Euphytica **140:** 107–120, 2004.

Potential of marker-assisted selection in hemp genetic improvement

G. Mandolino* & A. Carboni
Istituto Sperimentale per le Colture Industriali, Via di Corticella 133, 40128 Bologna, Italy;
*(*author for correspondence: e-mail: g.mandolino@isci.it)*

Key words: Cannabis, breeding, forensic, marker-assisted selection, molecular markers

Summary

The development and applications of molecular markers to hemp breeding are recent, dating back only to the mid-1990s. The main achievements in this field are reviewed. The analysis of Cannabis germplasm by RAPD, AFLP and microsatellites is discussed, with its consequence for the still debated species concept in Cannabis. DNA-based markers have also been exploited in the field of forensic science, in an attempt to discriminate licit from illicit crop. The main applications of the molecular markers to the breeding, however, have been achieved with the development of markers closely linked to the male sex and to some of the most relevant chemotypes. Active research is carried out by several groups in the field of identification and characterization of the genes involved in fiber quality and quantity, and in the determination of monoecy, another very important target of hemp breeding. Besides, markers associated to new, potentially useful chemotypes are being developed, for the marker-assisted breeding of pharmaceutical Cannabis.

Introduction

Recently, a great expansion of applications and developments in the field of agricultural biotechnology has occurred. Many of these developments rely on genomics' advancements, and a remarkable part of the achievements of this newly born science have been transferred and exploited in the field of plant breeding. However, the extent to which the biotechnologies were exploited in the different crops, depended not only on economic reasons (e.g. the area dedicated to the crop, the added value of the crop itself, the economy relying on the products obtained from that crop), but also on the importance of that crop in the most advanced countries. In this sense, the case of hemp is somehow peculiar. In the United States, one of the most advanced countries from the point of view of the development and applications of agricultural biotechnologies, hemp cultivation is simply not allowed (Morris, 2002), and in several countries the situation was no different, at least until 1998.

Despite the difficulties that hemp cultivation encountered in some countries, several advanced techniques were exploited in the study of this non-food species. Some of the most recent advancements will be reviewed in other contributions of this special issue; in this paper, we will focus on the state of the art of the applications of molecular markers and genomics to the study and genetic improvement of *Cannabis sativa* L.

A brief outline of the marker systems used in Cannabis studies

Several papers reviewed the application and the requirements of marker-assisted selection (MAS) in crop plants (Mohan et al., 1997). Molecular markers employed in MAS should: (i) co-segregate or be tightly linked (ideally, less than 1 cM) to the trait object of selection; (ii) it should lend itself to a mass-screening for the identification of the marker genotype in breeding lines and populations; (iii) its validity should be

recognized in a laboratory-independent manner, i.e. it should be reliable and reproducible in different laboratories.

Historically, the earliest molecular markers to be extensively used in population studies and in plant breeding have been isozymes (Lewontin and Hubby, 1966). Isozyme variation has been exploited to a good extent in plant breeding. The identification of isoenzymatic forms have been associated to specific traits such as disease resistance (e.g. to root knot disease in tomato; Rick and Fobes, 1974); the distribution of the different variant isoforms in natural and domesticated populations has been used to infer the genetic structure and heterozygosity level of plant populations (Brown and Weir, 1983). Besides, the codominant nature of isoenzyme markers (i.e. the heterozygous plants can be distinguished by both homozygous types for the contemporary presence in a starch gel of both the isoforms) made them particularly useful in the estimation of heterozygosity and in studies of gene flow from crop species to their wild relatives (e.g. Bartsch et al., 1999). Isozyme markers are quite reproducible, but the number of known loci corresponding to isoenzymatic systems, is not very high; besides, obtaining a readable pattern of enzymatic activity may be technically difficult in some plant material.

In Cannabis research, isozymes markers have been exploited only to a limited extent. In recent studies on the chemotaxonomy of the genus, in the frame of a germplasm survey, more than 150 Cannabis accessions were analyzed for isozyme variation (Hillig, 2004).

The second marker type used in plant breeding was restriction fragment length polymorphisms (RFLPs). RFLP markers rely on differences, at the genomic DNA level, in the target sequences of the restriction endonucleases; such differences lead to variant DNA fragment length upon restriction, usually visualized by agarose gel separation, followed by hybridization of the immobilized fragments to a labeled (usually radioactively) probe and autoradiography. An RFLP marker is codominant and identifies one single locus at a time; RFLP molecular maps were developed since the 1980s for several important crop species (Beckmann and Soller, 1983; Tanksley et al., 1989). However, the earliest applications of molecular biology techniques to the study of *Cannabis sativa* date back only to the mid-1990s; at that time, the main molecular tools used in plant breeding and genetics were PCR based. As a consequence, RFLP technology has not been widely exploited in hemp. As far as we know, no RFLP markers were developed for hemp, and the molecular maps available today do not contain any of these markers.

Since the introduction of PCR, the strategy of production of molecular markers and of genetic analysis changed. The earliest PCR-based markers to be extensively applied to plant breeding and MAS were the RAPD markers (Williams et al., 1990). In this case, PCR amplification is mediated by short decamer primers of random sequence. Such primers find with a certain frequency annealing sites on the opposite strands of the target DNA molecule, so that an amplification product can be produced from the intervening sequence. The annealing sites of the decamer primers can be variable, and consequently some of the amplified fragments will be polymorphic in the different DNAs. The nature of RAPDs (and of many PCR-based markers) is dominant and "multilocus". As it will be detailed in the next sections, RAPD markers have been successfully applied in Cannabis germplasm analysis and MAS research. The inherent limited between-laboratories reproducibility of many RAPD markers, however, makes in most cases necessary the isolation of the RAPD bands found associated to a trait of interest. The relevant RAPD fragments are gel-isolated, cloned and sequenced; specific 20-mer primers are then designed, amplifying only the sequence found by genetic analysis to be linked to the trait (SCAR markers: Sequence Characterized Amplified Region). A further improvement in marker producing techniques came from the development of amplified fragment length polymorphisms (AFLPs; Vos et al., 1995). The restriction digestion of genomic DNA is followed by a PCR amplification of the fragments obtained, mediated by labeled primers able to anneal to the ends of the fragments, and having a variable number of extra nucleotides randomly chosen, so to amplify only a subset of the total fragments obtained. The labeled amplification products are then run in polyacrilamide gels. These markers are much more reproducible than RAPD markers, are multilocus and dominant; in recent years, the increasingly widespread availability of automatic sequencers/genetic analyzers machines, allowed the scoring of fluorescently labeled AFLP fragments by capillary electrophoresis. The fragments are identified as peak signals detected by coupling charge devices (CCD), upon laser eccitation of the fluorochromes. AFLP markers have been widely used in hemp research, and one of the molecular maps produced has been composed with these markers (see later).

Microsatellites (or simple sequence repeats, SSR) are short sequences of two, three or more nucleotides

that are repeated for a variable number of times in the genome. These markers are codominant; primers are designed on the basis of the DNA sequences flanking the repeat stretch, able to amplify the intervening sequence by PCR. In general, one single locus is identified by each PCR reaction, but the number of alleles that can be identified is very high, as the variability in the repeated motif number is high in the plant genomes (Morgante and Olivieri, 1993). As it will be discussed in the next section, microsatellites have been only very recently identified and used in *C. sativa*.

The forensic applications of DNA markers

The aims of forensic scientists working on Cannabis were mainly two. First, they intended to develop methods capable of recognizing the presence of cannabis (an illicit material in most countries), distinguishing it from other plant sources. Secondly, there was an interest to exploit molecular markers in analyzing suspect plant material, so to reconstruct phylogenies of the different drug strains, and possibly to infer, from their relative distribution, the routes of diffusion of such illicit crops. It must be therefore acknowledged, that molecular tools have been exploited in Cannabis by forensic scientists well before than by plant breeders. The first aim has been accomplished by exploiting sequences from chloroplast DNA (ctDNA), and particularly a short intergenic sequence located between the chloroplast genes for the transport RNAs for the aminoacids leucine and phenylalanine. It was found that when specific primers flanking this sequence were used, amplified products occurred only if the template was made of *C. sativa* DNA; this simple test was developed in a patented product to be used by the law reinforcements organs (Linacre and Thorpe, 1998; Wilkinson and Linacre, 2000). Within this region, only limited nucleotide polymorphisms was detected; this is not surprising, as it is known that chloroplast DNA has both a very low structural evolution and nucleotide synonymous substitution rate. In fact, among the entries present in NCBI for this sequences, 34 are actually the same sequence obtained from different Cannabis sources, by different authors, and are practically coincident, with only a nucleotide missing in one sequence compared to the others (Kohjyouma et al., 2000).

Another approach to identify *C. sativa* at the DNA level exploited the properties of the Internal Transcribed Spacers I and II (ITS1 and ITS2) of the nuclear ribosomal genes. Once amplified and sequenced, this DNA region was found to distinguish univocally *C. sativa* from any other plant species, including the closely related hop; a limited number of single nucleotide polymorphisms (SNPs) were found examining five different Cannabis accessions (Siniscalco Gigliano et al., 1997). As an alternative to the discrimination based on sequencing, a cleaved amplified polymorphic sequence (CAPS) marker test of the amplified ITS1 was also devised, allowing identification of hemp samples, but showing no within-species polymorphisms (Siniscalco Gigliano and Di Finizio, 1997).

The approaches described earlier have been used to distinguish Cannabis from other plants; it is our opinion, however, that the real task, also from the forensic point of view, remains the discrimination of drug from non-drug strains. In a scenario in which *C. sativa* cultivation for fiber, oil or pharmaceutical purposes became widespread, such methods of discrimination of licit from illicit crops are likely to be of limited use, and to generate confusion, rather than to solve it. However, forensic researchers also contributed insights in the plant genetic structure that are useful for hemp breeders too. Protocols were established for the extraction of DNA for AFLP analysis from marihuana samples (Coyle et al., 2003), and RAPD (Gillan et al., 1995; Jagadish et al., 1996) and ISSR (Kojoma et al., 2002) markers were also used in an attempt to establish a more causative relation between the cannabinoid type of the plant and the markers identified. These attempts however were not fully successful, and no correlations of molecular markers with the gas chromatographic or HPLC cannabinoid's profiles of different Cannabis germplasm were detected.

Microsatellites and the genetic variability in Cannabis

Until recently, no information was available about microsatellite loci in *C. sativa*. In 2003, three different groups of forensic researchers reported the isolation, sequencing and use of different microsatellites in the analysis of Cannabis germplasm (Alghanim and Almirall, 2003; Gilmore and Peakall, 2003; Hsieh et al., 2003).

Microsatellite markers were most useful in describing the hemp germplasm. It was found that the most common repeated motif, the dinucleotide GA/CT, is also the most frequently detected not only in the close relative hop, but in general in the plant kingdom (Alghanim and Almirall, 2003; Jake et al., 2001; Toth

et al., 2000). Di- and tri-nucleotide repeats were the most frequent, with an allele number ranging from 2 up to 28. The highest number of alleles detected was found for an exanucleotide repeat (CACCAT), for which 30 alleles were detected in a survey of 108 plant samples (Hsieh et al., 2003); in this work, genotypes are reported showing up to four alleles of different size, suggesting that this particular SSR was multilocus. Table 1 lists the microsatellite loci described so far in Cannabis. In most cases, amplification by primers flanking the repeats yielded one or two bands; all the papers describing these loci, however, only examined Cannabis germplasm (Gilmore et al., 2003); therefore, no data about the heritability of the alleles identified were provided in these studies.

Irrespective of the potential application in the forensic field, some of these studies also provided insights on the genetic structure of hemp. The variability found was very high; in one study, out of 93 plants examined, only four (belonging to the same drug accession) were not distinguishable from each other (Gilmore et al., 2003), while all the other presented different microsatellite combinations at the different loci identified. The average allele number found in all the three studies was 9, but because the relatively limited number of plants examined, it is likely that the identified loci have more

Table 1. The microsatellites known in *Cannabis sativa* L.

Locus	Repeat	Size range (bp)	Number of alleles detected	Reference
Dinucleotide repeats				
ANUCS201	$(GA)_{26}$	161–223	18	Gilmore & Peakall, For. Sci. Intl., 2003
ANUCS202	$(GA)_{20}$	147–185	14	Gilmore & Peakall, For. Sci. Intl., 2003
ANUCS203	$(CT)_{50}$	169–267	28	Gilmore & Peakall, For. Sci. Intl., 2003
ANUCS204	$(CT)_{26}$	128–184	14	Gilmore & Peakall, For. Sci. Intl., 2003
ANUCS205	$(CT)_{21}$	172–242	18	Gilmore & Peakall, For. Sci. Intl., 2003
ANUCS206	$(AT)_{11}$	159–167	4	Gilmore & Peakall, For. Sci. Intl., 2003
C08-CANN2	$(GA)_{21}$	171–203	9	Alghanim & Almirall, Anal. Bioanal. Chem., 2003
H11CANN1	$(CT)_{18}$	285–297	7	Alghanim & Almirall, Anal. Bioanal. Chem., 2003
H09-CANN2	$(GA)_{15}$	204–224	6	Alghanim & Almirall, Anal. Bioanal. Chem., 2003
Trinucleotide repeats				
ANUCS301	$(TTA)_{15}$	209–261	13	Gilmore & Peakall, For. Sci. Intl., 2003
ANUCS303	$(GTG)_7$	141–156	5	Gilmore & Peakall, For. Sci. Intl., 2003
ANUCS305	$(TGG)_{10}$	141–162	7	Gilmore & Peakall, For. Sci. Intl., 2003
ANUCS307	$(ACC)_6$	105–108	2	Gilmore & Peakall, For. Sci. Intl., 2003
B02-CANN2	$(AAG)_{10}$	163–172	3	Alghanim & Almirall, Anal. Bioanal. Chem., 2003
E07-CANN1	$(CTA)_9$	105–111	3	Alghanim & Almirall, Anal. Bioanal. Chem., 2003
B05-CANN1	$(TTG)_9$	235–244	4	Alghanim & Almirall, Anal. Bioanal. Chem., 2003
D02-CANN1	$(GTT)_7$	105–111	3	Alghanim & Almirall, Anal. Bioanal. Chem., 2003
H06-CANN2	$(ACG)_7$	266–273	3	Alghanim & Almirall, Anal. Bioanal. Chem., 2003
Pentanucleotide repeats				
ANUCS501	$(TTGTG)_4$	80–95	3	Gilmore & Peakall, For. Sci. Intl., 2003
Exanucleotide repeats				
CS1F/CS1R	(CACCAT)	18–240	30	Hsieh et al., For. Sci. Intl., 2003
Heterogeneous repeats				
ANUCS302	$(CAA)_7$-$(CAA)_4$	140–173	10	Gilmore & Peakall, For. Sci. Intl., 2003
ANUCS304	$(TCT)_8TCA(TCT)_7$	167–230	15	Gilmore & Peakall, For. Sci. Intl., 2003
ANUCS306	$(GAT)_3$-$(GAT)_6$	92–95	2	Gilmore & Peakall, For. Sci. Intl., 2003
ANUCS308	$(TAA)_3$-$(AT)_5$	177–203	8	Gilmore & Peakall, For. Sci. Intl., 2003
C11-CANN1	$(GAT)_8(GGT)_7$	150–175	5	Alghanim & Almirall, Anal. Bioanal. Chem., 2003
B01-CANN1	$(GAA)_{13}A(GAA)_3$	323–339	5	Alghanim & Almirall, Anal. Bioanal. Chem., 2003
D02-CANN2	$(CTT)_6ATT(CTT)_{10}$	221–236	4	Alghanim & Almirall, Anal. Bioanal. Chem., 2003

alleles than detected. The heterozygosity, or gene diversity, was found to be 0.83 in one study (93 plants, 15 loci) and 0.568 in another (average of 11 loci for 41 plant samples). Alghanim and Almirall (2003) calibrated their neighbor joining tree analysis performed with SSR allele frequencies data, on the basis of previously known associations between the Cannabis samples they examined; they found a good agreement with AFLP data. The extent of variability found was always very high, except for samples probably deriving from clonal propagation of single plants. Given the high level of diversity found, the allele frequencies were on average quite low, usually below 0.30, except for a few alleles occurring at high frequencies in cultivars with a high extent of inbreeding, such as Fibrimon and a drug strain (Gilmore and Peakall, 2003). In this work, 15 accessions were examined, and the molecular data were used to estimate variance partition between and within accessions. The majority of the observed marker variation (73%) was attributable to individual differences within accessions, and only 21% of the variation was due to between-accession diversity. Besides, when the accessions were divided into the two groups "drug" and "non-drug", it was found that only 6% of the variation was attributable to the chemotype, and therefore it was concluded by the authors that there was no clear split or defined boundary between drug and non-drug materials. Conceptually similar results had been obtained also by our group in a survey of six different varieties or inbred lines (10 plants per variety) examined by RAPD markers (102 loci; Forapani et al., 2001); despite the fact that this analysis was conducted using dominant, multilocus markers, there was a basic agreement with Gilmore and Peakall's findings, essentially on two facts. First, the very high degree of variation within the cultivars or accessions, accounting for over 50% of the total observed variation, that was found still high even in inbred lines obtained after two cycles of inbreeding (31% of polymorphic loci in an inbred female line, while for Fibranova, a cross-bred variety, the value found was above 78%; see also Faeti et al., 1996). Secondly, despite several specific markers were found showing significantly higher frequencies in certain cultivars, and therefore potentially useful for variety discrimination, out of 102 markers scored, 68 had a calculated F_{ST} value below the average (0.48); this finding indicated that the majority of the loci had low discriminating power, and confirmed by independent means the existence of a widely shared gene pool in Cannabis, with limited cultivar boundaries and relatively poor loci segregation between different populations (Forapani et al., 2001). This holds true also for drug and non-drug material, as later confirmed by Gilmore and Peakall by SSR markers (2003).

The results reviewed show that different studies point out that no wide split between drug and non-drug accession is possible on the basis of the analysis of the genetic structure of the different varieties and accessions. While some cvs, like monoecious fiber cvs or drug inbred lines, do have a genetic base narrower than most fiber hemp accessions, no safe identification (for the purpose of illicit cultivations repression) of a single or limited number of drug or fiber plants is possible on the basis of the general distribution of molecular markers. This can only be accomplished based on the causal genetic determinants of the drug or non-drug phenotype (i.e. of the chemotype); studies like those described earlier, however, have a great interest from the population genetics and genomic point of view.

Molecular maps

The extremely high degree of variation and heterozygosis found by different authors, especially in dioecious hemp, suggested that many markers could be present at the heterozygous state, and therefore could potentially segregate in F1s, allowing the construction of linkage maps.

Two non-saturated molecular maps of Cannabis have been published so far. The first one derives from a cross between monoecious plants found during a screening for phenotypic traits in the accession CAN 19/86 (from Southern Italy, kindly provided by Dr. Graner, IPK germplasm bank, Gatersleben). Some of the plants having a female habit turned out to be monoecious, and were used as pollen parents in a cross with Carmagnola female plants (males were eradicated before they were recognizable by using a male-specific marker, see later). Other monoecious plants of the 19/86 accession were also isolated to produce S1-selfed seed. The F1 plants were raised, their sexual phenotype scored at maturity, and screened by RAPD markers (Carboni et al., 2000; Mandolino and Ranalli, 2002). In the 19/86 $\times$ Carmagnola cross, 182 markers out of 441 were polymorphic between the two parentals (genotypes *Aa* and *aa*, respectively), and segregated in the progeny. Among the 259 non-polymorphic loci, 31 segregated 3:1 in the progeny and therefore had genotype *Aa* in both parents. For a limited number of non-polymorphic, non-segregating loci, a complete genotyping was obtained from the data relative to the segregation of the same locus in the

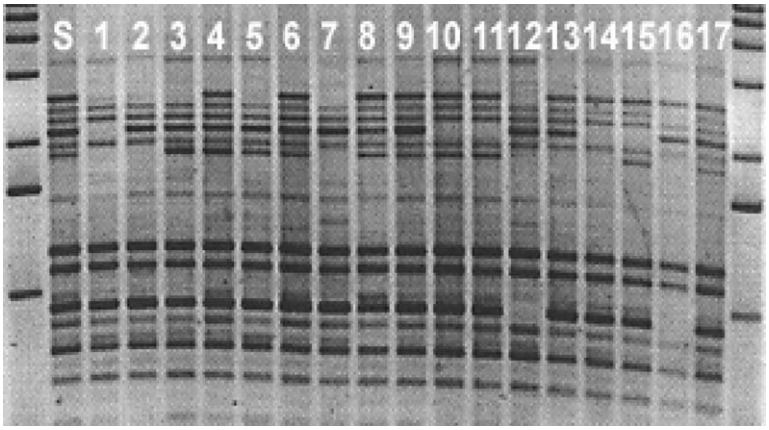

Figure 1. RAPD pattern of a monoecios plant belonging to the CAN 19 accession (lane S), and of part the selfed progeny obtained from it (lanes 1–17). Note the segregation of the markers present in heterozygosity in the parental plant. The figure is the negative of an ethidium-bromide stained gel. In the first and last lanes, the molecular weight standards.

selfing S1 progeny (Figure 1). The markers formed 11 linkage groups in Carmagnola; the monoecious trait, was found segregating 1:1 in the F1 offspring, but it was not placed on the map at standard LOD scores, along with several other markers.

The second map available for *C. sativa*, has been constructed from a cross between male and female plants of another germplasm accession, CAN 18 (Peil et al., 2000), and should therefore enclose the male Y chromosome. This AFLP map shows another characteristics quite commonly found in AFLP maps of plant genomes, i.e. the tendency of the markers to cluster in "hot spot" along the linkage groups. The early map was composed of five linkage groups, while a later version enclosed 10 linkage groups (122 markers; Peil et al., 2001). This latter work also shed some light on the structure and properties of the Y chromosome, as it will be discussed in the next section.

The high between-individual variability for molecular markers found in hemp by different authors, limits the practical utility of a genetic map for *C. sativa*; it is possible that markers appearing closely associated with important traits in one map will not be of general use in MAS when used in different populations. This is a general problem of many molecular markers, but it might reveal particularly important in hemp. Despite of this limitation, different mapping populations could be used for the construction of dense, saturated maps of specific regions of the genome, surrounding different genes of interest (e.g. monoecy, chemotype). From such regions of high marker density, it will probably be possible to identify and isolate the genes by positional cloning, and to obtain markers to be used in MAS.

DNA markers and sex phenotype

In nature, *C. sativa* is a dioecious plant; however, monoecious varieties were also developed, and proved to have distinctive advantages, such as their high seed yield, a higher uniformity compared to dioecious varieties, and ease of mechanical harvesting. Drawbacks of monoecy are a narrower genetic base, necessary to maintain the monoecious trait in a significant proportion of the plants, and the necessity of strict isolation and seed batch control, due to the lower competitiveness of monoecious pollen compared with pollen from male plants, with consequent high probability of unintended pollination from dioecious males. In dioecious varieties, it is in some cases necessary to identify and score the male plants for fiber quality, and to allow the pollination only to the best-scoring males (the Bredemann principle). Therefore, the practice of hemp breeding requires in several occasions the identification of the sexes (male, female, monoecious), with the male sex being especially important for the different strategies of improvement in dioecious and for quality controls in monoecious hemp.

It is now a well-established fact that *C. sativa* is endowed with heteromorphic chromosomes. The male sex is characterized by a Y chromosome, reported to be larger than the X chromosome, and presumably responsible for the extra 47 Mbp characterizing the male genome, as measured by flow citometry (1636 Mbp for females and 1683 Mbp for males; Sakamoto et al., 1998). These authors also found that the satellite and the long arm of Y chromosome were the earliest structure to condensate at the metaphase stage, suggesting

that its sequence composition might be peculiar and different from the short Y arm, the X chromosome and the autosomes. This characteristics is paralleled by different reports of DNA markers found to be specific of male plants, or showing male–female polymorphism. Sakamoto et al. (1995) found two male-associated RAPD markers after a screening of five male and five female plants with only 15 random primers. One of these markers (730 bp long) included a 164 bp sequence that, when used as an hybridization probe, yielded male-specific patterns in *Eco*RI, *Bam*HI and *Hind*III digests of male and female DNAs. The whole male-specific RAPD fragment was sequenced and named MADC1; it did not include a long ORF, and no significant homologies with other sequences present in gene banks were identified. Sakamoto et al. did not develop from this marker a SCAR marker or a rapid test, suitable for identification of male and female plants in MAS. Mandolino et al. (1997, 1998), upon RAPD screening of Cannabis germplasm, identified a 400 bp marker almost exclusively associated to the male phenotype (Figure 2a). This marker was cloned and sequenced, and specific 20-mer primers were designed. These primers were able to amplify a 391 bp fragment in all male plants, and two larger DNA fragments in female and monoecious plants (Mandolino et al., 1999). It was therefore a marker exploiting the organization of the genome regions surrounding the primer's annealing sites, rather than identifying a sequence solely present in male plants. This sex-associated SCAR marker was unable to distinguish male and female DNAs when it was used as an hybridization probe in Southern blots of DNA digested with several restriction enzymes. Double digestion of genomic DNA and hybridization to the isolated and cloned 400 bp fragment (Figure 2b) suggested the possibility that the sequence (named MADC2) corresponding to the marker was present in more than one copy in the hemp genome (Figure 2c).

The two sequences, MADC1 (Sakamoto et al., 1995) and MADC2 (Mandolino et al., 1999), shared some features, like the G + C content (39.9% for MADC1 and 40.4% for MADC2) and the absence of significant ORFs. Like MADC1, MADC2 has limited homology to other known sequences. However, all the homologies reported by BLAST analysis, were with repeated regions or retrotransposon-like sequences; this is also in agreement with the fact that MADC1 is a LINE-like retrotrasposon, physically located on the Y chromosome, as evidenced later by FISH (Sakamoto et al., 2000). It was known that DNA from sex chromosomes in different dioecious species are rich in retrotransposon-like sequences (Clark et al., 1993), and this also seems to be the case in *C. sativa*. The location of MADC2 on the Y chromosome was suggested not by direct evidence, but from genetic analysis; in fact, it was shown that in different F1 progenies of a cross between male and female plants, no recombination between the SCAR marker developed from MADC2 and the sexual phenotype occurred (Mandolino et al., 2002). Therefore, this marker satisfied the requirement of general association to the male phenotype, irrespective of the population screened; as a consequence, it was ideally suited for the development of a quick method of sex identification, based on direct PCR of leaf fragments. This "boil and amplify" method (Mandolino and Ranalli, 2002) allows the 100% safe identification of male plants in both dioecious and monoecious varieties, at stages as early as necessary, with no need of previous preparation of intact genomic DNA; only a few milligrams of plant tissue are necessary to identify the male plants.

Male-associated markers have been described also using AFLPs (Flachowsky et al., 2001; Peil et al., 2003). In this case, however, there have apparently been problems in the correct scoring of the sexual phenotype, especially in the case of field-grown plants. The reasons for these difficulties are not clear; the two accessions used for the study (CAN 17 and CAN 18) are from IPK germplasm bank, and are originated from Hungary and Germany, respectively. It is possible that, at least in some cases, sex in these accessions was subject to incomplete expression, or partial reversion of the phenotype; besides, it is known that the ploidy level can influence the expression of the sexual phenotype, despite the presence of a Y chromosome. Though these phenomena have indeed been described in hemp, it is difficult to explain why they should only occur in field-grown plants. Therefore, the usefulness of these markers in MAS is still uncertain. In a later work, Peil et al. (2003) assigned 43 more AFLP markers to the Y chromosome, on the basis of the complete linkage of these markers with the male sex in a progeny of 80 plants deriving from a male × female cross; besides, the observation of the skewed segregation ratios of several other markers led the authors to hypothesize the existence, on the Y chromosome of male plants, of a region (pseudo-autosomal region, PAR), able to recombine to some extent with the corresponding portion of the X chromosome, similarly to what observed in other dioecious species with sex chromosomes (Di Stilio et al., 1998). Other research groups described markers linked to the male sex (NCBI nucleotide sequences AF364954 and

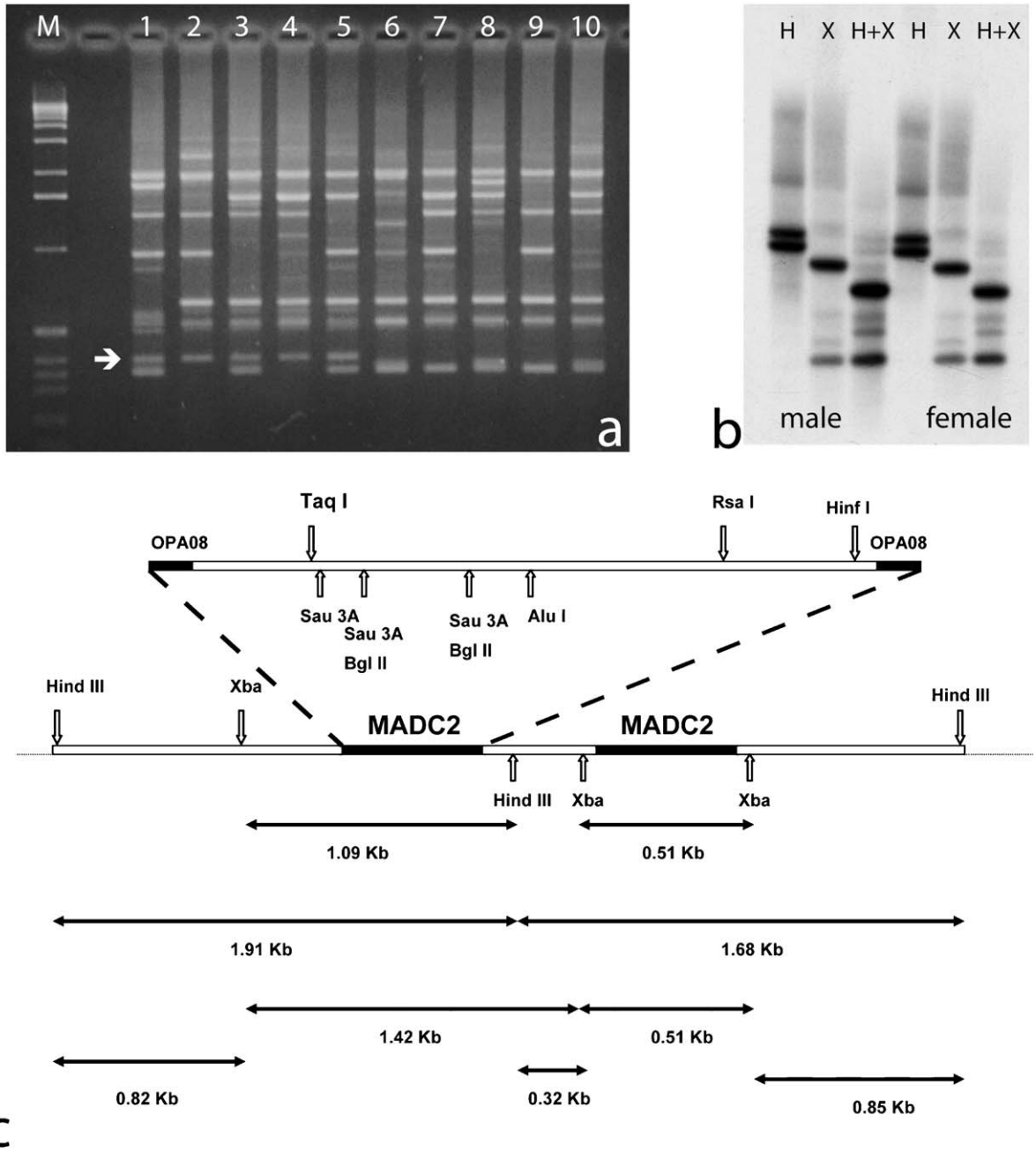

Figure 2. (a) RAPD pattern generated by the primer OPA08 (Operon Technologies, USA). The arrow indicate the 400 bp male-associated marker, present in the male plants (lanes 1–5) and absent in the female plants (lanes 6–10). M, molecular weight standard. (b) Hybridization pattern obtained using the cloned 400 bp RAPD fragment visible in a) on digested DNA from male and female plants restricted with *Hind*III (H), *Xho*I (X) and with both enzymes (H + X). Note the absence of male–female polymorphism. (c) Restriction map of the genome region containing the sequence corresponding to MADC2.

AF364955, MADC5 and MADC6, by Torjek et al.), and also female-associated RAPD markers were reported (Shao et al., 2003), though in the latter case no chromosome-specific location can be hypothesized, and it was not specified whether this marker was able to distinguish female from monoecious plants.

The monoecious trait is another very important character for which MAS would be extremely useful.

As stated earlier, it is already available a male-identifying marker that can be fruitfully used in keeping under strict control the number of "contaminating" male plants present in a monoecious stand or seed lot, allowing for instance the introduction of the "elite" or "superelite" terms for seed batches in a very precise and controllable manner. The marker identifying male plants can be therefore already exploited in MAS and in seed quality certification of monoecious varieties. However, the availability of a marker, presumably autosomic, tightly linked to the monoecious trait itself would be extremely important. Unfortunately, no reports of such a marker have yet been made until today; as described in the fifth section, the monoecious trait did not integrate, under the mapping conditions used, in the map of CAN 19 accession obtained so far (Carboni et al., 2000). The identification of monoecious-associated markers is also complicated by the strong environmental influences altering the expression of the male flowers in monoecious plants, so that under different conditions the ratio female:monoecious plants in a monoecious seed lot can vary quite strongly. In the CAN 19 (monoecious) × Carmagnola (female) cross described in fifth section, the monoecious trait segregated 1:1, and the segregating groups can be used for the identification of monoecious-associated markers through the bulk segregant analysis strategy (Paran and Michelmore, 1993). Alternatively, these segregating materials, in which the genome background is randomized into the two groups, could be used for isolating the pertinent genes, *via* reverse genetics approaches (see also Moliterni et al., in this special issue). The monoecious trait is presently one of the main targets for biotechnology applied to hemp research, and it is likely that we will assist to major breakthroughs in the next future; the availability of a male- and of a monoecious-specific marker, possibly combined in a single PCR assay, would allow the complete identification of the sexual phenotype of plants belonging to any variety.

Molecular markers for the chemotype

Today, it is generally accepted that three major "chemotypes" (i.e. "chemical phenotypes", plants characterized by a defined cannabinoid profile) exist in *C. sativa*; these chemotypes were formally defined about 30 years ago (Small and Beckstead, 1973) on the basis of the content ratio of the two major cannabinoids found in hemp, Δ^9-tetrahydrocannabinol (THC) and cannabidiol (CBD), expressed as percent of the inflorescence dry matter. The three chemotypes (I, prevalent THC; II, THC and CBD; and III, prevalent CBD) were found in several ecotypes and varieties, though some of the most inbred materials, such as the French monoecious cultivars, only showed two chemotypes (Fournier and Paris, 1980). The tripartite distribution of chemotypes (Figure 3) had suggested a simple genetic determinism, based on one or few genes (Becu et al., 1998). However, only recently accurate progeny analysis experiments demonstrated the existence of one locus, named *B*, with at least two alleles, B_T and B_D, each responsible for the synthesis of the two most common cannabinoids, THC and CBD, and probably coding for the respective synthases (De Mejier et al., 2003; Mandolino et al., 2003).

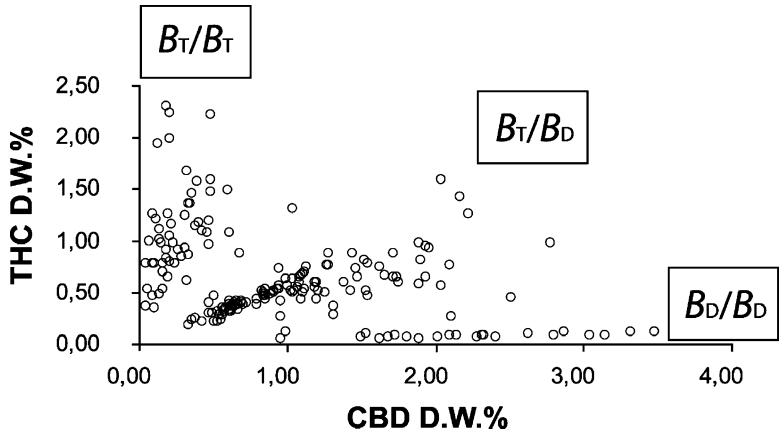

Figure 3. Chemotype distribution of the old Italian fiber cultivar "Eletta Campana". This tripartite distribution is typical of old ecotypes before the counter-selection for THC begun, and directly visualizes the distribution of B_D and B_T alleles.

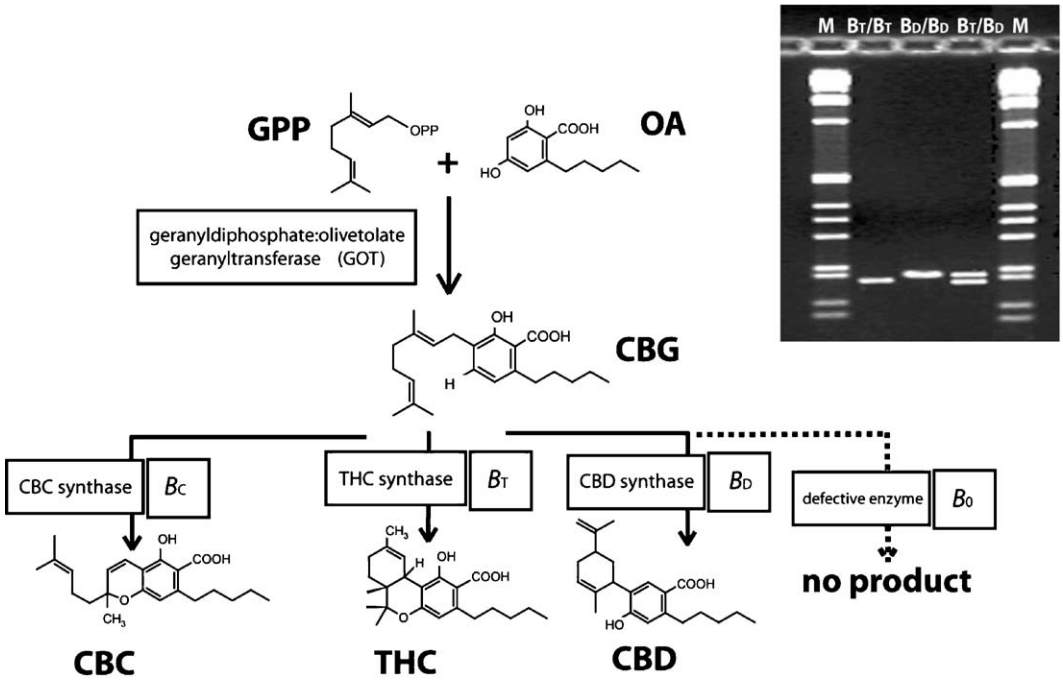

Figure 4. The biosynthetic pathway of the most common cannabinoids in Cannabis plants. For each step, the relative enzyme has been indicated (if known), and the state of the alleles at the B locus is proposed, accounting for the chemical phenotype. The inset shows the *B190/B200* markers, obtained amplifying Cannabis DNA with the RAPD-deriving SCAR primers. Note the codominancy of this marker.

The model illustrating the enzymes involved and the alleles reputed responsible for the different steps of cannabinoid biosynthesis, is shown in Figure 4. The condensation of geranygeraniol diphosphate with olivetolic acid (catalyzed by geranylgeraniol:olivetolate transferase, GOT; Fellermeier and Zenk, 1998) is the step leading to the first Cannabis' exclusive product, cannabigerol (CBG); this particular compound was also described as the prevalent cannabinoid in some plants (Fournier et al., 1987). These "mutants" could not therefore be considered belonging to any of the three formerly known chemotypes, and were assigned to a new chemotype (prevalent CBG, or chemotype IV; Figure 5). CBG is today widely accepted as the common precursor for the synthesis of both THC and CBD (Fellermeier et al., 2001). A further chemotype was found, with an undetectable amount of cannabinoids. This "zero cannabinoid" type, (we propose for it the creation of a chemotype V; see Figure 5 for its gas-chromatogram) has been described by some authors in different germplasm (G. Grassi and I. Virovets, personal communication), though it is not yet clear whether this absence was due to a metabolic block at the level of GOT, or rather to a very limited formation of glandular trichomes, the site of synthesis and accumulation of cannabinoids (Kim and Mahlberg, 2003). Usually, CBG is detected in very small amounts in Cannabis' extracts, probably because it is almost completely utilized as substrate by the downstream synthases (THC- and CBD-synthases), transforming it into THC, CBD or other less common end products (such as cannabichromene, CBC; Figure 4). The two synthases are respectively coded by B_D and B_T, the two alleles at the B locus, and have very similar K_m and V_{max} (Taura et al., 1995, 1996). This peculiarity explains the fact that, when both enzymes are present (i.e. when the genotype at the B locus is $B_D\,B_T$), the almost equal efficiency of oxidocyclization of CBG into the respective end products, leads to a ratio close to the unity in the THC:CBD ratio. This produces the distribution along the median line of the THC vs. CDB scatter plots observed for chemotype II plants, and for all the F1 progenies of pure chemotype parentals.

One of the main targets for fiber hemp has been for a long time the eradication of hemp plants bearing the B_T allele, and consequently synthesizing more or less "illegal" amounts of THC. There have been, in the

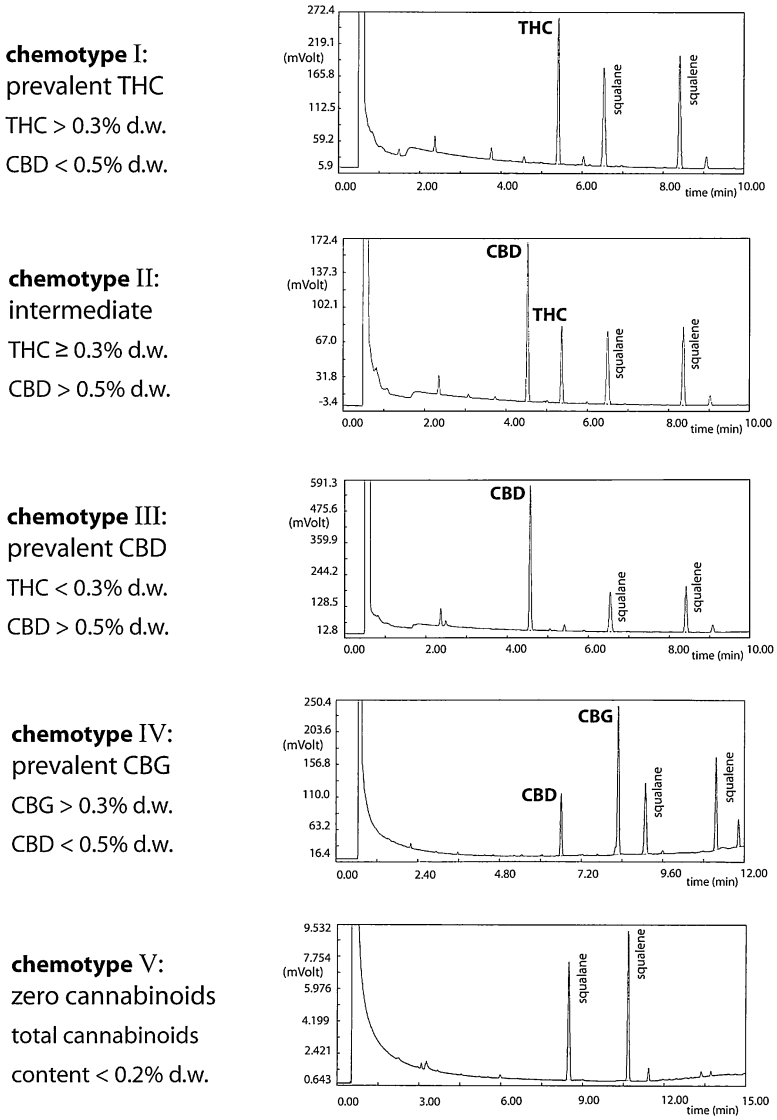

Figure 5. The gas chromatographic profiles of the different chemotypes of *Cannabis sativa* (see text for details).

hemp breeding community, proposals of total elimination of THC, irrespective of its amount (even if well below the 0.2% needed to get the EU subsidies), and rumors of creation of transgenic hemp plants in which THC synthesis was completely suppressed by antisense technology. It is our opinion that these approaches are unnecessary, and that standard selection practices, with the possible help of MAS, can be sufficient to limit the THC content well below the 0.2% of inflorescence dry weight. However, in some fiber ecotypes, like the one for which chemotypes distribution is shown in Figure 3 (an old Italian fiber ecotype, Eletta Campana), the number of plants that should be considered homozygous for THC was indeed not negligible; besides, the absolute amount of THC in the inflorescences, though not at the levels of the drug strains, is high enough to make this cultivar ineligible for EU subsidies. At the beginning of the 1990s, this situation was common for many dioecious fiber cultivars, and therefore the necessity arose, for an effective presence on the market, to "clean" the seed batches from THC-producing plants, without altering the overall genetic background. This has been the aim of the work leading to the identification of chemotype-associated molecular markers, made in our Institute in collaboration with E. de Meijer, breeder at Horta Pharm BV (The Netherlands) before,

and presently at GW Pharmaceuticals (UK). Different segregating F2s were obtained from initial crosses between inbred lines with contrasting chemotypes (I and III, i.e. almost pure THC and almost pure CBD); the genetic analysis of the gas-chromatographic data demonstrated that the F1 offspring was completely hybrid (chemotype II), while all three chemotypes were again present in the F2 generations, in a 1:2:1 proportion (pure THC:mixed THC + CBD:pure CBD) within each progeny; this finding was in agreement with the hypothesis of one gene and two codominant alleles (B_D and B_T) for chemotype determination. This hypothesis is not the only possible, but it is the simplest explaining of the presently available data. The F2 segregating groups were screened by RAPD markers using the bulk segregant approach, and several CBD- or THC-associated markers were identified. All these markers behaved as dominant, except one (named *B190/B200*; Figure 4 insert), deriving from a CBD-associated RAPD fragment, that once transformed into a SCAR marker, turned out to be codominant, and therefore able to genotype completely at the *B* locus the plants. The efficiency of correct identification of the chemotypes was 88% for pure THC plants, 95% for mixed chemotype plants, and 98% for pure CBD plants (de Meijer et al., 2003). However, these markers, very useful within the pedigrees created from the starting inbred lines, were not equally effective in unrelated materials, like the dioecious fiber varieties Carmagnola, Fibranova or Eletta Campana. Besides, despite the very good degree of association with the chemotype shown by marker *B190/B200*, it cannot be taken into consideration for the marker-assisted identification of illicit crops and for legal purposes (P. Cantin, personal communication). In this case, in fact, a marker must be 100% linked to the chemotype, for its exploitation as an effective and reliable drug repression tool. The only marker with these characteristics is of course the gene itself. In the NCBI database, there are the sequences corresponding to the genes for the THC- and CBD-synthases (entry numbers AB057805, E55107, E55108, E55090 and E55091); these sequences have been patented by a research group of the Taisho Pharmaceuticals Company, Japan.

The sequences of the genes coding for THC- and CBD-synthase show very high similarities; the identity along the 1635 bp coding sequence is 89.3%. The major difference is apparently a missing nucleotide triplet in the positions 757–759 of the THC-synthase sequence. The translated protein sequence is 545 and 544 aminoacids, for CBD- and THC-synthase, respectively. The THC-synthase has a missing aminoacid (SER) in position 253 of the sequence. Out of the 545 aminoacids stretch, only 87 (16%) are different between the two enzymes (including the missing one); about half of these variations, however, are between aminoacids of the same type. The aminoacid changes are quite evenly distributed throughout the sequence, the longest variant stretch consisting of six aminoacids in positions 491–496. These differences are large enough to allow the construction of specific primers, able to identify in the different chemotypes the allelic complement of each plant. In our laboratory, we devised a three-primers system able to amplify, in a single PCR reaction of leaf tissue fragments, the DNA sequences identifying the allelic status at the *B* locus (A. Carboni, unpublished).

Perspectives

Genetic analysis of heterozygous ($B_D\,B_T$) plants from different crosses, revealed that the THC:CBD ratio may vary slightly but consistently and heritably around the value of 1 (de Meijer et al., 2003). This suggests the possibility that several isoenzymatic forms of THC- and CBD-synthases exist in different germplasm. Confirmation of this hypothesis, presently in progress through the sequencing in our laboratory of these possible variants, could lead to the identification of further alleles of potential interest at the *B* locus. Besides, the identification, either by progeny analysis or by direct sequencing, of the alleles responsible for the synthesis of the several cannabinoids described in *C. sativa*, would open the possibility of assisted selection in Cannabis, bred not only as a fiber crop, but also for its pharmaceutical applications (see G. Guy and R. Pertwee contributions in this special issue). Chemotype IV and V, having CBG or no cannabinoids, are of remarkable interest for both fiber and pharmaceutical purposes; the identification of the alleles at the *B* locus responsible for the accumulation of CBG or for the absence of cannabinoids (Figure 4), would open the way to the development of molecular markers for these chemotypes. The knowledge of the genes and alleles responsible for the different chemotypes could also lead to the manipulation of the pathway in both plants and cell cultures; the availability of chemotype-specific cell cultures in which the cannabinoid biosynthesis is made active by manipulations of the key enzymes of their pathway, could lead to the development of bioreactors useful for the *in vitro* large-scale production of specific cannabinoids for the pharmaceutical industry.

However, *C. sativa* remains primarily a fiber crop, and a great deal of work is being done for improving

our knowledge of gene expression during fiber development. The physical properties of the fibers depend on their chemical characteristics and on the way they are assembled into bundles. Fiber quality is strongly influenced by the chemical composition of elementary fibers and cell wall. In the latest years, the completion of the genome sequence of *Arabidopsis thaliana* made available a great deal of information on gene sequences. About 15% of the about 25.000 Arabidopsis genes probably codes for functions correlated to the biosynthesis, assembling and modification of the cell wall (Carpita et al., 2001). Only a very limited number of the genes involved have been so far identified, though an increasing number of xylem-specific EST sequences is available from studies on pine and poplar, Arabidopsis and Zinnia (Boudet et al., 2003). A high cellulose content, a low degree of lignification and a reduced number of cross linking between the pectines and the structural components of the wall, are the main characteristics reputed to be important to obtain a easily extractable fiber and a good quality, for both textile and paper industry. The identification of the genes responsible for these characteristics will lead to the development of new varieties through genetic engineering strategies, and/or through the use of molecular markers associated to the allelic variants conferring the desirable traits during the selection or the germplasm screening work. Because the high number of genes probably involved, the technology of choice to identify them might be genome-wide microarray analysis (M. Toonen, unpublished data).

The renewed interest in *C. sativa* as a multi-purpose crop has been the main driving force for the increasing application of the genomic and molecular tools to the breeding of this species by several research group; the biotechnological products obtained from these researches are likely to lead in the near future to major advancements in the areas of fiber, oil, pharmaceuticals and food industry, and this in turn should hopefully widen the cultivated area dedicated to this crop.

References

Alghanim, H.J. & J.R. Almirall, 2003. Development of microsatellite markers in *Cannabis sativa* for DNA typing and genetic relatedness analysis. Anal Bioanal Chem 376: 1225–1233.

Bartsch, D., M. Lehnen, J. Clegg, M. Pohl-Orf, I. Schuphan & N. Ellstrand, 1999. Impact of gene flow from cultivated beet on genetic diversity of wild sea beet populations. Mol Ecol 8: 1733–1741.

Beckmann, J.S. & M. Soller, 1983. Restriction fragment length polymorphisms in genetic improvement: Methodologies, mapping and costs. Theor Appl Genet 67: 35–43.

Becu, D.M.S., H.D. Mastebroek & H.J.P. Marvin, 1998. Breeding for root knot nematode resistance in hemp. In: Proceedings of 'Bast Fibrous Plants Today and Tomorrow', 28–30 September, p. 149, St. Petersburg.

Boudet, A.M., S. Kajita, J. Grima-Pettenati & D. Goffner, 2003. Lignin and lignocellulosics: A better control of synthesis for new and improved uses. Trends Plant Sci 8: 576–581.

Brown, A.H.D. & B.S. Weir, 1983. Measuring genetic variability in plant populations. In: S.D. Tanksley & T.J. Orton (Eds.), Isozymes in Plant Genetics and Breeding: Vol. 1A. Developments in Plant Genetics and Breedings, pp. 219–240. Elsevier, Amsterdam.

Carboni, A., C. Paoletti, V.M.C. Moliterni, P. Ranalli & G. Mandolino, 2000. Molecular markers as genetic tools for hemp characterization. In: Proceedings of Bioresource Hemp, 13–18 September, Wolfsburg. On line at www.nova-institut.de/bioresource-hemp/op.htm.

Carpita, N., M. Terney & M. Campbell, 2001. Molecular biology of plant cell wall: Searching the genes that define structure, architecture and dynamics. Plant Mol Biol 47: 1–5.

Clark, M.S., J.S. Parker & C.A. Ainsworth, 1993. Repeated DNA and heterochromatine structure in *Rumex acetosa*. Heredity 70: 527–536.

Coyle, H.M., G. Shutler, S. Abrams, J. Hanniman, S. Neylon, C. Ladd, T. Palmbach & H.C. Lee, 2003. A simple DNA extraction method for marijuana samples used in amplified fragment length polymorphism (AFLP) analysis. J Forensic Sci 48: 343–347.

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

Di Stilio, V.S., R.V. Kesseli & D.L. Mulcahy, 1998. A pseudoautosomal random amplified polymorphic DNA marker for the sex chromosomes of *Silene dioica*. Genetics 149: 2057–2062.

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

Fellermeier, M. & M.H. Zenk, 1998. Prenylation of olivetolate by a hemp transferase yields cannabigerolic acid, the precursor of tetrahydrocannabinol. FEBS Lett 427: 283–285.

Fellermeier, M., W. Eisenreich, A. Bacher & M.H. Zenk, 2001. Biosynthesis of cannabinoids. Incorporation experiments with ^{13}C-labeled glucoses. Eur J Biochem 268: 1596–1604.

Flachowsky, H., E. Schumann, W.E. Weber & A. Peil, 2001. Application of AFLP for the detection of sex-specific markers in hemp. Plant Breed 120: 305–309.

Forapani, S., A. Carboni, C. Paoletti, V.M.C. Moliterni, P. Ranalli & G. Mandolino, 2001. Comparison of hemp (*Cannabis sativa* L.) varieties using Random Amplified Polymorphic DNA markers. Crop Sci 41: 1682–1689.

Fournier, G. & M. Paris, 1980. Détermination de chimiotypes à partir des cannabinoïdes chez le chanvre à fibres monoïque (*Cannabis sativa* L.) Possibilités de sélection. Physiologie Végétale 18: 349–356.

Fournier, G., C. Richez-Dumanois, J. Duvezin, J.-P. Mathieu & M. Paris, 1987. Identification of a new chemotype in *Cannabis sativa*: Cannabigerol-dominant plants, biogenetic and agronomic prospects. Planta Med 53: 277–280.

Gillan, K., M.D. Cole, A. Linacre, J.W. Thorpe & N.D. Watson, 1995. Comparison of Cannabis sativa by Random Amplified of Polymorphic DNA (RAPD) and HPLC of cannabinoids: A preliminary study. Sci Justice 35: 169–177.

Gilmore, S. & R. Peakall, 2003. Isolation of microsatellite markers in *Cannabis sativa* L. (marijuana). Mol Ecol Notes 3: 105–108.

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

Hillig, K., 2004. Genetic evidence for speciation in Cannabis (Cannabaceae). Gen Res Crop Evol, in press.

Hsieh, H.M., R.J. Hou, L.C. Tsai, C.S. Wei, S.W. Liu, L.H. Huang, Y.C. Kuo, A. Linacre & J.C.I. Lee, 2003. A highly polymorphic STR locus in *Cannabis sativa*. Forensic Sci Int 131: 53–58.

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

Jake, J., K. Kindlhofer & B. Javornik, 2001. Assessment of genetic variation and differentiation of hop genotypes by microsatellites and AFLP markers. Genome 44: 773–782.

Kim, E.S. & P.G. Mahlberg, 2003. Secretory vescicle formation in the secretory cavity of glandular trichomes of *Cannabis sativa* L. (Cannabaceae). Mol Cells 15: 387–395.

Kohjyouma, M., I.-J. Lee, O. Iida, K. Kurihara, K. Yamada, Y. Makino, S. Sekita & M. Satake, 2000. Intraspecific variation in *Cannabis sativa* L. based on intergenic spacer region of chloroplast DNA. Biol Pharm Bull 23: 727–730.

Kojoma, M., O. Iida, Y. Makino, S. Sekita, M. Satake, 2002. DNA fingerprinting of Cannabis sativa using Inter-Simple Sequence Repeat (ISSR) amplification. Planta Med 68: 60–63.

Lewontin, R.C. & J.L. Hubby, 1966. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. Genetics 54: 595–609.

Linacre, A. & J. Thorpe, 1998. Detection and identification of cannabis by DNA. Forensic Sci Int 91: 71–76.

Mandolino, G., V. Faeti, A. Carboni & P. Ranalli, 1997. A 400 bp marker tightly linked to the male phenotype in dioecious hemp. In: Proceedings of the 2nd Bioresource Hemp Symposium, 27 February–2 March, pp. 195–196, Frankfurt, Germany.

Mandolino, G., A. Carboni, S. Forapani & P. Ranalli, 1998. DNA markers associated with sex phenotype in hemp (*Cannabis sativa* L.). In: Proceedings of the Bast Fibrous Plants Today and Tomorrow Meeting, 28–30 September, pp. 197–201, St. Petersburg, Russia.

Mandolino, G., A. Carboni, S. Forapani, V. Faeti, P. Ranalli, 1999. Identification of DNA markers linked to the male sex in dioecious hemp (*Cannabis sativa* L.). Theor Appl Genet 98: 86–92.

Mandolino, G., A. Carboni, M. Bagatta, V.M.C. Moliterni & P. Ranalli, 2002. Occurrence and frequency of putatively Y chromosome linked DNA markers in *Cannabis sativa* L. Euphytica 126: 211–218.

Mandolino, G. & P. Ranalli, 2002. The application of molecular markers in genetics and breeding of hemp. J Ind Hemp 7: 7–23.

Mandolino, G., M. Bagatta, A. Carboni, P. Ranalli & E.P.M. de Meijer, 2003. Qualitative and quantitative aspects of the inheritance of chemical phenotype in Cannabis. J Ind Hemp 8: 51–72.

Mohan, M., S. Nair, A. Baghwat, T.G. Krishna, M. Yano, C.R. Bhatia & T. Sasaki, 1997. Genome mapping, molecular markers and marker-assisted selection in crop plants. Mol Breed 3: 87–103.

Morgante, M. & A.M. Olivieri, 1993. PCR-amplified microsatellites as markers in plant genetics. Plant J 3: 175–182.

Morris, D., 2002. Why has the hemp revolution bypassed the United States? J Ind Hemp 7: 61–65.

Paran, I. & R.W. Michelmore, 1993. Development of PCR-based markers linked to downy mildew resistance in lettuce. Theor Appl Genet 85: 985–993.

Peil, A., E. Schumann, H. Flachowsky, U. Kriese, M. El Ghani, M. Riedel & W.E. Weber, 2000. AFLP markers for male plants of hemp (*Cannabis sativa* L.). In: Proceedings of the Bioresource Hemp, 13–18 September, Wolfsburg, Germany. On line at www.nova-institut.de/bioresource-hemp/op.htm.

Peil, A., H. Flachowsky, E. Schumann & W.E. Weber, 2001. In: Proceedings of the International Conference on "Bast Fibrous Plants on the Turn of the Second and Third Millennium", 18–22 September, pp. 1–7, Shenyang, China.

Peil, A., H. Flachowsky, E. Schumann & W.E. Weber, 2003. Sex-linked AFLP markers indicate a pseudoautosomal region in hemp (*Cannabis sativa* L.). Theor Appl Genet 107: 102–109.

Rick, C.M. & J. Fobes, 1974. Association of an allozyme with nematode resistance. Tomato Genet Coop Rep 24, 25.

Sakamoto, K., K. Shimomura, Y. Komeda, H. Kamada & S. Satoh, 1995. A male-associated DNA sequence in a dioecious plant, *Cannabis sativa* L. Plant Cell Physiol 36: 1549–1554.

Sakamoto, K., Y. Akiyama, K. Fukui, H. Kamada & S. Satoh, 1998. Characterization, genome sizes and morphology of sex chromosomes in hemp (*Cannabis sativa* L.). Cytologia 63: 459–464.

Sakamoto, K., N. Ohmido, K. Fukui, H. Kamada & S. Satoh, 2000. Site-specific accumulation of a LINE-like retrotransposon in a sex chromosome of the dioecious plant *Cannabis sativa*. Plant Mol Biol 44: 723–732.

Shao, H., S.-J. Song & R.C. Clarke, 2003. Female-associated DNA polymorphisms of hemp (*Cannabis sativa* L.). J Ind Hemp 8: 5–9.

Siniscalco Gigliano, G., P. Caputo & S. Cozzolino, 1997. Ribosomal DNA analysis as a tool for the identification of *Cannabis sativa* L. specimens of forensic interest. Sci Justice 37: 171–174.

Siniscalco Gigliano, G. & A. Di Finizio, 1997. The *Cannabis sativa* L. fingerprint as a tool in forensic investigations. Bull Narcotics 1: www.unodc.org/unodc/en/bulletin/bulletin_1997-01-01_1_page007.html.

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

Tanksley, S.D., N.D. Young, A.H. Paterson & M.W. Bonierbale, 1989. RFLP mapping in plant breeding: new tools for an old science. Biotechnology 7: 264.

Taura, F., S. Morimoto & Y. Shoyama, 1995. First direct evidence for the mechanism of delta 1-tetrahydrocannabinolic acid biosynthesis. J Am Chem Soc 38: 9766–9767.

Taura, F., S. Morimoto & Y. Shoyama, 1996. Purification and characterization of cannabidiolic-acid synthase from *Cannabis sativa* L. J Biol Chem 271: 17411–17416.

Toth, G., Z. Gáspári & J. Jurka, 2000. Microsatellites in different eukaryotic genomes: Survey and analysis. Genome Res 10: 967–981.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper & M. Zabeau, 1995. AFLP: A new technique for DNA fingerprinting. Nucleic Acids Res 23: 4407–4414.

Wilkinson, M. & A. Linacre, 2000. The detection and persistence of *Cannabis sativa* DNA on skin. Sci Justice 40: 11–14.

Williams, J.G.K., A.R. Kubelik, K.J. Livak, J.A. Rafalski & S.V. Tingey, 1990. DNA polymorphism amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18: 6531–6535.