A Numerical Taxonomic Analysis of *Cannabis* with Special Reference to Species Delimitation[1]

ERNEST SMALL,[2] PERRY Y. JUI,[3] AND L. P. LEFKOVITCH [3]

Abstract. *Two thousand five hundred plants, representing 232 diverse populations of* Cannabis, *were grown under standard conditions in a garden, scored for 47 attributes, and the data used in a numerical taxonomic study of variation. Groups of interest included "nonintoxicant" and "semi-intoxicant" populations (collectively referable to C.* sativa), *"intoxicant" populations (sometimes called C.* indica), *fiber and oil cultivars (referable to C.* sativa), *"wild" populations (sometimes called C.* ruderalis), *and plants either containing or not containing cannabigerol monomethyl ether. Clustering methodology revealed only a limited tendency for the populations to separate into the above groupings. However, canonical analysis (equally weighted multiple discriminant analysis) of morphological characteristics only proved highly successful in delineating the groups. The analysis resulting from the comparison of wild and cultivated populations when applied to a large sample of populations failed to suggest two discrete groupings, and it is consequently concluded that wild and cultivated populations intergrade so greatly as to preclude recognition of wild plants as a separate species (the so-called C.* ruderalis). *Those morphological characteristics that successfully distinguish intoxicant populations from other populations in material raised under standardized garden conditions were sufficiently variable to suggest that the intoxicant potential of plants collected in nature cannot be reliably distinguished by morphology; consequently it is judged that there are no grounds for distinguishing intoxicant plants (the so-called C.* indica) *as a separate species. It is concluded that all plants of* Cannabis *are assignable to one species, C.* sativa.

The taxonomic treatment of *Cannabis* at the species level has recently become a contentious issue in North America. Most plant taxonomists have considered the genus *Cannabis* to comprise one species, C. *sativa* Linnaeus, and accordingly most North American legislation governing the proscriptions against marijuana and hashish -defines the controlled material as C. *sativa.* Recently some botanists have argued in courts on behalf of defendants charged with narcotic-related crimes that two "legal'' species of *Cannabis, C. indica* Lamarck and C. *ruderalis* Janischevsky, deserve recognition.

[1] We thank J. McNeill, B. Baum, B. K. Thompson, and M. R. Binns for their helpful criticism of the manuscript.

[2] Biosystematics Research Institute, Research Branch, Agriculture Canada, Ottawa, Ontario, Canada K1A OC6.

[3] Statistical Research Service, Agriculture Canada, Ottawa, Ontario, Canada K1A 0C6.

The present viewpoint of the most prominent of these botanists (Schultes et al., 1974; Emboden, 1974) contrasts with their previously published categorical statements that ***Cannabis*** is monotypic (Schultes, 1970; Emboden, 1972). The resulting confusion as to whether the name C. ***sativa*** includes all samples of marijuana has become the subject of an ongoing forensic debate that has been discussed elsewhere (Small, 1974, 1975 ***b,c,d, 1976).***

The present study represents part of a broad systematic survey of variation in ***Cannabis.*** Previous work (Small, 1972a) indicated that diverse populations from widespread geographical locations appear to be completely interfertile, so that this criterion provides no basis for taxonomic segregation. By contrast, studies of variation of the psychoactive constituents (Small 1972b; Small & Beckstead 1973*a,b*; Small et al., 1975) indicated the existence of chemical groupings within ***Cannabis,*** whereas a study of herbarium material revealed that characteristics of the achenes usually differ somewhat between wild and cultivated plants (Small, ***1975a).*** The considerations upon which taxa have been recognized in ***Cannabis*** make it imperative to clarify the relationships among "wild" plants, plants selected for fiber and oil properties, and plants selected for drug properties. This paper is concerned with the analysis, by numerical techniques, of the characteristics and taxonomic structure of a large and heterogeneous sample of populations of ***Cannabis*** cultivated under' standard conditions. Particular attention is devoted to the extent to which it is possible to distinguish morphologically wild from domesticated plants and plants of different intoxicant potential, because these groupings reflect the historical rationale for the recognition of species in ***Cannabis.***

Only the names C. ***sativa, C. indica,*** and C. *ruderalis* deserve consideration as possibly reflecting the existence of more than one species in ***Cannabis,*** but the following specific epithets have been validly published in the genus:

C. sativa Linnaeus, Sp. Pl. 1027, 1753.
C. ***indica*** Lamarck, Method. Bot. 1: 695. 1785. (Dated 1783 in publication, but Breistroffer (1948) determined that the relevant pages belong to Part 2 published in 1785.)
C. *macrosperma* Stokes, Bot. Mater. Med. 4: 539. 1812.
C. ***chinensis*** Delile, Ann. Sci. Nat. Bot. 12: 366. 1849.
C. *gigan tea* Vilmorin, Rev. Hort. 5: 109. 1851.
C. ***ruderalis*** Janischevsky, Učen. Zap. Saratovsk. Gosud. Cernyševskogo Univ. 2 (2): 3–17. 1924.
×*C. intersita* Soják, Novit, Bot. Délect. Seminum Horti. Bot. Univ. Carol. Prag. 1960: 20. 1960. (= C. ***ruderalis*** × ***C. sativa***).

Lamarck described C. ***indica*** in 1785 from fragmentary material. One of the characteristics that he used to distinguish it from C. ***sativa,*** the alleged alternate leaf arrangement in C. ***indica*** in contrast with the opposite arrangement in C. *sativa,* is without merit since all plants of ***Cannabis*** tend to have opposite lower leaves and alternate upper leaves. Other characteristics, for example the smaller size of C. ***indica*** coupled with more marked branching, "harder" stems, and relatively narrow leaflets led to a qualified acceptance

of such an entity (usually as a variety rather than a species); for the most part, however, botanists did not accept C. ***indica*** as a species. Considerable confusion resulted when the binomial C. ***indica*** was adopted by pharmacologists to designate officinal drug preparations from C. ***sativa.*** Although various botanists have altered or expanded Lamarck's concept of C. ***indica,*** no acceptable treatment has yet been produced.

Lamarck made specific mention of the inebriating potential of C. ***indica*** and contrasted this species with C. ***satiua,*** which he described as having much better qualities for fiber, for which it was cultivated, in contrast to C. ***indica*** cultivated for its drug content. Recently Stearn (1974) lectotypified C. ***satiua,*** interpreting Linnaeus' concept as based on European fiber stocks and designating a lectotype from the material studied by Linnaeus that represents such a plant. Chemical studies (Small & Beckstead, 1973*a,b*) indicate that fiber stocks are of limited intoxicant potential, as are most plants of ***Cannabis*** in Europe, northern Asia, and North America north of Mexico. These plants are here designated "nonintoxicant" when females yielded marijuana with less than 0.1% content of the psychoactive constituent, tetrahydrocannabinol (THC), and "semi-intoxicant" when females, but not males, yielded marijuana of higher THC content. Such plants contrast markedly with many originating from farther south (Mexico and southwards, Africa, India, southern Asia), which we term "intoxicant." In comparison with the less intoxicant categories (nonintoxicant and semi-intoxicant), intoxicant plants have large amounts of THC, large amounts of resin in the male plants (as well as in the females), and a requirement for a long growing season for sexual differentiation. The lectotypic specimen of C. ***indica*** is assignable to the intoxicant group. Accordingly, it would seem appropriate to recognize the less intoxicant plants and the intoxicant plants as *C.* ***sativa*** and C. ***indica*** respectively if sufficient morphological differentiation is found.

During our chemical studies we discovered that plants from northeastern Asia (for the most part) possessed trace amounts of a nonintoxicant "cannabinoid" (the cannabinoids are the class of terpenoid chemicals to which THC is assigned), cannabigerol monomethyl ether (CBGM), whereas no other populations contained detectable CBGM. Although only seven populations of this chemical race were available, we also assessed the comparative morphology of this group.

Janischevsky described ***C.*** *ruderalis* as a new species in 1924. His C. *ruderalis* amounted to plants growing spontaneously or indigenously in southeast-central Russia as opposed to cultivated (domesticated) plants to which the name C. ***sativa*** was restricted. He distinguished the two taxa on the basis of achene characteristics that can be interpreted as having arisen in the wild as a result of selection (or conversely, as having arisen in cultivation as a result of relaxed selection). Wild plants exhibited achenes with morphological features favoring I) dissemination: attenuated bases, sharp basal abscission zone, and possibly some proliferation of oily tissue attractive

to the bug *Pyrrhocoris apterus* Linnaeus; 2) camouflage: a layer covering the achene, homologous with the perianth, bearing pigmented areas resulting in a marbled appearance; and 3) ensured propagule formation when dwarfed by the frequently inhospitable habitat of this weedy plant: small achene size. Janischevsky noted that the utility of his treatment could only be decided by more comprehensive geographical sampling than he had conducted.

Our use of the term "wild" requires qualification. We recognize the distinctions between native or indigenous plants on the one hand and escaped or spontaneous plants on the other. It has been speculated that ***Cannabis*** is native to central Asia (references is Schultes, 1970; Small et al., 1975), and it is clear that it has been introduced into the New World and elsewhere (both for fiber and drug purposes) where it escapes and forms reproducing populations that can be termed naturalized. Naturalized populations may also result from a range extension of indigenous populations. Compounding the situation, cultivars escape continually and form "spontaneous" populations, which gradually evolve characteristics adapted to existence in the wild, and populations in the wild are continually taken up, domesticated, and selected for particular characteristics. Further complicating the situation has been the evolution of weedy biotypes, perhaps the result of interaction between the wild and domesticated phases. *Cannabis* is one of the oldest of domesticated plants, and genetic exchanges between the wild and cultivated phases have occurred for perhaps 8,500 years (Schultes & Hofmann, 1973). It is unlikely that there are populations of ***Cannabis*** extant anywhere, including the putative central Asian center of origin, that have escaped very substantial introgression from the effects of cultivation. We perceive no noncircular and reliable method of distinguishing plants found in nature as spontaneous, naturalized, or indigenous (at least in Asia). (Compare Anderson's (1952) discussion of the problem of identifying "wild" apple trees.) We have labelled a number of populations in our studies as "wild" simply on the basis that they were collected in nature, apparently not originating from plantation activities.

MATERIALS AND METHODS

Two hundred thirty-two populations of ***Cannabis*** established in a plantation described by Small (1972*b*), and Small and Beckstead (1973a) were studied. Seed stocks were obtained from a wide variety of sources (Appendix 1). Male plants were collected at anthesis, females at initial fruit maturation, generally several weeks after male anthesis. A few populations, particularly those of the "intoxicant" group, failed to flower before frost so that only vegetative plants were available. Several populations produced mostly monoecious plants. At least ten plants of each population were collected, usually five males and five females. Parallel studies evaluated the chemistry of the psychoactive constituents (Small & Beckstead 1973*a,b*),

fiber content of stems (after Bredemann, 1922), and oil content of achenes (after Appelqvist, 1967).

For each specimen fresh weight and height were measured in the field. A section of stem from the middle of the plant, a leaf from the middle of the plant, and the central inflorescence were preserved. The root was discarded after study because of its bulk and dearth of characters. Achenes of the original acquisitions were studied and are preserved at DAO. Variation in achene morphology is treated more intensively elsewhere (Small 1975a). Specimens are deposited at DAO, BM, NY, USF, and us.

We were rather surprised but somewhat encouraged (because of the experience of others studying different sexes or life-history stages by numerical methods) that the separate cluster analyses of the sexes showed little discordance. Since each analysis is based on fewer characters than the combined set, we decided to average the data for each population prior to analysis and to use these means to represent the OTU's. We recognize the conceptual difficulties of averaging male and female organisms but consider this solution sufficient for our purpose of assessing overall taxonomic structure. Furthermore, the means were based on different numbers for each OTU, depending on the number of different specimens originally measured, and so each is a better or worse estimate of the true value. However, the numbers were approximately the same, and small, so they are comparable. Two thousand five hundred plants were analysed in terms of 232 OTU's. The 47 attributes examined are listed in Appendix 2.

Clustering was carried out using a dissimilarity coefficient based on the Gower (1971) similarity coefficient. Although some OTU's were not scored for all attributes because of limited material or failure of material to mature, no triangular inequalities were found after checking all possible distance triangles, where distance was equated with dissimilarity, and hence the metric space was Euclidean. All 47 attributes were used in clustering procedures. Clustering algorithms included nearest neighbor, furthest neighbor, unweighted centroid, weighted centroid (median), average linkage (unweighted pair-group method using arithmetic averages), incremental sum of squares, and flexible sort analysis.

In addition, to clustering methodology, a Q-technique canonical variate analysis was attempted in order to examine the extent to which several groups of interest could be separated. (By canonical analysis we mean a multiple discriminant analysis in which the prior probabilities of the groups are equal and not proportional to the sample sizes.) The groups were defined on the basis of intoxicant potential, presence or absence of CBGM, and whether wild, cultivated, or not reliably classifiable as either. Eighteen groups were represented by different combinations of the defining characteristics (Table 1). Separate canonical analyses were conducted first among the three groups defined on intoxicant potential (24 intoxica-nt populations, 48 semi-intoxicant populations, 160 nonintoxicant populations), second between a group of seven populations of the CBGM-containing race

TABLE 1. Numbers of populations of each kind recognized by all criteria examined in the present study. CBGM = Cannabigerol monomethyl ether.

	Wild	Cultivated	Unclassified	Totals
With CBGM				
Nonintoxicant	0	1	0	1
Semi-intoxicant	0	3	2	5
Intoxicant	0	0	1	1
Totals	0	4	3	7
Without CBGM				
Nonintoxicant	11	77	71	159
Semi-intoxicant	1	7	35	43
Intoxicant	5	0	18	23
Totals	17	84	124	225

and another of all remaining populations, and third between a group of 17 populations collected in the "wild" and another of some 88 cultivar populations. The remaining 127 populations, not reliably classified as wild or cultivated, were considered in the framework of the wild-cultivated canonical variate axis to examine to what extent the considerable morphological discrimination found. between the selected samples of wild and cultivated plants could be used to distinguish two such groups.

Canonical variate analysis is limited in the number of characters that can be simultaneously evaluated to one fewer than the smallest sample size of the classes examined. Because the purpose of our canonical analyses was to examine the extent of morphological differentiation, only the morphological characters listed in Appendix 2 were used. Furthermore, only the characters for which all populations were scored were chosen, thus eliminating floral characters because some populations failed to flower. The use of t-tests for the continuously scored characters and χ^2 values derived from contingency tables for a number of multiple-state characters further provided a crude indication of the potential value of the characteristics. Multiple-state-ordered characters usually cannot be assessed by a t-test, but if the means of such variables are considered to be normally distributed, then a t-test applied to them is perhaps effective. Characters 13, 14, 21, and 22 were of this kind, and the differences between groups were examined by t-tests. On the other hand, because the multiple-state characters 6, 25, 26, 27, 28, 29, 45, 46, and 47 were represented by single values for each OTU, the χ^2 test from a contingency table was used for these variables. The final choice of characters for the canonical analyses, made by employing a forward stepwise procedure on the remaining characters, is given in Table 3.

RESULTS

Cluster analysis tended to associate populations defined by intoxicant potential, content of CBGM, and wild or cultivated origin. The seven methods adopted produced similar results. The dendrogram for the nearest-

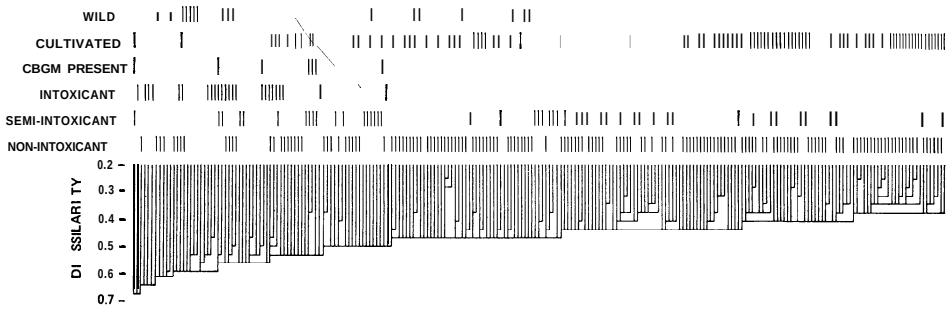


FIG. 1. Nearest-neighbor dendrogram of 232 populations of *Cannabis,* showing relationships of the six groupings examined in this studv.

neighbor method, which seemed to allocate the populations so that they coincide closest to the previously defined groupings, is presented in Fig. 1. There is no evident discrete structuring of the populations but rather a continuum of dissimilarity exists. The most similar populations are cultigens that are nonintoxicant (Fig. 1) whereas 74 of the 232 populations show a dissimilarity of 0.5 or greater with the remaining ones. The dendrogram has the general form for groups, which, if they exist at all, are poorly differentiated.

The results of t-tests or of χ^2 tests for the attributes for the two comparisons of principal interest, firstly wild US. cultivated, and secondly intoxicant US. less intoxicant (semi-intoxicant and nonintoxicant) populations, are given in Table 2. In both cases, more than half the characters differ significantly.

The order of importance of characters from the subsets of characters chosen for use in the canonical analysis for each of the five analyses conducted is given in Table 3. It should be borne in mind that this order will not necessarily be consistent with that suggested by single character t-tests or χ^2 tests. In contrast to the clustering methodology adopted, canonical analysis achieved high levels of description.

The two canonical variates for the three groups of populations based on intoxicant potential are shown in Fig. 2. While the 24 intoxicant populations are fairly well separated, the 160 nonintoxicant populations overlap with the 48 semi-intoxicant ones. The generalized (Mahanolobis) distances among the centroids of the three groups, based on the 23 morphological characters used (Fig. 3), further shows that the nonintoxicant populations are very similar to the semi-intoxicant ones but that both (arbitrarily distinguished) groups are well separated from the intoxicant populations.

On an intuitive basis one might expect that semi-intoxicant plants represent introgression of nonintoxicant plants with intoxicant plants and perhaps should be morphologically intermediate between these. This is not obviously the case as is shown in Fig. 3 where the centroid of the semi-

TABLE 2. Means of character sample means and values of t or χ^2 for comparisons between wild and cultivated populations and intoxicant and less intoxicant populations (see Table 1.) * = significant at $P = 0.05$; n as in column heading unless otherwise specified.

Character	Wild ($n = 17$)	Cultivated ($n = 88$)	t	χ^2	Intoxicant ($n = 24$)	Less Intoxicant ($n = 208$)	t	χ^2
1	180.0	197.9	-1.78	—	189.9	182.4	0.90	—
2	992.1	841.2	1.20	—	1039.0	788.0	2.80*	—
3	2.08	2.11	-0.14	—	2.24	2.00	2.08*	—
4	37.9	42.6	-1.18	—	10.0	44.0	–13.4*	—
5	11.3	11.2	0.18	—	16.8	10.8	13.48*	—
6	0.00	0.36	—	9.08*	0.08	0.18	—	1.56
7	8.23	8.34	-0.47	—	8.47	7.87	2.44*	—
8	13.0	17.3	–7.58*	—	14.2	16.3	–4.33*	—
9	1.30	1.99	–8.59*	—	1.44	1.87	–5.51*	—
10	9.45	11.23	–3.81*	—	8.90	10.54	–3.51*	—
11	1.75	2.06	-3.47 *	—	2.06	2.02	0.61	—
12	14.4	15.3	-2.22 *	—	15.1	15.0	0.35	—
13	2.27	2.49	–2.53*	—	2.30	2.36	-0.68	—
14	1.60	1.57	0.49	—	1.53	1.62	-1.58	—
15	0.14	0.12	2.70*	—	0.11	0.13	–2.58*	—
16	0.72	0.98	-1.78	—	0.74	0.89	-1.77	—
17	2.68	3.29	–4.12*	—	2.38	3.15	-5.98 *	—
18	7.75	10.64	–9.36*	—	8.23	10.16	–6.17*	—
19	20.1	18.8	1.05	—	29.3	19.4	9.47*	—
20	13.2	15.2	–3.84*	—	13.0	14.5	–2.81*	—
21	1.57	1.40	2.31*	—	1.48	1.59	-1.25	—
22	1.99	1.32	6.14*	—	1.58	1.59	-0.06	—
23	1.15	1.41	–3.47*	—	1.28	1.29	-0.18	—
24	0.38	0.68	-6.09 *	—	0.50	0.55	-0.94	—
25	2.88	4.97	—	31.54*	3.88	4.35	—	2.63
26	2.53	3.45	—	16.52*	3.75	3.41	—	4.97
27	1.94	2.44	—	5.66	3.50	2.71	—	7.46
28	2.41	2.31	—	0.43	2.13	2.19	—	6.07*
29	1.82	2.63	—	23.15*	2.29	2.39	—	9.95
30	9.76	19.69	–11.59*	—	11.12	17.50	–6.92*	—
31	5.79	6.56	–3.81*	—	5.55 ($n = 2$)	6.37	-1.71	—
32	51.0	57.5	-1.94	—	58.3 ($n = 2$)	58.4	-0.03	—
33	3.95	3.12	3.65*	—	4.25 (n = 2)	3.21	1.79	—
34	64.3	63.5	0.53	—	72.5 ($n = 2$)	62.8	2.62*	—
35	5.44	6.17	-1.72	—	7.69 ($n = 12$)	5.54	4.26*	—
36	3.74	4.12	–3.41*	—	3.73 (n = 12)	4.04	–2.09*	—
37	1.73	1.70	0.37	—	1.71 ($n = 12$)	1.70	0.14	—
38	40.2	38.2	1.38	—	42.2 ($n = 12$)	38.7	2.42*	—
39	3.20	3.47	–3.10*	—	3.15 ($n = 12$)	3.39	–2.08*	—
40	0.83	0.84	-0.22	—	0.76 ($n = 12$)	0.85	-1.28	—
41	2.29	2.01	1.11	—	3.56 ($n = 13$)	2.14	4.06*	—
42	26.8($n = 5$)	33.3	-4.25 *	—	27.5 ($n = 13$)	32.9	–5.24*	—
43	9.98	14.10	–4.85*	—	11.07	11.96	1.30	—
44	1.20	1.22	-0.26	—	0.98 ($n = 4$)	1.25	–2.10*	—
45	0.00	0.06	—	1.01	0.04	0.03	—	0.82
46	0.24	0.15	—	1.42	0.92	0.10	—	26.71*
47	0.25	0.06	—	6.03*	0.93 ($n = 14$)	0.05		35.97 *

TABLE 3. Order of importance of variables as evaluated by forward stepwise procedure. The values in the body of the table give the ranking of importance after adjusting for those previously selected.

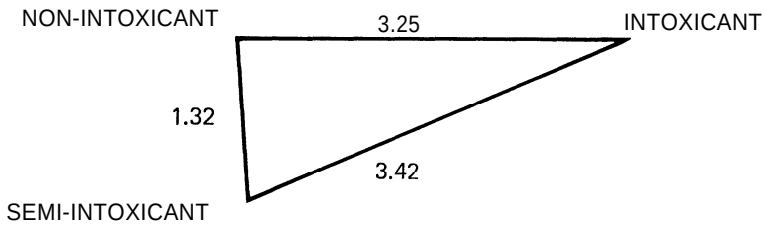
Character	Wild *vs.* cultivated	Intoxicant *vs.* semi-intoxicant *vs.* nonintoxicant	Intoxicant *vs.* semi-intoxicant + nonintoxicant	Semi-intoxicant *vs.* nonintoxicant	CBGM present *vs.* absent
1	—	10	10	5	—
2	—	11	18	6	—
3	—	12	19	7	—
6	3	—	—	—	—
7	—	6	3	20	—
8	10	23	20	21	—
9	4	2	2	4	3
10	6	8	5	12	—
11	5	5	4	9	—
12	—	9	9	2	1
13	—	16	11	17	4
14	—	22	23	22	—
15	—	19	12	13	5
16	—	21	13	23	—
17	9	20	16	18	—
18	8	4	17	3	—
19	—	1	1	14	2
20	14	7	8	15	—
21	13	3	6	1	—
22	7	—	—	—	6
23	16	—	—	—	—
24	15	17	22	11	—
25	2	14	15	8	—
26	12	18	14	16	—
27	—	13	7	19	—
29	11	15	21	10	—
30	1	—	—	—	—

intoxicant plants appears to resemble that of the intoxicant plants even less than that of the nonintoxicant plants.

In an attempt to examine nonintoxicant and semi-intoxicant populations in the absence of the intoxicant ones, we compared these by canonical analysis alone. A discrimination of 75% was achieved (52 of 208 populations were misclassified). Although no notable differences were found by considering only these two groups, there were significantly larger values in nonintoxicant populations for characters 6, 18, and 43 and smaller values for characters 2, 12, 19, 21, 45, and 46.

Because of the relative similarity of nonintoxicant and semi-intoxicant populations, these were grouped for further comparison with the intoxicant populations. The frequency histogram on the canonical axis (Fig. 4) reveals that the two groupings can be separated with a remarkable level of correct classification of 94% (14 of 232 populations misclassified).

It should be stressed that the number of populations containing CBGM



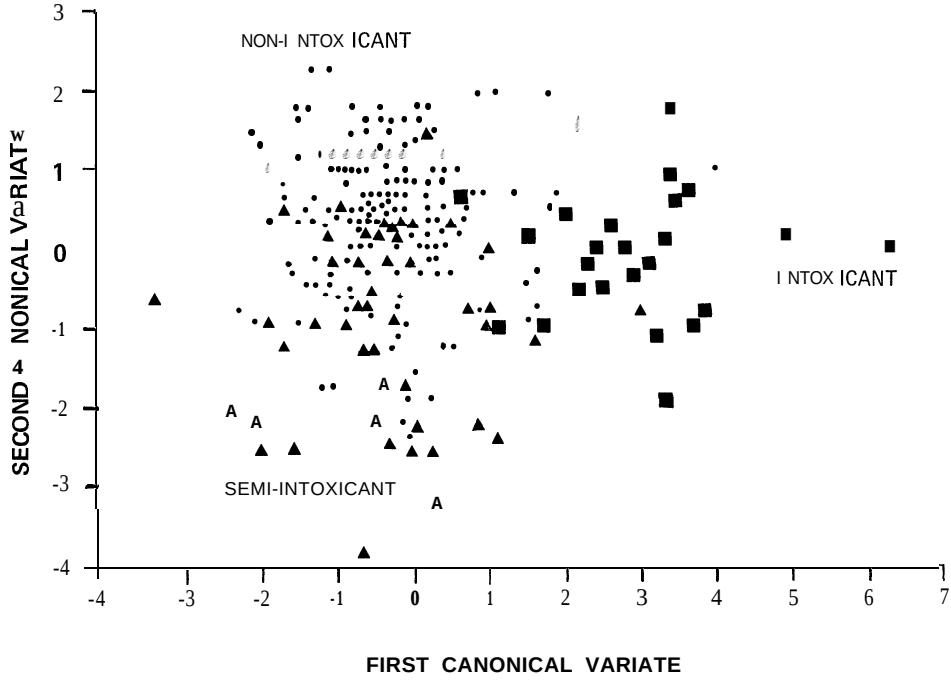


FIG. 2-3. — 2. *(Below.)* Canonical analysis of nonintoxicant, semi-intoxicant, and intoxicant populations (indicated by dots, triangles, and squares, respectively), showing relationships with respect to the first two canonical variates. — 3. (*Above.*) Generalized morphological distances of the intoxicant groups based on mean of means of 23 morphological features examined in canonical analysis.

was small ($n = 7$), especially in comparison with the number of populations not containing detectable CBGM ($n = 225$), although a very high level of discrimination was achieved by canonical variate analysis (91 % correct classification, 21 of 232 populations misclassified). This is less surprising because of the relative group sizes. The following differences were significant: populations with detectable CBGM had smaller values for characters

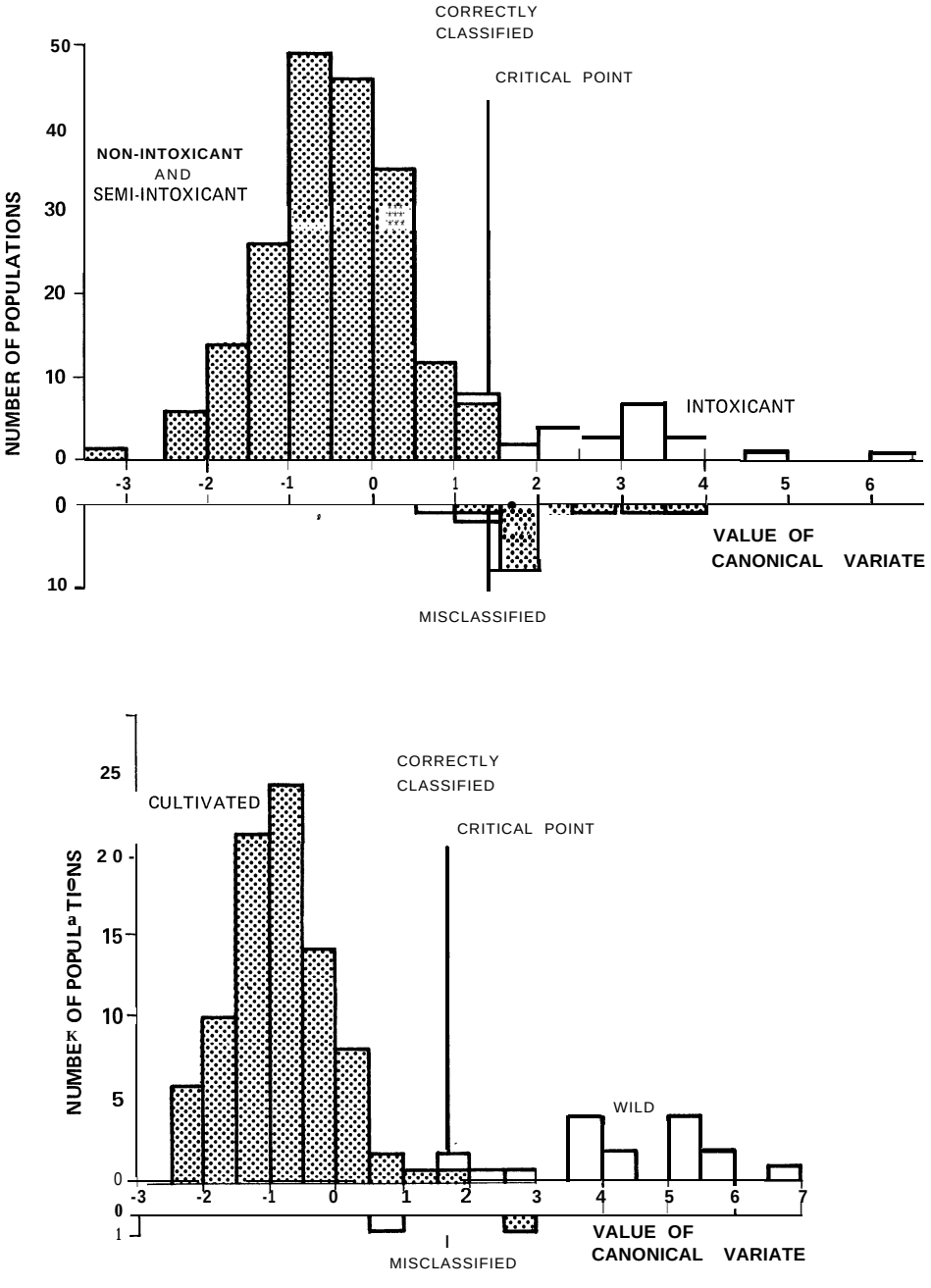


FIG. 4-5. — 4. (Above.) Canonical analysis of intoxicant and less intoxicant (non-intoxicant + semi-intoxicant) populations for the canonical axis, showing histograms of numbers of populations correctly and incorrectly classified by the analysis. — 5. ***(Below.)*** Canonical analysis of wild and cultivar populations for the canonical axis, showing histograms of numbers of populations correctly and incorrectly classified by the analysis.

15, 22, 23, and 43 and larger values for characters 2, 9, 12, 13, 19, 24, 32, 35, 37, 41, 45, 46, and 47.

A frequency histogram for 88 reliably identified cultivar populations and 17 reliably identified wild populations is projected on the canonical axis in Fig. 5. Only one misclassification in each group was obtained, achieving 98.1 % correct classification. However, the remaining 127 populations, which were less reliably interpretable as wild or cultivated, when scored on the same canonical axis failed to show two discrete groups on different sides of the critical point (Fig. 6).

DISCUSSION

Cluster analysis failed to suggest that there are groupings of interest sufficiently separated to warrant taxonomic recognition. Insofar as such methodology represents an "objective" approach, weighting all characteristics equally without assuming any groupings, it may be stated that little support was found for the recognition of taxa in ***Cannabis***. Canonical analysis, in contrast, assumes ***a priori*** that groups exist and allows ***a posteriori*** judgments to be made on them, a defensible scientific procedure. It is of interest that Sneath and Sokal (1973) recommend a combination of "classification" (particularly clustering) and ordination techniques in numerical-taxonomic studies. Our ordination approach (by canonical analysis) which has been to verify structure within ***Cannabis*** postulated by other botanists, has proved particularly useful in this study. However, a number of considerations require evaluation in attempting to translate the inferences obtained from the substantial levels of discrimination achieved into taxonomic conclusions.

It should be stressed that prior chemical studies made it clear that several basic chemical groups, with geographical integrity, existed. The two widespread chemical groups—the northern, little-intoxicating, early-maturing category with little resin in the male plants, and the southern, highly intoxicating, late-maturing category with high resin content in male plants—constituted logical candidates for specific recognition, corresponding to the two putative species C. *sativa* and C. *indica*. It is of interest that two groupings within the less intoxicant category, the nonintoxicants and the semi-intoxicants, can be distinguished in the material studied with 75% accuracy on strictly morphological grounds and likewise the finding that plants with trace amounts of CBGM and usually originating from northeastern Asia could also be distinguished. These findings, however, are of peripheral interest to the discovery that there might be morphological grounds for distinguishing plants with high intoxicant potential from those with limited intoxicant potential.

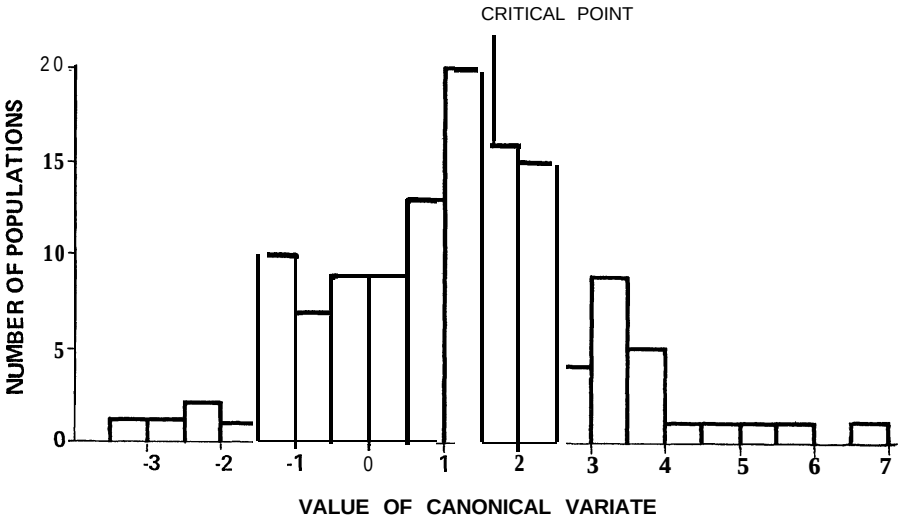
Despite the considerable level of discrimination between intoxicant plants and the less intoxicant plants—94%—it has become evident that a practical combination of morphological features that can consistently distinguish the two groupings in nature, and thereby serve to characterize two species, has yet to be discovered. Two considerations govern this impracticality. First,

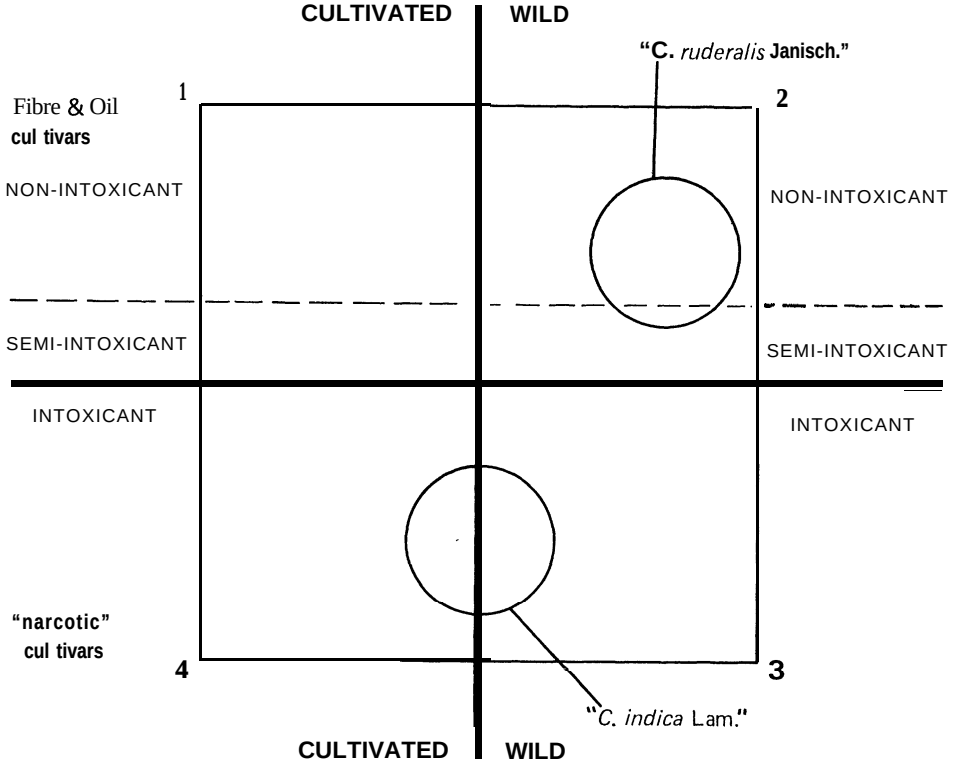
the characters are too plastic for consistent identification of material in nature. The most obvious distinguishing features in intoxicant plants, higher secretory gland density and smaller leaflet dimensions, are subject to great environmental modification like many morphological characteristics of ***Cannabis.*** (Achene characteristics are comparatively conservative; note Small, 1975*a*). It is evident that the considerable morphological distinction that we are able to perceive arose from the fact that all plants had approximately identical growth conditions in our standard garden. Specimens collected in nature are much more variable as a result of environmental modification and cannot be identified as to intoxicant potential on the morphological grounds successfully seen in our garden material. We are also suspicious that the late maturation of intoxicant plants in our garden may have caused consistent differences in morphology and that this may have contributed to the substantial morphological distinction perceived. Second, the high level of discrimination was achieved not only by standardizing growth conditions but also by measuring a large number of characteristics (23) none of which alone or even in small numbers possessed much discrimination. We doubt that even the most committed practitioner of numerical taxonomy, let alone the majority of plant taxonomists, would advocate the use of such a cumbersome mathematical function to discriminate species. Given the greater variability of plants in nature-not in terms of a single locality but over the whole range-it would be most unlikely that a similar study of plants in nature could provide a less complex combination to be useful for taxonomic diagnosis.

In terms of the relative selective forces for fiber and oil in the less intoxicant category of plants, many of which are oil and fiber cultivars, it is not surprising that such plants have a higher content of fiber in the stems or a higher content of oil in the achenes than those of intoxicant strains. The higher secretory gland density in intoxicant plants appears to be an obvious concomitant of selection for drug content. Several of the characters variously ascribed to C. *indica,* such as smaller size, smaller and narrower leaflets, greater branching, and smaller achenes, are seen to have some, but only limited, validity as descriptions of the intoxicant group of plants.

For the canonical analyses of the combinations of groups defined on intoxicant potential, all available populations were evaluated at one time. In the comparison of wild and cultivated plants, selected samples of the two groups of interest were first examined and the resulting allocation rule was tested by applying it to the remaining plants. The latter has the disadvantage of being based on a smaller sample but the advantage of providing a test of external predictability. The latter procedure was more appropriate for the comparison of wild and cultivated plants because we could not confidently assign all plants to one or the other group.

The comparison of 1'7 wild populations and 88 cultivars revealed a high level of discrimination (98%). Agreeing with Janischevsky (1924)


CRITICAL POINT
NUMBER OF POPULATIONS
20
15
10
5
0
-3
-2
-1
0
1
2
3
4
5
6
7
VALUE OF CANONICAL VARIATE


CULTIVATED
WILD
"C. ruderalis Janisch."
1
2
Fibre & Oil
cul tivars
NON-INTOXICANT
NON-INTOXICANT
SEMI-INTOXICANT
SEMI-INTOXICANT
INTOXICANT
INTOXICANT
"narcotic"
cul tivars
4
3
"C. indica Lam."
CULTIVATED
WILD

in his limited studies, we found that in general our wild plants contrasted with cultivars in having small achenes with attenuated bases and a sharply defined abscission zone and a marbled covering perianth. Additional features characterized wild plants: smaller leaflet dimensions, slimmer stems with a more restricted hollow in the internodes, and greater branching. Relative lack of branching, hollower stems, and greater fiber content of the stems obviously reflect artificial selection. Not unexpectedly, oil content was higher in the achenes of the cultivars, and several cultivars were monoecious. (Selection for monoecious fiber cultivars overcomes the inconvenient differential maturation times of the sexes.)

When the efficacy of the morphological discrimination analysis was tested (Fig. 6) it became apparent that there are numerous intermediate plants and that two groupings cannot be recognized when the total variation pattern is considered. This conclusion was reached by Vavilov (1926), who noted that in the USSR populations of ***Cannabis*** exhibit all degrees of intermediacy between Janischevsky's conceptions of wild and domesticated plants.

A major finding of the present study is that there is insufficient evidence on which one could recognize more than one species in ***Cannabis.*** It has been shown that there are insufficient morphological grounds for distinguishing so-called wild from domesticated plants consistently. While intoxicant plants, in contrast to less intoxicant plants, can be distinguished by a numerical analysis of many characters of plants grown under standard circumstances, the much greater variability of plants collected in nature militates strongly against the practical identification of the intoxicant potential of plants of ***Cannabis*** on morphological grounds alone.

Cannabis is an extremely widespread organism in both the cultivated and wild phases and various races repeatedly have been introduced into cultivation and have escaped to form spontaneously spreading populations. It is not suprising, therefore, that the variation pattern is substantially continuous, at least insofar as exomorphic features are concerned. It should also be stressed that artificial selection by man for fiber, oil, and drug has provided the stimulus for the evolution of races with characteristics that led to taxonomic segregation at the species level in the past and that such artificially selected attributes provide a questionable basis for the recognition of species.

Nevertheless, four comprehensive groups can be recognized (Fig. 7) based

←

FIG. 6–7. — 6. (*Above.*) Histograms of scores of populations not included in analysis of Fig. 5. Populations scored along same canonical variate axis as in Fig. 5. Absence of biomodal distribution indicates absence of two discrete groups. (See the text.) — 7. *(Below.)* An interpretation of variation in *C. sativa* with respect to the variation axes wild-cultivated and nonintoxicant-intoxicant. Four basic biological groups (quadrats 1–4) are indicated. The positions of *C. ruderalis* and *C. indica,* as circumscribed by their authors, are shown.

on the four possible combinations., of limited and pronounced intoxicant potential and on whether the plants are domesticated or wild, The question of appropriate taxonomic rank for these groupings is of interest. In a forthcoming publication these will be recognized as varieties and a full examination of the advisability of delimiting taxa in ***Cannabis*** will be presented (Small & Cronquist, 1976). The present investigation leads us to conclude that there is little merit, given the evidence available, for the recognition of more than the one species, C. *sativa*, in the genus ***Cannabis.***

APPENDTX 1

SOURCES OF THE POPULATIONS STUDIED

In the listing below A indicates that detectable cannabigerol monomethyl ether is present, C indicates a named fiber cultivar, and W indicates an authentic wild population. A detailed list of the sources and of herbaria containing vouchers is available from the senior author.

Intoxicant Populations: Cambodia: 154. Gambia: 186. India: 194 (W), 163 (W), 165 (W), 166 (W), 167 (W), 164, 26. Jamaica: 66. Malawi: 300, 301. Mauritius: 70. Mexico: 284, 41 (A). Rhodesia: 235. Sierra Leone: 63. South Africa: 11, 74, 162, 273. Thailand: 10. Uganda: 76, 77.

Semi-intoxicant Populations: Argentina: 36 (C). Bulgaria: 134 (C). Canada: 51, Chile: 80 (C). Czechoslovakia: 287 (C), 203, 191, 259. England: 32. France: 43, 206, 218, 293. Germany: 182, 283, 53 (A), 221, 37, 69. Hungary: 266 (C), 7 (C). Italy: 34, 156. Japan: 152 (A,C), 153 (A,C), 160 (A,C), 29. Korea: 170 (A). Mexico: 24. Netherlands: 190, 269, 270, 207 (C). Poland: 222, 241. Romania: 81. Spain: 209. Sweden: 230, 231, 232, 237. Turkey: 280. USSR: 264, 56, 82 (W), 60, 58. Yugoslavia: 95.

Nonintoxicant Populations: Argentina: 298 (C), 299 (C). Bulgaria: 102 (C), 103 (C), 104 (C), 105 (C), 108 (C), 113 (C), 114 (C), 138 (C), 139 (C), 140 (C), 141 (C), 142 (C), 144 (C), 109 (C). Canada: 48, 9 (W), 20 (W), 38, 30 (W), 189 (W). China: 282 (A,C). Cyprus: 42 (C). Czechoslovakia: 214, 285 (C), 286 (C), 288 (C), 4, 85, 86, 15 (C), 17, 18, 176 (C), 177 (C), 208, England: 202, 83, 75. France: 217, 211, 145, 171, 5, 27, 291, 292, 294, 295, 296, 244 (C), 245 (C), 246 (C), 247 (C), 248 (C), 249 (C), 250 (C), 251 (C), 252 (C), 253 (c), 254 (C), 255 (C), 256 (C), 257 (C), 258 (C). Germany: 216, 28 (W), 62, 150 (C), 178 (C), 179 (C), 180 (C), 181 (C), 40, 71, 210, 226, 21, 183, 219, 220. Hungary: 265 (C), 267 (C), 268 (C), 184 (W), 6 (C), 8 (C), 87 (C), 88 (C), 89 (C), 236, 237 (C), 173, 238, 175, 239, 172, 240, 174. Northern Ireland: 72. Israel: 74, 45, 46 (C). Italy: 155 (C), 157 (C), 39, 146, 147, 148. Netherlands: 224, 61, 35 (C). Poland: 242, 243 (W), 67 (C), 68 (C), 274 (C), 65 (C), 185 (C), 201 (C), 192, 212. Portugal: 271. Romania: 275 (C), 204, 168 (C), 169 (C), 195 (C), 196(C), 197 (C), 198 (C), 199 (C), 188 (C). Syria: 149. Turkey: 193, 278, 279, 19, 23, 12. USA: 79 (W), 297 (W), 25 (W). USSR: 260, 262, 263, 57, 59, 55, 158 (C), 159 (C), 213, 234, 91, 92 (C), 93 (C), 94 (C).

APPENDIX 2

LIST OF CHARACTERS

Miscellaneous: 1, height (cm); 2, fresh weight (g); 3, maximum root diameter (cm); 4, latitude of origin (degrees); 5, time elapsed from germination to male anthesis (weeks, estimated as 20 weeks for vegetative populations); 6, monoecious (1) or dioecious (0).

Leaf (midstem, middle leaflet): 7, number of leaflets; 8, leaflet length (cm); 9, leaflet width (cm); 10, petiole length (cm); 11, length of terminal, unserrated portion of leaflet

(cm); 12, number of serrations Oh one side of leaflet; 13, degree of forward serration curvature (1 = uncurved, 2 = moderately curved, 3 = strongly curved); 14, location of widest part of leaflet (1 = distal, 2 = middle, 3 = proximal); 15, leaflet thickness (mm); 16, petiole thickness (mm); 17, length of shorter (inner) part of serration (mm); 18, length of longer (outer) part of serration (mm); 19, gland density, abaxial leaflet (mm^{-2}).

Stem (at mid-stem): 20, length subtending four leaves (cm) (measures degree of branching); 21, leaf arrangement (1 = opposite, 2 = subopposite, 3 = alternate); 22, degree of ribbing (1 = very conspicuous, 2 = moderate, 3 = inconspicuous); 23, diameter (cm); 24, diameter of central hollow portion of internode (cm).

Achene: ***25,*** darkness (1 = black, 6 = light gray); 26, proportion covered by perianth (1 = much, 6 = little); ***27,*** length of longest perianth spots (1 = long, 5 = short); 28, proportion of achehes with at least trace perianth at middle (1 = small, 3 = large); 29, elongation of base (1 = much, 3 = little); 30, weight (mg).

Fruit ***Bract:*** 31, length (mm); 32, gland density, abaxial surface (mm^{-2}); 33, cystolith hair density, abaxial surface (mm^{-2}); 34, percentage of glands which are stalked.

Male Flowers: 35, length of representative bract on main stem of inflorescence (mm); 36, tepal length (mm); 37, tepal width (mm); 38, percentage of tepal lacking green color (tepal margins lack pigment); 39, anther length (mm); 40, anther width (mm); 41, pedicel length (mm).

Chemical: 42, percentage oil in achenes; 43, percentage fiber in stalks, (dry weight); 44, ratio of fiber, male/female plants; 45, CBGM present (1) or absent (0); 46, THC comprising more than 50% of cannabinoids of resin (1), or less than 50% (0); ***47,*** ratio resin content female/male plants greater than 1.4 (0), or less than 1.4 (1).

LITERATURE CITED

ANDERSON, E. 1952. Plants, Man, and Life. Boston.

APPELQVIST, L.-Å. 1967. Further studies on a multisequential method for determination of oil content in oilseeds. J. Amer. Oil Chem. Soc. 44: 209–214.

BREDEMENN, G. 1922. Über Faserausbeutebestimmungen bei Hanfzüchtung. Angew. Bot. 4: 223-233.

BREISTROFFER, M. 1948. Sur les dates exactes de parution de quelques livres botaniques publiés par D. Villars et de Lamarck. Proc. Verb. Mens. Soc. Dauph. Ethn, Arch. 24 (Nos. 182-184).

EMBODEN, W. A, 1972. Ritual use of Cannabis *sativa* L.: a historical-ethnographic survey. Pp. 214-236 *in* P. T. Furst (ed.). Flesh of the Gods. New York.

———. 1974. A botanical history of the genus ***Cannabis*** with reference to legislation. California Attorneys for Criminal Justice No. 5, Aug./Sept. Suppl. l-5.

GOWER, J. C. 1971. A general coefficient of similarity and some of its properties. Biometrics **27: 857-871.**

JANISCHEVSKY [= YANISCHEWSKII], D. E. 1924, Forma konopli na sornykh mestakh v Yugovostochnoi Rossii. Učen. Zap. Saratovsk. Gosud. Cernyševskogo Univ. 2: 3-17.

LAMARCK, J. B. DE. 1785 ("1783" in the publication). Encyclopedique Méthodique. Botanique. 1 (2): 694-695.

SCHULTES, R. E. 1970. Random thoughts and queries on the botany of ***Cannabis. Pp.*** 11-38 ***in C.*** R. B. Joyce & S. H. Curry (eds.). The Botany and Chemistry of ***Cannabis.*** London.

——— & A. HOFMANN. 1973. The Botany and Chemistry of Hallucinogens. Springfield.

———, W. M. KLEIN, T. PLOWMAN & T. E. LOCKWOOD. 1974. ***Cannabis:*** an example of taxonomic neglect. Bot. Mus. Leafl. 23: 337–367.

SMALL, E. 1972*a*. Interfertility and chromosomal uniformity in ***Cannabis.*** Canad. J. Bot. ***50:*** 1947-1949.

———. 1972*b*. The hemp problem in Canada. Greenh. Gard. Grass 11: 46-52.

———. 1974. American law and the species problem in *Cannabis*. Microgram 12: 131-132.
———. 1975a. Morphological variation of achenes of *Cannabis*. Canad. J. Bot. *53: 978-987.*
———. 1975*b*. Letter to the editor. J. Forensic Sci. 20: 739-741.
———. 1975c. American law and the species problem in *Cannabis:* science and semantics. Bull. Narc. 27 (3): l-20.
———. 1975*d*. On toadstool soup and legal species of marihuana. Pl. Sci. Bull. 21: 34-39.
———. 1976. The forensic taxonomic debate on *Cannabis:* semantic hokum, J. Forensic Sci. 21: 239–251.
——— & H. D. BECKSTEAD. 1973*a*. Common cannabinoid phenotypes in 350 stocks of *Cannabis*. Lloydia 35: 144-165.
——— & ———. 1973*b*. Cannabinoid phenotypes in *Cannabis*. Nature 245: 147–148.
———, ——— & A. CHAN 1975. The evolution of cannabinoid phentotypes in *Cannabis*. Econ. Bot. 29: 219-232.
——— & A. CRONQUIST. 1976. A practical and natural taxonomy for *Cannabis*. Taxon. *(In press.)*
SNEATH, P. H. A. & R. R. SOKAL. 1973. Numerical Taxonomy. San Francisco.
STEARN, W. T. 1974. Typification of *Cannabis sativa* L. Bot. Mus. Leafl. 23: 325-336.
VAVILOV, N. 1. 1926. The origin of the cultivation of "primary" crops, in particular of cultivated hemp. Trudy Prikl. Bot. Selekc. 16: 221-233.