–402.
Lemeshev N., Rumyantseva L. and Clarke R.C. 1994. Maintenance of *Cannabis* germplasm in the Vavilov Research Institute gene bank – 1993. J. Int. Hemp Assoc. 1(1): 3–5.
Li H.-L. 1974. An archaeological and historical account of *Cannabis* in China. Econ. Bot. 28: 437–448.
Mandolino G. and Ranalli P. 2002. The applications of molecular markers in genetics and breeding of hemp. J. Indust. Hemp. 7: 7–23.
de Meijer E.P.M. 1994. Diversity in *Cannabis*. Doctoral thesis, Wageningen Agricultural University, Wageningen, The Netherlands.
de Meijer E.P.M. 1995. Fibre hemp cultivars: a survey of origin, ancestry, availability and brief agronomic characteristics. J. Int. Hemp Assoc. 2: 66–73.
de Meijer E.P.M. 1999. *Cannabis* germplasm resources. In: Ranalli P. (ed.), Advances in Hemp Research, Haworth Press, Binghamton, NY, pp. 133–151.
de Meijer E.P.M. and Keizer L.C.P. 1996. Patterns of diversity in *Cannabis*. Genet. Resour. Crop Evol. 43: 41–52.
de Meijer E.P.M. and van Soest L.J.M. 1992. The CPRO *Cannabis* germplasm collection. Euphytica 62: 201–211.
Migal N.D. 1991. Genetics of polymorphic sex evolution in hemp. Genetika 27: 1561–1569. (Translated in Soviet Genetics, March 1992: 1095–1102.)
Morden C.W., Doebley J. and Schertz K.F. 1987. A Manual of Techniques for Starch Gel Electrophoresis of Sorghum Isozymes. Texas Agric. Exp. Sta. Misc. Publ. 1635, College Station, TX.
Nei M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.
Pickersgill B. 1988. The genus *Capsicum*: A multidisciplinary approach to the taxonomy of cultivated and wild plants. Biol. Zentralbl. 107: 381–389.
SAS Institute, 2002. JMP Statistics and Graphics Guide. SAS Institute, Cary, NC.
Schultes R.E. 1970. Random thoughts and queries on the botany of *Cannabis*. In: Joyce C.R.B. and Curry S.H. (eds), The Botany and Chemistry of *Cannabis*. J. and A. Churchill, London, pp. 11–38.
Schultes R.E. 1973. Man and marijuana. Nat. Hist. 82: 58–63, 80, 82.
Schultes R.E., Klein W.M., Plowman T. and Lockwood T.E. 1974. *Cannabis*: an example of taxonomic neglect. Harvard Univ. Bot. Mus. Leafl. 23: 337–367.
Serebriakova T.Ya. and Sizov I.A. 1940. Cannabinaceae Lindl. In: Vavilov N.I. (ed.), Kulturnaja Flora SSSR, vol. 5, Moscow-Leningrad, USSR, pp. 1–53.
Shields C.R., Orton T.J. and Stuber C.W. 1983. An outline of general resource needs and procedures for the electrophoretic separation of active enzymes from plant tissue. In: Tanksley S.D. and Orton T.J. (eds), Isozymes in Plant

Genetics and Breeding. Part A, Elsevier Science Publishers, Amsterdam, pp. 443–516.
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*, vol. 1, Corpus Information Services, Toronto.
Small E. and Beckstead H.D. 1973. Common cannabinoid phenotypes in 350 stocks of *Cannabis*. Lloydia 36: 144–165.
Small E. and Cronquist A. 1976. A practical and natural taxonomy for *Cannabis*. Taxon 25: 405–435.
Small E., Jui P.Y. and Lefkovitch L.P. 1976. A numerical taxonomic analysis of *Cannabis* with special reference to species delimitation. Syst. Bot. 1: 67–84.
Soltis D.E., Haufler C.H., Darrow D.C. and Gastony G.J. 1983. Starch gel electrophoresis of ferns: A compilation of grinding buffers, gel and electrode buffers, and staining schedules. Am. Fern J. 73: 9–27.
Vavilov N.I. 1926. The origin of the cultivation of 'primary' crops, in particular cultivated hemp. In: Studies on the Origin of Cultivated Plants, Institute of Applied Botany and Plant Breeding, Leningrad, USSR, pp. 221–233.
Vavilov N.I. and Bukinich D.D. 1929. Konopli. Zemledel'cheskii Afghanistan. Tr. Prikl. Bot. Genet. Sel. 33(Suppl.): 380–382, 474, 480, 584–585, 604 Reissued in 1959 by Izdatel'stuo Akademii Nauk SSSR, Moskva-Leningrad.
Weeden N.F. and Wendel J.F. 1989. Genetics of plant isozymes. In: Soltis D.E. and Soltis P.S. (eds), Isozymes in Plant Biology, Dioscorides Press, Portland, OR, pp. 46–72.
Wendel J.F. and Weeden N.F. 1989. Visualization and interpretation of plant isozymes. In: Soltis D.E. and Soltis P.S. (eds), Isozymes in Plant Biology, Dioscorides Press, Portland, OR, pp. 5–45.
Wiley E.O. 1981. Phylogenetics. John Wiley and Sons, New York.
Zeven A.C. and Zhukovsky P.M. 1975. Cannabidaceae [sic]. In: Dictionary of Cultivated Plants and their Centres of Diversity, Centre for Agricultural Publishing and Documentation, Wageningen, The Netherlands, pp. 62–63, 129–130.
Zhukovsky P.M. 1964. Hemp (*Cannabis* L.) In: Cultivated Plants and their Wild Relatives, 2nd edn., Ch. 9. Publishing House Molos, Leningrad, USSR, pp. 456–458.