

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